Cultural differences in the persuasiveness of evidence types
in France and the Netherlands

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. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op woensdag 18 januari 2006
des ochtends om 10.30 uur precies

door

Jos Mathieu Albert Hornikx
geboren op 17 februari 1979
te Eindhoven

Promotores:	prof. dr. H. Hoeken
	prof. dr. P.J. Schellens
Copromotor:	dr. M. Starren
Manuscriptcommissie:	prof. dr. H. Beentjes
	prof. dr. F.H. van Eemeren (Universiteit van Amsterdam)
	prof. dr. M. Gerritsen
	prof. dr. H. Jacobs
	prof. dr. W. Spooren (Vrije Universiteit Amsterdam)

Publisher F&N Boekservice
Cover Jos Hornikx
Set in Book Antiqua
ISBN 10: 90-7383859-2
ISBN 13: 978-90-7383859-8

Preface

In the period that I have worked on this dissertation, I have been surrounded by many colleagues and friends, who have made this an enjoyable phase in my life.

Throughout the project, Hans Hoeken has proven to be a brilliant supervisor. Hans, your great knowledge of research methodology, your innovative ideas, and the speed with which you provide simple solutions to complicated problems has helped and stimulated me to improve myself. Our conversations were always constructive and unbelievably efficient. Thank you for your inspiration. Peter Jan Schellens, our collaboration was short, but your ideas and suggestions greatly helped me to improve and refine my manuscript in the past year. Marianne Starren, I am grateful for the way you encouraged me to do a PhD, and for your listening ear when I was disappointed in my own tight schedules involving transport, hotels, and experiments. Carel Jansen, thank you for putting trust in me, and for your supervision during the first three years of the project.

If anyone is thinking of doing a PhD, I recommend that they do so at the Department of Business Communication Studies in Nijmegen, where a pleasant work ambiance and high academic quality go hand in hand. I would like to thank all my colleagues for their interest in my work but mention a few in particular. Renske van Enschot and Elizabeth de Groot, thank you for our discussions about writing a PhD thesis, and its consequences for the everyday things in life. Rogier Crijns, Lettica Hustinx, and Frank van Meurs, I am very grateful for your curiosity for my work, and for the happy moments we shared. Evelyne Vos-Fruit, thank you for opening the door to meeting new French colleagues. Margot van Mulken, I much appreciated our discussions about cultural differences between France and the Netherlands, and your knowledge in this area. Brigitte Planken, thank you very much for your quick help in correcting my English. Finally, Rob le Pair, our shared interest in sports and humour have made us firm friends. If you appreciated my company only half as much as I enjoyed yours, you must have had as good a time as I had.

I am grateful to a number of colleagues outside the department, in particular Bas Andeweg, Wim Blokzijl, Fons Maes, and Jan Renkema for their help with the

data collection, and Judith Kamalski for our extensive discussions. Similarly, I would like to thank Thomas Beaufils, Joop Bindels, Jean-Pierre Brisacier, Chantal Claudel, Geneviève Felten, Christine Guislin, Claudia Huisman, and Philip Scheiner, and the Lequime family for their help with the data collection in France.

Finally, I would like to thank my parents for encouraging me to educate myself. I thank my brother Maarten, my sister Katrien, my family, and my friends for their support. Lisette, my love, you are mentioned last, but that is meant as a compliment. Thank you so much for being there when I needed it the most.

Nijmegen, November 2005

Contents

Introduction

In this chapter, I will argue that the study of cultural differences in the relative persuasiveness of evidence types is important to increase our understanding of argument quality in persuasive effects research. I will first introduce argument quality and the role of evidence (1.1). Subsequently, I will explain why culture may affect the relative persuasiveness of evidence types, which results in the formulation of the research questions (1.2). This chapter ends with an overview of the present study (1.3).

1.1 Argument quality

There is a long tradition in the study of persuasion. The ancient Greek philosophers were probably the first to intensively reflect on the persuasion process. Aristotle's work in particular has had a major influence on our thinking about how messages should be designed in order to successfully persuade people. The Greek approach to studying persuasion was theoretical, and rather intuitive. However sound most of the Greeks' reflections may appear, they were not submitted to empirical tests until the 20th century. From then on, however, scholars have conducted persuasion studies in a systematic and empirical way. The approach to studying persuasion has therefore changed dramatically (see McGuire, 2000; Perloff, 2003), but the main objective of such studies has remained the same since the Greeks: gaining insight into the persuasion process.

Aristotle distinguished between three instruments of persuasion: ethos, pathos, and logos (Aristotle, trans. 1932). Persuaders using ethos try to create a positive, competent image of themselves through their words and gestures. The strategy of pathos relies on the arousal of emotions of the persuadees. Finally, the instrument of logos implies setting up a line of reasoning with constellations of claims and supporting arguments. Current persuasion studies are still interested in the ways the use of ethos, pathos, and logos affects persuasive outcomes (see

Hoeken, 2001c). Fishbein and Ajzen (1981, p. 359) underscored the importance of studying logos, and especially the persuasive effectiveness of arguments. They stressed that the lack of research attention to the persuasiveness of arguments is "probably the most serious problem in communication and persuasion research". Recently, the social psychologist McGuire (2000) compared current persuasive effects research with the ancient Greek study of persuasion. He remarked that current researchers have neglected a few important topics in persuasion that were already studied by the Greeks. One of these topics is the relative persuasiveness of different types of argument.

Most of the research that has been done on the effectiveness of arguments was carried out within the framework of the Elaboration Likelihood Model (ELM), one of the most influential persuasion models of the last decades (Petty & Cacioppo, 1986)[1]. People in principle want to hold correct beliefs and attitudes. If they are confronted with a persuasive message, they will therefore want to assess the correctness of the claim in such a message. The ELM's central assumption is that the extent to which people are willing to base their beliefs and attitudes on a critical assessment of the arguments in the message is variable. The likelihood that people will elaborate a message depends on individual and situational factors. People who are highly motivated and able to scrutinize a persuasive message are predicted to follow the central route to persuasion. They will examine the arguments in the message. Strong arguments will persuade them, and weak arguments will not. If people lack motivation and/or the ability to scrutinize a message, they are predicted to take the peripheral route, that is, they will rely on a more superficial way of processing the information. They will use heuristics, or rules of thumb, such as 'there are a lot of arguments, so it must be true'. Heuristics can only be used, if cues are present in the persuasive message, such as a number of arguments. Strong cues (a large number of arguments) will lead to more persuasion than weak cues (a small number of arguments). The quality of the arguments is not important in the peripheral route: strong arguments are as persuasive as weak arguments.

Although the ELM has stimulated studying the conditions under which argument quality comes into play, it has failed to answer the question as to what really constitutes a strong argument. What did the strong and weak arguments look like that were used in ELM experiments? Consider examples (1), a strong argument, and (2), a weak argument, used in the typical persuasive message in ELM studies about the implementation of a senior comprehensive exam:

(1) "Data from the University of Virginia, where comprehensive exams were recently instituted, indicate that the average starting salary of graduates increased over $4000 over the two-year period in which the exams were begun. At comparable universities without comprehensive exams, salaries increased only $850 over the same period [...]." (Petty & Cacioppo, 1986, pp. 56-57)

(2) "A member of the Board of Curators has stated publicly that his brother had to take a comprehensive exam while in college and now he is manager of a large restaurant. [...] He also indicated that the university has received several letters from parents in support of the exam. In fact, 4 of the 6 parents who wrote in thought that the exams were an excellent idea [...]." (Petty & Cacioppo, 1986, p. 58)

What distinguishes the quality of these two arguments? Is it the difference between the University of Virginia and the member of the Board of Curators? Is it the discrepancy between a dataset versus one restaurant manager? Or is an average starting salary more relevant than the approval of parents? As these arguments differ in various ways, it is difficult to determine what characterizes a strong or a weak argument. Recently, the problematic status of argument quality in the ELM has repeatedly been emphasized (e.g., Areni, 2002; O'Keefe, 2002; Van Enschot-Van Dijk, Hustinx, & Hoeken, 2003). O'Keefe (1995, 2002) has regularly discussed the shortcomings of Petty and Cacioppo's (1986) approach to argument quality in the ELM. His starting point is the empirical definition of argument quality. In the ELM approach, participants were given a series of arguments that they had to rate for persuasiveness. Other participants were then asked to think about such arguments, and to record their thoughts during the processing of the arguments. These arguments were then labeled as strong or weak:

> "We define a 'strong message' as one containing arguments such that when subjects are *instructed* to think about the message, the thoughts that they generate are predominantly favorable. [...] the arguments in a weak message are such that when subjects are instructed to think about them, the thoughts that they generate are predominantly unfavorable" (Petty & Cacioppo, 1986, p. 32)

In a subsequent experiment, other participants were given a persuasive message that included these strong or weak arguments. They were asked to evaluate the message and to record their thoughts. Typically, strong arguments were found to be more persuasive, and to elicit more favorable thoughts than the weak

arguments. However, strong arguments had already been manipulated as being more persuasive than weak arguments. Therefore, O'Keefe (2002, p. 155) argued that a circular reasoning is in play in this approach to argument quality:

> "Thus to say, 'Under conditions of high elaboration, strong arguments have been found to be more effective than weak arguments' is rather like saying, 'Bachelors have been found to be unmarried'. Researchers did not need empirical research to find these things out"

Unfortunately, the empirical approach to a definition of strong and weak arguments gives no answer to the question of what specific qualities an argument must have to be persuasive.

There is a need for deeper insight into argument quality. With the current ELM understanding of argument quality, "it is not yet possible to provide much direction to persuaders about just how to compose effective messages under conditions of high elaboration likelihood" (O'Keefe, 2002, p. 147). In order to be able to distinguish strong and weak arguments, it is important to have a clearer notion of argument itself. In the ELM, arguments were defined as "bits of information [...] that are relevant to a person's subjective determination of the true merits of an advocated position" (Petty & Cacioppo, 1986, p. 16). This means that 'argument' was defined very vaguely. Booth-Butterfield and Welbourne (2002, p. 167) correctly remark that the ELM fails to "provide significant conceptual understanding of a most basic construct – argument"[2].

A solution to the problem of an empirical approach to argument quality is suggested in argumentation theory (O'Keefe, 1995, 2002; cf. Allen, et al., 2000). Argumentation theorists have a normative approach towards the quality of arguments. They address the question as to what characteristics should make an argument strong. Also, they make a distinction between different types of argumentation, of which each has its own normative criteria for good argumentation. These criteria are formulated as evaluation questions that can be asked to assess the quality of the arguments. An argument is strong, if it meets normative criteria. Under the condition in which people carefully scrutinize the arguments in a persuasive message – ELM's central route – norms for acceptability should matter: normatively strong arguments should be persuasive, and normatively weak arguments should not be persuasive. A normative approach to argument quality hence provides a framework for studying the persuasive effectiveness of arguments that are normatively strong or weak.

The most common type of argumentation that is employed in persuasive communication, such as advertising (Schellens & Verhoeven, 1994), and public information campaigns (Schellens & De Jong, 2004), is pragmatic argumentation, or argumentation from consequences. It was also used in the persuasive messages of ELM studies (see, e.g., Petty & Cacioppo, 1986). The prototypical example of pragmatic argumentation is the recommendation of an action on the basis of its positive consequences. In an advertisement, for example, the use of specific fitness equipment may simply be recommended by referring to the consequence of better stamina. Two important evaluation questions for good pragmatic argumentation are (Feteris, 2002; Schellens, 1985): (a) how probable is it that the action will lead to the consequence, in this case, that using this equipment will certainly lead to better stamina, and (b) how desirable is the consequence, in this case, the desirability of better stamina? The text in the ad about this fitness equipment is unclear about the probability and the desirability of better stamina. Ads like these may not be persuasive, because "consumers may neither automatically believe specific product claims to be true nor perceive them as desirable" (Munch, Boller, & Swasy, 1993, p. 294).

In order to make the ad more persuasive, the probability and desirability have to be made evident. This can be done by providing evidence, "data (facts or opinions) presented as proof for an assertion" (Reynolds & Reynolds, 2002, p. 429). Text writers may use different types of evidence. Hoeken and Hustinx (2003b) distinguish anecdotal, statistical, causal, and expert evidence. Anecdotal evidence consists of one case, such as 'Since Andrew Clark regularly uses this fitness equipment, his stamina has improved'. Statistical evidence consists of numerical information about a large number of cases, such as 'A study among 426 people shows that their stamina has improved after having regularly used this equipment'. Causal evidence consists of an explanation, for instance that 'The wireless heart rate monitor in the equipment enables the users to organize their workout, so that their stamina will improve effectively'. Finally, expert evidence consists of a confirmation by an expert, such as 'Professor Delacroix, a specialist in sports medicine, underscores that the regular use of this fitness equipment will improve the users' stamina'.

From a normative point of view, an argument is strong if the evaluation questions related to it can be answered positively. A strong pragmatic argument, therefore, is an argument that makes it highly probable that the consequences will occur, and that makes the consequences highly (un)desirable (cf. Hoeken, 1997). The stronger the evidence in support of the probability and the desirability is, the

stronger the pragmatic argument will be. Different types of evidence are not equally persuasive (see, e.g., reviews of Baesler & Burgoon, 1994; Reinard, 1988). In order to increase our understanding of argument quality, the study of the relative persuasiveness of evidence types is therefore important. A number of studies have investigated the relative persuasiveness of evidence types, but there is still much to be learned about which type of evidence is the most persuasive (O'Keefe, 2002; Reynolds & Reynolds, 2002). O'Keefe suggests that the mechanisms that explain or affect research findings on the effects of evidence should be identified. Some evidence studies have already started to investigate different aspects, such as involvement (Slater & Rouner, 1996), and the vividness of the evidence (Baesler & Burgoon, 1994). However, another aspect has remained untouched: the cultural background of the receiver of evidence.

The question about the possible effect of culture on the relative persuasiveness of evidence types was asked in the first review of studies on evidence effects: "Does evidence function the same way in various cultures?" (McCroskey, 1969, p. 176). Research on this aspect has been called for in studies on evidence (Allen & Preiss, 1997; Bradac, Sandell, & Wenner, 1979; Greene & Brinn, 2003; Reynolds & Reynolds, 2002), and argumentation (Hollihan & Baaske, 1998; Sanders, Gass & Wiseman, 1991), but no article has ever been published about the role of culture on the relative persuasiveness of evidence types. In Reinard's (1988) review of studies on the relative persuasiveness of evidence types, there is only one unpublished study that investigated reactions of people from different cultures to types of source evidence in propaganda messages. Recently, Parrott, Silk, Dorgan, Condit, and Harris (2005) examined the effect of culture on the perceived persuasiveness of evidence, but they compared two forms of one type of evidence (statistical evidence with a visual or a verbal representation). These types of evidence were perceived as equally persuasive by African Americans and European Americans.

1.2 The role of culture

From a normative point of view, the persuasiveness of evidence should depend on the degree to which evidence respects the normative criteria for the support of a claim. If people carefully evaluate the persuasiveness of statistical evidence, for instance, argument theory supposes that the norm of a large sample size comes into play. The sample size in the statistical evidence, such as 426 people who used the fitness equipment, has to be sufficiently large in order to generalize to the claim

about the whole population. Processes, such as causal attribution, induction, and prediction, describe the ways in which people deal with these norms. These psychological processes have been investigated from a cultural perspective (e.g., Fiske, Kitayama, Markus, & Nisbett, 1998). Such cultural studies show that processes may differ between people from various cultures. A recent framework for cultural differences and similarities in reasoning is that of analytic versus holistic thought (Nisbett, 2003; Nisbett, Peng, Choi, & Norenzayan, 2001).

Nisbett et al. (2001) focus on cultures influenced by ancient Greece, and cultures influenced by the ancient Chinese culture, as these are known to differ in their ways of reasoning. Cultures influenced by ancient Greece (e.g., Europe, United States) are characterized by analytic thought, whereas cultures affected by the ancient Chinese culture (e.g., China, Japan) are characterized by holistic thought. Analytic thought is a rule-based system that has "a tendency to focus on attributes of the object to assign it to categories, and a preference for using rules about the categories to explain and predict the object's behavior" (p. 293). Holistic thought is a system "involving an orientation to the context or field as a whole, including attention to relationships between a focal object and the field, and a preference for explaining and predicting events on the basis of such relationships" (p. 293). Nisbett et al. discuss studies that show that this ancient distinction in reasoning between the West and the East still exists. One of these studies is Norenzayan, Smith, Kim, and Nisbett (2002, study III), in which the evaluation of deductive arguments was investigated in two cultures. Deductive arguments reason from general to particular. Two types of deductive arguments were distinguished: atypical and typical arguments. In the typical arguments, such as (3), the category in the conclusion is typical for the overall category in the premise. In the atypical arguments, such as (4), the category in the conclusion is not typical for the overall category in the premise.

(3)	All Europeans are nice people	[premise]
	Therefore, *Germans* are nice people	[conclusion]
(4)	All Europeans are nice people	[premise]
	Therefore, *Polish* are nice people	[conclusion]

From the perspective of analytic thought, the two types of arguments should be equally persuasive. Once the unexpressed premise is detected – namely that Germans or Polish are Europeans – the reader will consider the argument a valid deductive argument. From a holistic point of view, attention is paid to the

similarity relations between the category, here the overall category in the premise, and the exemplar, here the category in the conclusion. The exemplar can be either typical or atypical for the category. With holistic reasoning, the typical exemplars should be more convincing than the atypical exemplars. Norenzayan et al. (2002, study III) gave these two types of arguments to European Americans and Koreans. A manipulation check showed that both cultural groups rated the typical exemplars as more typical than the atypical exemplars. With respect to the hypotheses, Koreans proved to be more persuaded by the typical exemplars than the atypical counterparts, whereas the two types of exemplars were equally persuasive to the Americans. In conclusion, for Eastern participants the typicality of the category in the conclusion is important for the assessment of the persuasiveness of a deductive argument. Studies such as these demonstrate that cultures may differ in basic cognitive processes. If processes related to the evaluation of arguments may be culture-dependent, it is also possible that there are cultural differences in the relative persuasiveness of evidence types.

General research questions

Kline (1971a) argued that a distinction should be made between actual persuasiveness (How persuasive are the types of evidence in reality?), and expected persuasiveness (How persuasive do people expect the types of evidence to be?). A comparison of both types of persuasiveness is important for the understanding of the persuasion process:

> "Such studies of the [actual persuasive] effects of evidence on audiences are useful, but they are not sufficient for an adequate understanding of communication, since an audience responding to a message is only one of the important parts of the process. Among other things, we must also study the encoding part, e.g., the way in which sources select materials for their messages" (Kline, 1971a, p. 190)

Studying expected persuasiveness should not replace that of actual persuasiveness, but it is an useful complement (O'Keefe, 2002). The two general research questions of this dissertation were therefore the following:

General RQ1 Are there cultural differences in the relative *expected* persuasiveness of anecdotal, statistical, causal, and expert evidence?

General RQ2 Are there cultural differences in the relative *actual* persuasiveness of anecdotal, statistical, causal, and expert evidence?

Specific research questions

On the basis of available cultural research on analytic versus holistic thought, there are no straightforward indications whether one type of evidence is more persuasive in one culture than in another. This kind of prediction may be possible, nevertheless, when values are taken as a possible source of cultural differences in the relative persuasiveness of evidence types.

There is growing interest in the impact of culture, and cultural values in the field of persuasion (see, e.g., Aaker & Maheswaran, 1997; Fitch, 2003; Le Pair, Crijns, & Hoeken, 2000), and argumentation (e.g., Aldrich, 2003; Warnick & Manusov, 2000). Fitch remarks that it is "commonly recognized that persuasion is fundamentally shaped by culture" (p. 100). The interest in the effect of culture goes back, argues Fitch, to the foundation of rhetoric with Aristotle. Simply stated, persuasive attempts should be consistent with the values and beliefs that are shared by the audience. If an audience with another cultural background has other values, the persuasive message may have to be adapted to these values. Research has indeed shown that cultures may differ in their value hierarchies, that is, in the relative importance that they attach to basic values such as freedom, pleasure, and creativity. The best-known evidence comes from Hofstede's (1980, 2001) extensive survey among people from 50 countries and three regions. Analyses of the large dataset revealed a pattern of four value dimensions: individualism – collectivism, uncertainty avoidance, power distance, and masculinity – femininity. Each of the countries (cultures) investigated was assigned a score on these dimensions. Cultures were shown to have specific value hierarchies ranging from values that are relatively important to values that are relatively unimportant. In the United States, for instance, individualist values (e.g., independence) are important, whereas in Chinese culture collectivist values (e.g., loyalty) are prioritized.

As different cultures may have different values priorities and beliefs, messages should be adapted to be consistent with the values and beliefs of a specific culture. This ancient Greek suggestion has received some research attention in advertising research. Advertisements often refer to values, such as loyalty or safety. A number of experimental studies have been set up to investigate whether advertisements are more effective if the values that they appeal to are adapted from culture to culture. Le Pair et al. (2000) provide a review of such empirical studies. In one of these studies, Han and Shavitt (1994), an ad with an individualist value appeal (referring to the importance of self) was more persuasive than an ad with a collectivist value appeal (referring to the importance of groups and society) in the United States, whereas the reverse was true for Korea. In five of the six published experiments

that Le Pair et al. reviewed, advertisements were more effective when the values they appealed to were important for a specific culture than when they were not. In the other experiment, the unadapted value appeal was more persuasive than the adapted appeal.

The value dimensions approach appears to be the most popular strategy that is used to characterize and study culture (Fiske, et al., 1998). In fact, values are a useful tool for understanding cultures, because value priorities of individuals can be related to attitudes and behavior (Smith & Schwartz, 1997). This means that, apart from the studies on cultural value adaptation in persuasive messages, culture can be studied by relating cultural values to attitudes and behavior. The importance that cultures attach to certain values may be an indication of the kinds of information that they find important or persuasive. If one is to search for possible links between cultural values, and the four types of evidence, the relationship between power distance and expert evidence is the most obvious (see Jansen, 1999).

The concept of power distance originates from social psychology. According to the power distance theory of Mulder (1977), power is the potential to determine the behavior of other people. Power distance is the degree to which power is distributed equally. A large power distance between two persons means that one person has more potential to determine the behavior of the other person than the other way around. In contrast, a small power distance between two persons implies that both persons have an equal potential to influence the behavior of the other. In his study on cultural values, Hofstede (1980, 2001) showed that cultures may differ in values that refer to power distance. He defines power distance as "the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally" (2001, p. 98). Power is not strictly defined, and may vary from social status to prestige and wealth (Hofstede, 2001). Walton (1997) distinguished between two types of authorities that argumentation may appeal to: an administrative authority, whose power is related to rules, and a cognitive authority, whose power comes from a high level of knowledge. A cognitive authority, or an epistemic authority (Kruglanski, 1989) is a source people may rely on when they acquire knowledge or make decisions. For expert evidence to be persuasive, the receiver will have to accept that the cognitive authority (the expert) possesses more knowledge about the topic in question. Kruglanski et al. (2005) suggested that the influence of experts on people depends on the perceived gap between their own knowledge and that of the expert. As a gap in knowledge may be accepted more easily in large

versus small power distance cultures, it could be argued that expert evidence may be more persuasive in these large power distance cultures.

Indirect evidence for a relationship between power distance and the reliance on experts was given by Pornpitakpan and Francis (2001). They investigated whether participants from a collectivist, large power distance, and strong uncertainty avoidance culture (Thailand) were more influenced by source expertise than participants from an individualist, small power distance, and weak uncertainty avoidance culture (Canada). Participants were given arguments (strong or weak in ELM terms) and an expert cue (high, moderate, or low expertise) in a shampoo ad. Source expertise tended to have more impact on the persuasiveness of the ad for the Thai than for the Canadian participants. The unique contribution of the dimension of power distance to this effect, however, could not be determined.

In order to effectively investigate the effect of culture on the persuasiveness of expert evidence, at least two cultures have to be compared that are as similar as possible on all aspects except the one under investigation (Van de Vijver & Leung, 1997), in this case power distance. This reduces the number of alternative explanations, and enhances the validity of the effect of culture on the persuasiveness of expert evidence. Two cultures that meet this criterion are the Dutch and the French culture. These European cultures differ considerably on the dimension of power distance (Hofstede, 1980, 2001; Koopman, et al., 1999). In general, French people are said to accept and expect large power distances between people with a relatively low power (e.g., lay people), and people with a relatively high power (e.g., experts). France scored 68 on a scale of 0 (smallest power distance) to 100 (largest power distance). In contrast, people in the Netherlands generally have a relatively small power distance: they scored 38 on the same power distance scale. This leads to two specific research questions:

Specific RQ1 Is the *expected* persuasiveness of expert evidence higher for the French than for the Dutch?

Specific RQ2 Is the *actual* persuasiveness of expert evidence higher for the French than for the Dutch?

The main objective of this dissertation was to investigate whether the cultural background of receivers affects the relative persuasiveness of anecdotal, statistical, causal, and expert evidence. This investigation concerned two cultures (the Dutch and the French), and four types of evidence, but focused in particular on one type

of evidence: expert evidence. In this dissertation, I will present five empirical studies. Studies 1 and 2 investigated the role of culture on *expected* persuasiveness, whereas Studies 3 and 4 examined the *actual* persuasiveness of the evidence types in France and the Netherlands. Study 5, finally, was set up for further clarification of the results of Study 4.

1.3 Overview of the study

The studies that will be presented in this dissertation were set up to investigate the role of culture on the relative persuasiveness of evidence types[3]. The starting point of these studies is the normative approach to argument and evidence quality. In Chapter 2, I will further explain the role of evidence within the framework of pragmatic argumentation. I will present the normative criteria for the persuasiveness of the four types of evidence. Chapter 3 is the empirical counterpart of Chapter 2. I will discuss whether what should be persuasive according to the normative criteria is also persuasive in practice by discussing the relevant empirical research on this matter. Such a comparison will also be carried out from a cultural perspective. I will also provide a review of studies that investigated the *actual* relative persuasiveness of evidence types.

The empirical studies that were conducted to answer the research questions presented above will be presented in Chapters 4 – 6. In Chapter 4, I will review the few empirical studies that have been conducted on the *expected* persuasiveness of evidence types. The first of the two empirical studies that I conducted on the expected persuasiveness of evidence types will be presented in this chapter. This study concerned a corpus analysis of professional text writers' use of the four types of evidence in Dutch and French persuasive information brochures.

Chapter 5 will discuss the second study in this domain, an experiment on the expected persuasiveness of evidence types. The presentation of this study will be preceded by a discussion of the experimental approach to Studies 2 – 4, and of the instruments that were used to measure the effect of culture on the relative persuasiveness of evidence types. In Chapter 6, I will present two studies on the relative *actual* persuasiveness of evidence types. Study 3 compared the persuasiveness of anecdotal, statistical, causal and expert evidence evidence in the two cultures. Where possible, these evidence types were normatively strong. Study 4 compared the persuasiveness of normatively strong and normatively weak statistical and expert evidence. The chapter will end with a supplementary study

conducted to throw light on a finding from Study 4. Finally, in Chapter 7, I will present my conclusion on cultural differences in the relative expected and actual persuasiveness of evidence types in the Dutch and the French culture.

Notes

1. Another dual-process model, Chaiken's (1987) Heuristic-Systematic Model, resembles the ELM a lot (see for an overview of differences and similarities, Kruglanski & Thompson, 1999a). I will concentrate on the ELM in this dissertation, because it is the most widely used model of the two.
2. A solution to the problem of what is an argument may also come from the Unimodel, a new view on the modeling of the persuasion process introduced in reaction to the vagueness of the concept of argument in the ELM (Kruglanski & Thompson, 1999a, 1999b). The underlying starting point of the Unimodel is a reflection on the question of which information serves as an argument. Persuasion is related to the process of judgment formation during which beliefs are formed on the basis of evidence. According to the Unimodel, "cues and message arguments all constitute forms [...] of persuasive evidence" (Kruglanski & Thompson, 1999a, p. 92). The Unimodel does not distinguish arguments and cues, but regards both types of information as an argument, or more precisely, as evidence for the message conclusion or claim. Although the Unimodel does not offer a framework to study evidence, it "reintroduces the concept of evidence as a central feature of persuasion" (Kruglanski & Thompson, 1999b, p. 192). The model is a further encouragement to study the persuasive effects of evidence, because "contemporary persuasion models neglected the role of evidence" (Kruglanski & Thompson, 1999b, p. 192).
3. Some of these studies and parts of this dissertation have been published elsewhere: parts of Chapter 3 in Hornikx (in press-a), parts of Chapter 4 in Hornikx, Starren, and Hoeken (2003), parts of Chapter 5 in Hornikx (in press-b), Study 1 in Hornikx (2003, 2004), Study 2 in Hornikx (2005), and Study 3 in Hornikx and Hoeken (2005).

A normative approach to the persuasiveness of evidence

Evidence has been investigated in different fields of study in the humanities and the social sciences, such as argumentation and debate (Klopf & McCroskey, 1969), decision making (Bradac, Sandell, & Wenner, 1979), cognitive psychology (Kahneman, Slovic, & Tversky, 1982), mass communication (Brosius & Bathelt, 1994), advertising (Dickson, 1982), educational communication (Koballa, 1986), health communication (Slater & Rouner, 1996), and policy communication (Hoeken, 2001b). Given this widespread and varied interest in evidence, it is important to specify the approach to evidence in this dissertation. As I mentioned in Chapter 1, a study of the persuasive effects of evidence is highly relevant within the framework of argument quality in the Elaboration Likelihood Model (ELM). In this ELM, argument quality has been approached empirically, which hampers our understanding of what characterizes strong and weak arguments. A normative approach is a solution to this problem. The argumentation used in persuasive messages is usually pragmatic argumentation. The normative criteria for strong pragmatic argumentation will be presented in Section 2.1. Evidence plays an important role here, because the normative criteria can be met by including strong evidence. I will introduce different types of evidence (2.2), and I will present normative criteria for strong evidence that constitute a normatively strong pragmatic argument (2.3). This chapter will end with a discussion of how the normative approach to argument and evidence quality relates to a cultural approach (2.4).

2.1 Pragmatic argumentation

Persuasive messages, which aim to influence people's attitude towards an object or behavior, often use argumentation to persuade them. These messages usually contain pragmatic argumentation. This is the case for advertising (Schellens &

Verhoeven, 1994), public information (Schellens & De Jong, 2004), and policy communication (Schellens & Lagerwerf, 2003). The persuasive messages in ELM studies also employed pragmatic argumentation to convince the participants of the implementation of a comprehensive exam (see Petty & Cacioppo, 1986). In pragmatic argumentation, an action is recommended on the basis of its favorable consequences (positive variant) or advised against on the basis of its unfavorable consequences (negative variant) (Feteris, 2002; Schellens, 1985). The simplest form of the positive variant of pragmatic argumentation looks like (1), and the negative variant like (2):

(1) action A leads to B
B is desirable
therefore: action A is desirable

(2) action A leads to B
B is not desirable
therefore: action A is not desirable

The implementation of a comprehensive exam, for instance, may be recommended by referring to higher starting salaries for students. As higher salaries are desirable, so is the comprehensive exam. Schellens (1985) demonstrated that there are different kinds of pragmatic argumentation, but that these can all be reduced to the simple form of (1) or (2). Another commonality of pragmatic argumentation is the often implicit rule "From the given number of alternative actions, the action that leads to the highest benefits at the lowest cost should be preferred" (p. 154). There might, for instance, be alternative behavior that guarantees higher starting salaries, such as participations in courses that prepare students for the labor market. The more available alternative actions there are, and the more negative and positive consequences relate to these actions, the more complicated pragmatic argumentation will be. The implementation of the comprehensive exam might, for instance, also have a negative consequence, namely a considerable investment. Case (3) is an example of a more complex form of pragmatic argumentation, of which (4) is a concrete example.

(3) action A leads to B
A costs X
B is desirable
B counterbalances X
therefore: action A is desirable

(4) the implementation of a comprehensive exam leads to higher starting salaries
the implementation is an expensive operation
higher starting salaries are desirable
higher starting salaries counterbalance the high costs
therefore: the implementation of a comprehensive exam is desirable

When argumentation theorists evaluate forms of argumentation, they use a series of normative, critical questions. On the basis of these evaluation questions, the quality of an argument is assessed. If an argument meets the criteria, it is considered as a strong argument. Schellens (1985) lists six evaluation questions for pragmatic argumentation:

1. Is B desirable?
2. Does action A lead to B?
3. Is action A feasible?
4. Is action A acceptable?
5. Do B and the possible other positive consequences of A counterbalance the disadvantages of A?
6. Is A the action that leads to B at the lowest cost.

Of these questions, the first two appear essential (Feteris, 2002; Van Eemeren & Grootendorst, 1992). The first one is about the desirability of the consequence: are higher starting salaries desirable? The second one is about the probability of the consequence: does the implementation of a comprehensive exam really lead to higher starting salaries? A pragmatic argument is normatively strong if these critical questions can be answered affirmatively. A strong argument in pragmatic argumentation, therefore, is an argument that makes it highly probable that the consequence will occur, and that indicates that the consequence is highly (un)desirable (Hoeken, 1997). In everyday argumentative discourse, pragmatic argumentation is usually employed in its simplest form, such as (1) or (2). Such simplified forms of argumentation may not supply sufficient information to answer the questions about the desirability and the probability of the consequence (cf. Munch, Boller, & Swasy, 1993). In order to make pragmatic argumentation normatively stronger, evidence can be employed as support for the probability and the desirability. Evidence is defined as "data (facts or opinions) presented as proof for an assertion" (Reynolds & Reynolds, 2002, p. 429). Evidence has been distinguished from other types of data that can be used to support claims. According to Klopf and McCroskey (1969), there are three distinct types of data: first-order data (audience opinion or audience knowledge), second-order data

(opinions or information asserted by the person that puts forward the claim), and third-order data. Third-order data are the most frequently used data and are called evidence.

From a normative point of view, the probability and the desirability of the consequence are equally important to a strong argument. In practice, the use of evidence appears to be less necessary for the desirability than for the probability of the consequence. Schellens and De Jong (2004) suggested that evidence for the desirability of consequences, such as a higher salary or good health, may be redundant if these consequences are in the people's own interest. People are often capable themselves of judging the desirability of consequences. This suggestion was supported in a corpus study of persuasive information brochures (Hornikx, Starren, & Hoeken, 2003). Evidence was given over three times as often in support for the probability than for the desirability of consequences. This does not mean, however, that providing support for the desirability is always redundant. Support appears necessary when the desirability of the consequences is not self-evident, for instance with consequences that are related to others. For example, people may not immediately grasp the desirability of better public transport in the Western region of Nigeria. In a fundraising brochure about better public transport in this region, evidence in support for this desirability could be necessary to convince the reader to donate to a specific fund. In the corpus study of Hornikx et al., evidence in support for the desirability of the consequences was indeed more frequent in brochures with an interest for other people than in brochures with a direct personal interest.

With this normative framework of pragmatic argumentation, it is possible to gain a better understanding of argument quality. In the ELM studies, the implementation of a senior comprehensive exam was recommended on the basis of several arguments. From the perspective of pragmatic argumentation, each argument consisted of a consequence of the exam's implementation, such as a higher starting salary. If the message had consisted of just this single consequence, the whole line of argumentation would have come down to (5) or, in a shorter formulation, (6):

(5) the implementation of a comprehensive exam leads to higher starting salaries
higher starting salaries are desirable
therefore: the implementation of a comprehensive exam is desirable

(6) The implementation of a comprehensive exam is a good thing, because it will allow students to have a higher starting salary.

The strong and weak arguments in the ELM, however, were more elaborate than (5) and (6). Example (7), mentioned in Chapter 1, was a strong argument:

(7) "Data from the University of Virginia, where comprehensive exams were recently instituted, indicate that the average starting salary of graduates increased over $4000 over the two-year period in which the exams were begun. At comparable universities without comprehensive exams, salaries increased only $850 over the same period [...]." (Petty & Cacioppo, 1986, pp. 56-57)

In my own reformulations, this argument consists of a desirable consequence of the exam (7a), and evidence that the implementation of the exam will indeed lead to this consequence (7b)[1].

(7a) Students who have passed a comprehensive exam will have a high average starting salary.

(7b) A study has shown that the starting salaries of graduates at a university with comprehensive exams are much higher than these of graduates of comparable universities without comprehensive exams.

An argument in ELM studies can therefore be defined as a desirable or undesirable consequence of a proposed action (7a), supported by evidence about the probability that the action will lead to the consequence (7b). Note that the desirability of the consequence was not supported by evidence in the typical ELM argument. If evidence is used in pragmatic argumentation, evidence for the probability of the consequences appears to be the most important.

After having described the function of evidence in pragmatic argumentation, it is important to take a look at the characteristics of the claims that evidence supports. In fact, there are different kinds of claims: evaluative claims, normative claims, and descriptive or factual claims (Rieke & Sillars, 1984; Schellens & Verhoeven, 1994). An example of an evaluative claim is 'The implementation of a comprehensive exam is bad'. An illustration of a normative claim is 'The implementation of a comprehensive exam should be forbidden'. Descriptive claims are the kinds of claims that evidence supports in pragmatic argumentation, because these claims describe the occurrence of effects as a result of an action. However, since they describe a causal relation, these claims can also be regarded as causal claims. An example of a descriptive, causal claim is (8):

(8) The implementation of a comprehensive exam leads to higher starting salaries.

In addition, claims such as (8) are general: the phenomena are described independently of a context, or of a reference to time and space. In contrast, singular or specific claims deal with particular instances that occur at specific points in time and/or place, such as in example (9).

(9) The implementation of a comprehensive exam at the University of Rotterdam will lead to higher starting salaries.

The distinction between general and specific claims is often made in related fields of study, such as judgment and decision making (e.g., Anderson & Kellam, 1992), and causal cognition (e.g., Oestermeier & Hesse, 2001). This distinction is important, because it affects the set of normative criteria for strong evidence that supports these claims (see Section 2.3). In sum, the claims in the present study of the persuasive effects of evidence are descriptive, causal, and general. The evidence that can be employed in support of these claims has various forms. In the next section, I will present different types of evidence.

2.2 Four types of evidence

Different types of evidence can be used to support descriptive, causal, and general claims in pragmatic argumentation. In Section 2.2.1, I will give a historical overview of the distinction between anecdotal, statistical, and expert evidence. Anecdotal evidence consists of a specific instance (Rieke & Sillars, 1984, p. 92), statistical evidence is a numerical summary of a series of instances (p. 94), and expert evidence consists of the testimony of an expert (p. 94). In Section 2.2.2, I will discuss causal evidence, and its difference with the three other types of evidence.

2.2.1 Historical overview

There are a lot of different typologies of evidence types. Reynolds and Reynolds (2002, p. 437) are correct to posit that "Argumentation and persuasion scholars would be well-served by extended efforts at conceptualizing and testing different classifications of evidence types". In order to give a rationale for the distinction between anecdotal, statistical, and expert evidence, I have traced the origins of the classifications of evidence types used in persuasion studies to debate textbooks.

Table 2.1 provides an overview of the three types of evidence in these two domains. I have used an old and a relatively new debate textbook (Klopf & McCroskey, 1969; Warnick & Inch, 1989), and reviews on the relative persuasiveness of evidence types that compared at least two of these evidence types (Allen & Preiss, 1997; Baesler & Burgoon, 1994; Reinard, 1988).

Table 2.1 Anecdotal, statistical, and expert evidence in debate textbooks and persuasion studies

	anecdotal	statistical	expert
debate	[1]factual examples	statistics	opinions, testimony
	[2]reports, descriptions	statistics	opinions
persuasion	[3]reports of events, examples	statistics	testimonial assertions
	[4]report	statistical	-
	[5]narrative	statistical	expert testimonial assertions

[1]Klopf & McCroskey (1969), [2]Warnick & Inch (1989), [3]Reinard (1988), [4]Baesler & Burgoon (1994), [5]Allen & Preiss (1997)

Debate textbooks, which have become popular in the United States since the end of the 19th century, were aimed at educating students for a legal career or simply for participation in debates (see, e.g., Van Eemeren, Grootendorst, & Snoeck Henkemans, 1997). In these textbooks, evidence is regarded as "facts, opinions, and objects that are used to generate proof" (Freeley, 1976, p. 74)[2]. Klopf and McCroskey (1969) distinguished four types of evidence, of which the first one – tangible objects, such as murder weapons – is rarely used in argumentative texts. The other types of evidence are opinions or testimony, factual examples, and statistics. Two decades later, Warnick and Inch (1989) gave a similar division of evidence types. They distinguished opinion evidence, reports and descriptions, and statistics. Debate textbooks thus suggested that testimonial and factual (statistics or examples) evidence can be used in written communication.

Experimental studies conducted to examine the relative persuasiveness of evidence types have generally concentrated on similar types of evidence. Initially expert evidence alone (McCroskey, 1969) was investigated. Later, the persuasiveness of expert, statistical and anecdotal evidence (Reinard, 1988) was studied. Most recently, the focus was on anecdotal and statistical evidence (Allen & Preiss, 1997; Baesler & Burgoon, 1994). Anecdotal and statistical evidence have been studied most extensively in persuasion studies. Concerning expert evidence, it should be noted that it was usually called source evidence. It appears legitimate

to label source evidence as expert evidence, because in the studies on source evidence reviewed by McCroskey sources in the experimental conditions differed in their perceived authoritativeness[3].

2.2.2 The case of causal evidence

The study of the persuasiveness of causal evidence is relatively new. In argumentation and debate, there appears to be a distinction between causal evidence, and other types of evidence. For Klopf and McCroskey (1969), for instance, evidence is third-order data, whereas causal information is considered as second-order data that the debater gives himself. Also, Rieke and Sillars (1984) actually do not speak of 'causal evidence' in their book on argumentation, but present it as the argument by cause in another chapter than anecdotal, statistical, and expert evidence. The distinction between causal and the other types of evidence can also be expressed in terms of deductive and inductive reasoning (see, e.g., Van Eemeren, et al., 1997). Causal evidence in support of a causal claim is characterized by deduction: if the causal relationship in causal evidence is true, then the claim it supports is *true*. The three other types of evidence are characterized by induction: if the evidence is true, then the claim is *probable*.

This distinction between deductive and inductive reasoning leads to the discussion of the fundamental difference between causal evidence and other evidence in the field of reasoning and thinking. Kuhn (1991) argues that there are two types of evidence that prove a claim such as 'The implementation of a comprehensive exam leads to higher starting salaries'. The first type is genuine evidence, which establishes an empirical relationship between an antecedent and an outcome. Such evidence is "distinguishable from description of the causal sequence itself" (p. 45). Anecdotal and statistical evidence can be placed in this category. In contrast, causal evidence is called pseudoevidence, which "cannot be sharply distinguished from description of the causal sequence itself" (pp. 65-66). Pseudoevidence provides a causal theory, that is, the mechanism by which a cause produces an effect. Kuhn therefore concludes that "pseudoevidence must be regarded not as evidence at all, but as a part of the theory" (p. 95). Note that expert evidence seems to take a position between genuine and pseudoevidence. On the basis of Kuhn, it may be considered pseudoevidence, but on the basis of the distinction between deductive and inductive reasoning, it may be grouped with anecdotal and statistical evidence in the category of inductive reasoning. In this dissertation, expert evidence is considered genuine evidence, since experts are expected to provide genuine evidence when they are asked to.

If causal evidence differs fundamentally from anecdotal, statistical, and expert evidence, why should its persuasive effect be compared to that of genuine evidence? A straightforward argument is that pseudoevidence and genuine evidence can both be used as proof. Brem and Rips (2000, p. 574) argue that both "explanations and evidence can enhance an argument by encouraging arguers to accept or reject argument claims". In fact, causal evidence appears to be an important source of information that people use to evaluate and support the probability of claims. First, concerning the *evaluation* of claims, a number of studies have shown that people make use of cause-effects relationships in order to coherently interpret and assess the likelihood of events (Ajzen, 1977; Tversky & Kahneman, 1980). Ajzen specifically introduced the causality heuristic to stress this importance: "It is proposed that especially in the case of judgments concerning human behavior and its effects, people often rely on their intuitive understanding of the factors that cause the event in question" (pp. 303-304).

Second, pseudoevidence is frequently employed to *support* claims. Kuhn (1991) had her participants generate causal theories for the occurrence of a phenomenon, such as unemployment. Subsequently, she asked these participants to provide evidence in support of their causal theory. All participants provided pseudoevidence for their own causes, but only 40% of them complemented it with genuine evidence. Brem and Rips (2000) demonstrated that this apparent preference for pseudoevidence as proof for causal claims is influenced by the availability of evidence. In a series of four studies, they showed that genuine evidence is used if it is available, and that people switch to explanations (pseudoevidence) if genuine evidence is unavailable. People also proved to be aware of their substitution of evidence for explanations when evidence is unavailable. Still, both Kuhn, and Brem and Rips demonstrated that pseudoevidence is an important source that people use to support causal claims. Pseudoevidence provides a mechanism for the cause-and-effect relation in the claim. Such mechanism information was found to be the most preferable causal information in paper-and-pencil experiments (Ahn, Kalish, Medin, & Gelman, 1995), and to be the most frequently used causal information in causal arguments in journal articles (Oestermeier & Hesse, 2000).

2.3 Normative criteria

From a normative point of view, a strong argument in pragmatic argumentation is an argument that makes it highly probable that a consequence will occur, and that indicates that this consequence is highly (un)desirable. When evidence is employed to support the probability, a line of argumentation is built, because evidence serves as an argument in support of a claim. Each line of argumentation can be characterized by the way in which the argument supports the claim. This characterization is called the argumentation scheme, "a more or less conventionalized way of representing the relation between what is stated in the argument and what is stated in the standpoint" (Van Eemeren & Grootendorst, 1992, p. 96). One of these schemes was already presented in Section 2.1, namely the scheme of pragmatic argumentation. Each type of evidence that is used as an argument for a claim is related to an argumentation scheme. Surprisingly, evidence types and argumentation schemes have been treated separately in argumentation textbooks (e.g., Klopf & McCroskey, 1969; Warnick & Inch, 1989). The argumentation scheme, however, determines which critical questions have to be asked in order to determine the quality of evidence. Some textbooks did include criteria for evidence (e.g., Rieke & Sillars, 1984), but these criteria apparently were not based on any theoretical account. Such a basis is provided here. I will first present different argumentation schemes, before I will indicate what critical questions are relevant to the different types of evidence.

Throughout the years, a lot of different classifications of argumentation schemes have been proposed. As a consequence, there are differences in the number of schemes, subschemes, and critical questions. It is beyond the scope of this dissertation to discuss these classifications. Instead, the starting point is the argumentation schemes and critical questions that are distinguished and formulated by Garssen (1997) and Schellens (1985), who both provided a review of classifications of argumentation schemes. When necessary, their schemes will be complemented with those of other argumentation theorists.

Which argumentation schemes are related to the four types of evidence? The answer to this question depends on the type of claim. In fact, the argumentation scheme is an expression of the relationship between the argument (the type of evidence) and the claim. A change in the type of claim may alter the argumentation scheme in question. In Section 2.1, specific claims were distinguished from general claims. Evidence can be employed to support the general, descriptive claim 'The implementation of a comprehensive exam leads to

higher starting salaries', and the specific, descriptive claim 'The implementation of a comprehensive exam at the University of Rotterdam will lead to higher starting salaries'. As Table 2.2 demonstrates, the argumentation scheme for statistical and anecdotal evidence is not the same when these evidence types support a general claim than when they support a specific claim[4].

Table 2.2 Argumentation schemes in relation to type of evidence and type of claim

type of evidence	general claim	specific claim
statistical	argument by generalization	argument by classification
anecdotal	argument by generalization	argument by comparison
causal	argument by cause	argument by cause
expert	argument by authority	argument by authority

In support of a general claim, statistical evidence is related to the argument by generalization. In an argument by generalization "you look at a series of instances and from them claim a general principle" (Rieke & Sillars, 1984, p. 72). A statement about the higher starting salaries that result from the implementation of comprehensive exams at a large number of universities may be used as evidence to generalize this result to universities in general. What critical questions can be asked to assess the quality of statistical evidence that is employed in support of a general claim? The critical questions for statistical evidence, as for the other types of evidence, will be presented in the form of normative criteria. I will present normative criteria that (a) are shared by most classifications (that is, they are listed in the overviews of Garssen, 1997 and/or Schellens, 1985), and/or that (b) are relevant in the framework of evidence supporting descriptive, causal claims in written persuasive communication. Appendix A provides a list of these normative criteria. Statistical evidence in support of general claims is normatively strong, if it meets the following two normative criteria: (a) the cases in the evidence should be representative of the class of cases in the claim, (b) the number of cases in the evidence should be sufficiently large in order to allow valid generalizations about the class of cases in the claim. Note that these normative criteria correspond to the rules of sampling theory about representativeness and sample size.

In support of a specific claim, statistical evidence is related to the argument by classification. The scheme of classification is as follows: "what is true of the items reported in the evidence is also true of a hitherto unexamined item under consideration that is known (or believed) to fall within that class" (Ehninger & Brockriede, 1963, p. 145). The same evidence about the higher starting salaries at a

large number of universities is not generalized to universities in general, but to the University of Rotterdam in particular. This scheme of classification implies two subsequent steps: generalization from a sample of cases to a class of cases, and a comparison between the sample and the single case in the claim. The normative criteria that apply here are similar to those that were just discussed (step 1), complemented with a criterion about similarity between sample and case (step 2). In fact, the sample in the evidence may be large and representative of the *population*; this does not necessarily imply that the sample and the *case* are similar. The University of Rotterdam, for instance, may have a specific characteristic that is not shared by a sample of Dutch universities that – as a whole – is representative of all the Dutch universities. Therefore, the similarity criterion has to be included. The normative criteria therefore are: (a) the cases in the evidence should be representative of the class of cases to which the case that is presented in the claim belongs, (b) the number of cases in the evidence should be sufficiently large in order to allow valid generalizations about the class of cases to which the case in the claim belongs, (c) the cases in the evidence should be highly similar to the case presented in the claim.

Next, anecdotal evidence is related to the argument by generalization when a general claim is concerned. The evidence that the implementation of a comprehensive exam at the University of Amsterdam has resulted in higher starting salaries may be used to support the general claim about higher salaries as a consequence of this comprehensive exam. The two normative criteria for the argument by generalization – sample size and representativeness – were presented for statistical evidence, and also apply to anecdotal evidence. By definition anecdotal evidence does not meet the criterion of large sample size. The representativeness criterion, however, can be met: a single case is representative of a group of cases, if they are highly similar. Similarity plays a more important role when anecdotal evidence is used to support a specific claim. Anecdotal evidence is then related to the argument by comparison. The same statement about the University of Amsterdam may be used to support the probability of the benefits of the comprehensive exam at the University of Rotterdam. In an argument by comparison, the argument is "presented as if there were a resemblance, an agreement, a likeness, a parallel, a correspondence or some kind of similarity between that which is stated in the argument and that which is stated in the standpoint" (Van Eemeren & Grootendorst, 1992, p. 97). Anecdotal evidence in support of specific claims is normatively strong, if it meets the following two normative criteria: (a) there should be a large number of relevant similarities

between the case in the evidence and the case in the claim, and (b) there should be no relevant dissimilarities between the case in the evidence and the case in the claim.

When causal evidence supports a general or specific causal claim, the argument by cause is in play. In this argumentation scheme, the argument is "presented as if what is stated in the argumentation is a means to, a way to, an instrument to or some other kind of causative factor for the standpoint, or *vice versa*" (Van Eemeren & Grootendorst, 1992, p. 97). In pragmatic argumentation, causal evidence supports a claim that is causal itself. Causal evidence, therefore, provides a mechanism that explains the underlying causal relation in the claim (see Figure 2.1). With causal evidence, a causal relation is created from the cause in the claim (C1) to the effect in the claim (E1). An example is the explanation that, if comprehensive exams are introduced (C1), students will have more proof for their expertise (E2); this proof (C2) enables them to get higher salaries more quickly (E1).

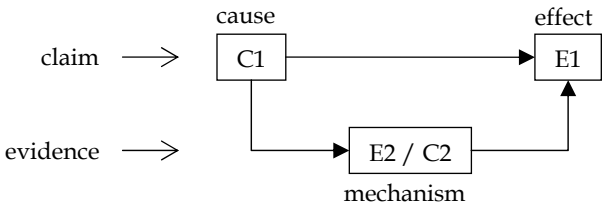

Figure 2.1 Causal evidence as a causal mechanism in support of the claim C1 > E1

The causal mechanism in Figure 2.1 consists of two relations, but the number of relations may be higher. Causal evidence is normatively strong, if it meets three normative criteria for each relationship, such as from C1 to E2, or from C2 to E1. These normative criteria are (see, e.g., the rules for causality in social sciences in Elifson, Runyon, & Haber, 1998): (a) the cause and the effect should covary, (b) the cause should come into play before the effect occurs, and (c) there should not be other variables that trouble the causal relationship in a positive way (the relationship only holds if another variable is present), or in a negative way (the relationship does not hold if another variable is present).

Finally, in an argument by authority "a statement is defended by pointing out the fact that an authoritative person or institution suscribes to it" (Schellens, 1985, p. 179). The simple fact that, for instance, Professor Jackson underscores that the implementation of a comprehensive exam leads to higher starting salaries, may

serve as support for the probability of this claim. Expert evidence can be considered as a form of genuine evidence, because experts are expected to be able to provide empirical facts (cf. Section 2.2.2). Facione and Scherer (1978, p. 315) clearly formulate this point: "Instead of the whole proof the reliable authorities give their word. They vouch for the conclusion's being true, but they could have given the proof. They should be able to supply an acceptable proof upon demand". Walton (1997) provided a detailed discussion of the argument by authority. He showed that there are two different types of authority: the administrative authority and the cognitive authority. An administrative authority has "the right to exercise command over others or to make rulings binding on others through an invested office or recognized position of power" (p. 76). Examples are a minister and a mayor. When a cognitive authority is concerned, there is "a relationship between two individuals where one is an expert in a field of knowledge in such a manner that his pronouncements in the field carry a special weight of presumption for the other individual" (p. 77). Expert evidence is most related to the cognitive authority. Expert evidence as support for general or specific claims is normatively strong, if it meets the following three normative criteria: (a) the expert should be credible, (b) the expert should be reliable, and (c) the expert should have a high expertise in the field that is relevant to the claim.

2.4 Different cultures, different norms?

In Section 2.3, I presented normative criteria that are expected to affect the persuasiveness of the four types of evidence. Normatively strong evidence should be more persuasive than normatively weak evidence if people carefully scrutinize the evidence in a persuasive message. These normative criteria were developed by American and European argumentation theorists, and scholars in the social sciences. There are no research findings to date that demonstrate that norms related to the persuasiveness of evidence types differ or not from culture to culture. Still, if norms should be culture-independent, cultures may react differently to these norms (see Section 3.1.2). Apart from this lack of empirical evidence, philosophers and argumentation theorists have stressed the importance of cultural differences in the evaluation of argument quality (see, e.g., MacIntyre, 1988; McKerrow, 1990). According to the multiculturalism perspective, the evaluation of argument quality depends on culturally specific beliefs and values. This perspective contradicts the normative approach to argument quality, which

argues that an argument is strong if it justifies the claim that it supports. In this view, argument quality is not dependent on the evaluator, but is an intrinsic characteristic of the argument. Siegel (1999) critically analyzed both perspectives, and concluded that the multiculturalist perspective fails against the normative approach. The multiculturalists claim that there are no transcultural, objective principles. Without these principles, however, argues Siegel, they can not criticize other perspectives, including their own normative perspective.

Apparently, there are different opinions as to whether argument quality is an intrinsic characteristic of an argument or whether it depends on receivers' cultural characteristics. In addition, if argument quality is culture-dependent, it is still an open question as to whether the norms are universal and people's reactions to them are culture-dependent, or as to whether norms themselves may be culture-dependent. Empirical research is needed to gain insight into this question. A number of studies have investigated whether norms do matter in practice, but not from a cultural perspective. In the next chapter, I will review such studies on the norms related to the four types of evidence.

Notes

1. Other researchers have also observed the distinction between desirability and probability in the arguments used in the ELM (e.g., Areni & Lutz, 1988; Van Enschot-Van Dijk, Hustinx, & Hoeken, 2003). Hoeken (1997) observed that this distinction is also made in rational decision making. When people adopt the central route to persuasion, their behavior should resemble that of a rational decision maker. In that case, models of rational decision making can be consulted. These models claim that people should base their decisions upon the consequences that result from these decisions (Abelson & Levi, 1985). Such decisions are based on two assessments: the desirability and the probability of the consequences of the decision.
2. It should be noted that 'opinion' has to be regarded as the *fact* that a person other than the debater or text writer has an opinion (e.g., in expert evidence). Such a fact can be observed and verified.
3. After McCroskey's (1969) review of studies on expert evidence, the interest in expert evidence has diminished considerably. The set of studies that Reinard (1998) selected for his review on expert evidence ('testimonial assertion evidence'), for instance, dated from 1955 to 1984. This current lack of interest in expert evidence might be explained by the popularity of the dual-process models as the ELM, in which expert assertions are considered as a cue, and not as an argument. As a consequence, evidence studies might have left expert evidence aside.
4. As each type of evidence is related to an argumentation scheme, one could argue that the study of the persuasiveness of evidence types could also be labeled the study of the persuasiveness of different argumentation schemes. In the framework in which evidence supports the probability of causal claims in pragmatic argumentation, however, type of evidence does not correspond to argumentation scheme for two

reasons. The first reason is that the argument by generalization is related to both statistical and anecdotal evidence (see Table 2.2). There is, therefore, no exact one-to-one relationship between evidence types and argumentation schemes. The second reason is that anecdotal and statistical evidence can each be linked to two distinct argument types depending on the type of claim. More generally, there is a difference in perspective of argumentation theorists, interested in schemes and criteria for strong argumentation, and persuasion scholars, interested in the persuasive *effect* of different types of evidence. More important than a detailed comparison of the two perspectives, however, is that the study of the relative persuasiveness of evidence types can benefit from insights of argumentation theory.

A review of studies on the actual persuasiveness of evidence types

Whereas Chapter 2 addressed the question what characteristics *should* make arguments persuasive, this chapter will answer the question which characteristics *do* make them persuasive. In Section 3.1, I will discuss empirical studies that investigated whether meeting normative criteria actually leads to a more persuasive outcome. I will show that only little attention has been paid in research to the role of culture in this area. Subsequently, I will discuss explanations other than the normative criteria that have been given for the persuasiveness of evidence types (3.2). A number of experiments have been conducted to test whether some types of evidence are more persuasive than others. In the last section of this chapter (3.3), I will present a review of such studies on the relative actual persuasiveness of evidence types.

3.1 Normative criteria and persuasiveness

In the condition of high elaboration, under which people are expected to carefully scrutinize the arguments in a persuasive message, normative criteria for strong evidence should matter. In this section, I will answer the questions as to whether what should be persuasive is also persuasive in practice (3.1.1), and whether cultures differ in this respect (3.1.2).

3.1.1 Do normative criteria for strong evidence matter?

Below, I will review studies that have tested whether meeting the normative criteria affects the persuasiveness of evidence in practice. It should be noted at the outset that only a minority of these studies were studies on evidence. This can be explained by the lack of interest in normative criteria for evidence. Most of the studies were (cognitive) psychological studies that examined whether people's

judgments were affected by information that did or did not meet the normative criteria relevant to the types of evidence. The criteria for strong *causal* evidence will not be discussed, because they have not been investigated.

Sample size

One of the normative criteria for strong statistical evidence in support of a general claim is the presence of a sufficiently large sample size of cases that allow valid generalizations about the population in the claim. There are hardly any empirical studies that have examined this sample size criterion. Doosje, Spears, and Koomen (1995, p. 643) correctly posit that "surprisingly little is known about the processes of generalizing sample information to a target population". I found only two studies that investigated the persuasive impact of different sample sizes. In Darke et al. (1998), participants were given the Petty and Cacioppo (1986) text about the implementation of a comprehensive exam. The text included the general opinion of either 10 or 1 000 students, of which 80% were in favor of the implementation or of which 80% were against. Participants were highly or moderately involved in the topic through a manipulation of the introduction date of the exam (in 1 year or in 10 years). Low-involved participants proved to be influenced by the majority opinion regardless of the sample size. However, sample size mattered for the high-involvement participants: they were only influenced by the majority opinion when the poll size was large.

In Doosje et al. (1995, study II), participants had to generalize behavior of a small or a large sample to their own ingroup or to an outgroup. Sample information was either positive (e.g., 'I gave a tourist directions'), or negative ('I jumped the queue at the baker's'). When the sample information was positive for the participants' ingroup, samples were generalized irrespective of their size. When the sample information was negative for the participants' ingroup, the sample size mattered. In fact, there was a less strong generalization with a small sample than with a large sample. Finally, these two studies suggest that large sample sizes may have more impact than small sample sizes under certain motivational conditions. In Darke et al. (1998), larger sample size only had an effect when the participants were highly involved, and in Doosje et al. the effect of a larger sample size was higher when the likelihood estimation of the negative behavior concerned the participants' ingroup.

It should be noted that a large number of studies have been conducted in cognitive psychology on the way people use sample size information to make decisions. Classic experiments showed that people are insensitive to sample size.

In Kahneman and Tversky (1973), for instance, participants were given the following base-rate information (comparable to statistical evidence): psychologists have interviewed 30 engineers and 70 lawyers. The participants were then given five descriptions of persons (comparable to anecdotal evidence), which were said to be randomly chosen from the 100 interviews. The participants were asked to judge for each description whether the person was an engineer or a lawyer. The results showed that the participants underused statistical information and its sample size information, but that they based their judgment on the individual's characteristics. People apparently did not use the so-called law of large numbers, which reads "the larger a sample is the more likely it is to be representative of the population from which it is drawn" (Kunda & Nisbett, 1986, p. 351).

In studies such as the one by Kahneman and Tversky (1973), statistical and anecdotal evidence were simultaneously presented to participants. In order to test the understanding of the law of large numbers, participants' reactions to sample sizes without competing individuating information should be investigated. A considerable number of studies did this, and demonstrated that people do understand the law of large numbers (e.g., Kunda & Nisbett, 1986; Nisbett, Krantz, Jepson, & Kunda, 1983; Well, Pollatsek, & Boyce, 1990). In Nisbett et al. (study I), for instance, participants had to make a judgment about a population on the basis of samples of 1 person, 3 persons, and 20 persons. Their estimations were consistent with the law of large numbers: "they were more willing to assume that the population resembles the sample when N is larger" (p. 349). On the basis of such studies, Well et al. concluded that people's understanding of the law of large numbers is not likely to result from a thorough knowledge of sample sizes, but from a more simple heuristic 'bigger is better'. Regardless of how the effect of different sample sizes may be explained, the studies discussed in this section provide reasons to believe that evidence with a larger sample size may be more persuasive than evidence with a smaller sample size[1].

Representativeness

The other normative criterion for strong statistical and anecdotal evidence in support of a general claim, is that the case(s) in the evidence should be representative of the population in the claim. Representativeness has to do with homogeneity of the characteristics of cases, and with random selection of a sample. Concerning homogeneity, the more homogeneous a population is with respect to its relevant characteristics, the higher the chances that even a small sample will be representative. A few studies have investigated the way people generalize as a

function of the homogeneity of samples and populations. Nisbett et al. (1983), for instance, demonstrated that people generalized more when the *population* was homogeneous with respect to relevant characteristics than when it was more heterogeneous. Doosje et al. (1995, study I) examined the effect of the degree of homogeneity of *samples* on the ease of generalization to a population. When the sample information was negative for the participants' ingroup, there was a less strong generalization with a heterogeneous sample than with a homogeneous sample. When the sample information was positive for the participants' ingroup, however, samples were generalized irrespective of their degree of homogeneity. Thus, under certain conditions, people understand that the more homogeneous a sample is, the more it can be generalized to a population. This understanding not only holds for reasoning from sample to population, but also from individual to sample. Park and Hastie (1987, study I) showed that participants generalized more easily from an individual to a homogeneous sample than to a heterogeneous sample. Representativeness can not only be assessed through the sample's or population's homogeneity, but also through information that the sample has been randomly selected or not (Kassin, 1979). Random selection implies that the sample should be representative. People have been shown to understand the relevance of random selection. In Wells and Harvey (1977, study I), participants who were told that the sample had been randomly selected made a more correct generalization than participants who were not.

A few studies in cognitive psychology have also investigated the effect of the representativeness of cases on people's generalization of these cases to the population (e.g., Hamill, Wilson, & Nisbett, 1980; Nisbett & Borgida, 1975). Representativeness was operationalized by indicating that cases were typical or atypical of the population. Typicality was found not to affect the generalization. Again, case information was given together with statistical population data in these studies. This design favored the impact of the anecdotal information, and therefore is not an unbiased test of the effect of representativeness. A more detailed discussion of these studies is given in Appendix B. In conclusion, a limited body of empirical research suggests that – under certain conditions – people may respect the normative criterion of representativeness when generalizing from case or sample to population.

Similarity

The criterion of high similarity between the case(s) in the evidence, and the case in the claim is relevant for anecdotal and statistical evidence in support of a specific claim. There should be a large number of relevant similarities, and there should be no relevant dissimilarities. Only two studies have investigated the effect of the degree of similarity on the persuasiveness of anecdotal evidence. Hoeken (2001b) gave his participants an anecdote about a town that was similar or dissimilar to the town in the claim. The dissimilar anecdotal evidence proved to be as persuasive as the similar anecdotal evidence. Hoeken (2005) reports on a more recent experiment by Hoeken and Hustinx that investigated the same similarity criterion, not with one claim with evidence, but with several claims with evidence. Similar anecdotal evidence proved to be more persuasive than dissimilar anecdotal evidence. In another study, people's likelihood estimation in function of similarity of cases was investigated. Read (1983) had his participants predict a specific performance of a target individual who was either similar or dissimilar to another individual who had performed the action. The participants were more likely to predict the performance when the target was similar to the other individual than when it was dissimilar.

A larger number of studies in cognitive psychology have examined similarity, which is essential in judgment and decision making (e.g., Smith & Zárate, 1992; Tversky, 1977). There are a number of different models of similarity, but they agree on the central characteristic of similarity: a shared set of properties. The more properties two cases share, the more similar they are. In a series of experiments, Tversky demonstrated that people perceive a higher similarity between two cases when common features are added or when distinctive features are deleted. In conclusion, a few empirical findings suggest that the normative criterion of similarity is relevant in practice: judgments based on dissimilar cases differ from those based on similar cases.

Credibility and reliability

Two of the three normative criteria for expert evidence are high credibility and high reliability. Source credibility is one the most frequently studied source factors in persuasive effects research, and has been divided into source expertise and source trustworthiness (see, for an overview, O'Keefe, 2002). As source credibility has been operationalized through the source's field of expertise (the third normative criterion), a discussion of studies on credibility would also imply a discussion of field of expertise. It is possible, however, to discuss studies on the

effect of credibility and reliability of sources that have the same field of expertise. The difference between high credibility / reliability and low credibility / reliability can be expressed through a difference in the title of the source: a professor or a doctor versus a researcher or a student. A number of studies have investigated the effect of the source's title on the persuasiveness of these sources. In general, credible and reliable sources (e.g., professors) have been found to be more persuasive than less credible and reliable sources (e.g., students) (see Petty & Cacioppo, 1986). This effect, however, needs to be qualified: it only occurs if participants have a low ability and/or a low motivation to scrutinize a persuasive message (but see Kruglanski & Thompson, 1999a). If participants are able and motivated to scrutinize a persuasive message, there is no effect for credibility and reliability. In conclusion, under conditions of low elaboration, sources who are highly credible and reliable are more persuasive than sources who are credible and reliable to a lesser degree.

Field of expertise

The other criterion for strong expert evidence is that the expert's field of expertise should be relevant to the field of the claim. A number of experimental studies have examined the effect of field of expertise on the source's persuasiveness, and have shown that sources with a relevant field of expertise are more persuasive than sources with an irrelevant field of expertise (e.g., Luchok & McCroskey, 1978; Maddux & Rogers, 1980; Pornpitakpan & Francis, 2001). In Maddux and Rogers, for instance, participants read a text about a source who stated that people should sleep only four hours a night. The source was Dr. James Campbell, an authority on sleep and sleep research, or on music during the Baroque period. The first Campbell was perceived as significantly more expert on the topic of sleep, and as more persuasive than the second Campbell. These kinds of studies suggest that expert evidence is more persuasive to the extent that the field of expertise of the expert corresponds to the domain of the claim.

In sum, the studies discussed above demonstrate that people react differently depending on the degree to which the normative criteria are met. People's judgments are different when the sample size is larger, when the representativeness of the sample or case is higher, when the similarity increases, when sources are more credible and reliable, and when their field of expertise is more relevant. It is important to be cautious about these conclusions, however,

because they are based on a limited body of research, and because the effects sometimes only occurred under certain conditions.

3.1.2 Cultural differences

In Section 2.4, I argued that cultures may differ in the normative criteria they have for the persuasiveness of evidence types. No studies were available to confirm or disconfirm such differences, but – even if the norms are the same – cultures may also vary in the way they deal with these norms. Nisbett et al. (1983) already suggested that culture may affect people's tendency to apply statistical rules to judgments that require these rules. I found only one study that tested a normative criterion from a cultural perspective. Norenzayan, Smith, Kim, and Nisbett (2002, study II) examined the use of rules for similarity judgments in American and East-Asian cultures. American participants were expected to make rule-based similarity judgments, applying the so-called unidimensional rule. According to this rule, an object is considered as similar to a category of objects if they all share a single feature (cf. Section 3.1.1). In a bouquet of seven flowers, all flowers may have the same kinds of leaves, but different colors. Asian participants were expected to make these judgments more intuitively on the basis of family resemblance. In this case, an object is similar to a category if a large number of features are shared by some of the objects, although no single feature characterizes all the objects. In a bouquet of seven flowers, each flower may share a characteristic with four other flowers, but there is no characteristic that all flowers share (except for the fact that they are all flowers). The hypotheses were confirmed: the Americans' reasoning followed the unidimensional rule, and Easterners' reasoning the family resemblance. For anecdotal evidence, this may imply that an anecdote that is dissimilar to a given case for an American participant might be similar to this case for an Asian participant.

In conclusion, relatively little attention has been paid in research to the role of culture on how people deal with norms such as large representativeness, high similarity, or high source credibility. Whether culture affects the way people deal with these norms is an issue that should be addressed in future research.

3.2 Other explanations for the persuasiveness of evidence types

If people carefully scrutinize a persuasive message, normative criteria for strong evidence, such as the criterion of a large sample size, should matter. Under this condition, statistical evidence could be expected to be more persuasive than anecdotal evidence, because the sample size in statistical evidence is larger. In some studies, however, anecdotal evidence was found to be more persuasive than statistical evidence (see reviews of Baesler & Burgoon, 1994; Reinard, 1988). The persuasiveness of evidence therefore can not be explained by normative criteria only. A few studies have provided other explanations why anecdotal, causal, and expert evidence can be persuasive. An overview of such studies is given below. For statistical evidence, I have not found other explanations than the normative criteria.

Anecdotal evidence

Several researchers have tried to explain the persuasive power of anecdotal evidence (see, e.g., Baesler & Burgoon, 1994; Sherman, Beike, & Ryalls, 1999). Anecdotal evidence may be persuasive, (a) because it evokes similar cases or (b) because it stimulates causal thinking about the causal claim that anecdotal evidence supports. Both processes contribute to the ease with which people generalize information about anecdotes to a population.

The first reason concerns the availability heuristic. In numerous experiments, an anecdote was operationalized as a vivid story with a beginning, a course of action, and an end (e.g., in Slater & Rouner, 1996). If people process such an anecdote, it will remind them of similar events (Strange & Leung, 1999). As a result, the probability of the event in general will be perceived as higher. This process is explained through the availability heuristic, which holds that people judge the probability of an event on the ease with which instances of similar events come to mind (Tversky & Kahneman, 1973). The more easily past occurrences of events can be retrieved from memory, the more probable the events are judged. The presentation of an anecdote facilitates retrieval of other events, and enhances the generalizability of the events. Anecdotal evidence does not necessarily possess a narrative structure. O'Shaughnessy and O'Shaughnessy (2004) distinguish the narrative from the anecdote. Whereas a narrative is "a story that accounts for the sequence of events" (p. 31), an anecdote is short and illustrative. Example (1), for

instance, is relatively pallid anecdotal evidence as support for the claim about the implementation of comprehensive exams.

(1) The implementation of a comprehensive exam at the University of Rotterdam has led to higher starting salaries.

The exemplar-based model of social judgment (Smith & Zárate, 1992) suggests that even relatively pallid anecdotal evidence such as (1) facilitates the process of evoking similar events. This model assumes that people retrieve information about similar instances whenever they encounter a new target instance, such as in anecdotal evidence. These similar instances in memory are cognitive representations, and are called exemplars. Exemplars may be very detailed or very minimalistic. This implies that even for examples such as (1) similar events can be evoked, and that, if this retrieval is easy, the generalizability of the event is facilitated.

The second reason why anecdotal evidence can be persuasive is that it may lead to spontaneous causal processing. Anderson (1983) gave his participants information that was similar to either anecdotal or statistical evidence that suggested a positive or a negative relationship between fire fighters' preference for risk and their success. After being debriefed about the fictitious nature of the anecdotal or statistical evidence, participants were asked to write down their beliefs about this relationship. Anecdotal evidence had more impact: participants who had been given anecdotal evidence believed more in the causal relationship than participants who had read the statistical evidence. A vividness explanation (anecdotal evidence was more vivid, so it was better recalled) was ruled out. In contrast, anecdotal evidence proved to stimulate causal thinking. Participants with anecdotal evidence wrote down more causal explanations for the relationship than participants with statistical evidence, and this made them believe more in this relationship. The impact of causal thinking will be elaborated on in the next section about causal evidence.

Causal evidence

A large number of studies have explained the impact of causal reasoning and thinking in judgments and decision making. What makes causal evidence persuasive? Some studies show that causal thinking about the occurrence of an event enhances the perseverance of the belief in this occurrence, even after discrediting the data that created this belief. Causal evidence appears to borrow its

persuasive power from the fact that it is difficult to refute it (e.g., Anderson, 1983; Anderson, Lepper, & Ross, 1980; Ross, Lepper, Strack, & Steinmetz, 1977). In a series of experiments, Ross et al., for instance, showed that people's subjective likelihood of the occurrence of an event increases when they are asked to give an explanation for this event. In study I, participants were given a clinical case history about a patient, and the information of an event in the patient's life. Half of the participants were then asked to provide an explanation why the event had occurred on the basis of the available case history. Subsequently, all participants were told that the event had not occurred. Next, they were asked to judge the likelihood of five events of which one event was the event they had read before. Participants who had been asked to provide explanations for the first event considered this event as more likely than the participants who had given no explanations, even though this first event was said not to have occurred. The results were corroborated in study II, in which participants were told at the outset that the first event had not occurred, and that it was hypothetical. Even under this circumstance, hypothetical events that had been explained were considered as more likely than events that had not been explained. Studies such as Ross et al. indicate (a) that beliefs for which an explanation is generated are perceived as more probable, and (b) that causal information is difficult to refute when people generate it themselves. This last effect is so strong that Brem and Rips (2000) call this the explanation effect.

Expert evidence

Expert evidence usually consists of an expert's repetition of the claim. An explanation for the persuasiveness of expert evidence may be related to people's expectation that experts are capable of giving supplementary support when they are asked to (cf. Facione & Scherer, 1978). Similarly, people may expect experts to provide strong (as opposed to weak) arguments. This suggestion was supported in Wiener, LaForge, and Goolsby (1990), and in Bohner, Ruder, and Erb (2002). In the study of Bohner et al., for instance, participants were given a short text about a professor (expert) or a student, and were asked to write down their expectations about that source (e.g., How convincing will the source's arguments be?). The subsequent message of the source contained weak or strong arguments, as defined in the ELM tradition (see Section 1.1). Participants who had received the message with weak arguments were far less persuaded when the source was a professor than when it was a student. This means that when people are asked to explicitly think about experts, they expect them to be capable of providing strong arguments.

In conclusion, these studies lend support to the suggestion that expert evidence may be persuasive because experts are expected to be in a position to provide strong supplementary support.

Until now, I have shown that the normative criteria presented in Chapter 2 may affect the persuasiveness of evidence types under certain conditions (3.1), and that there are other explanations for the persuasiveness of the types of evidence (3.2). In the next section, I will not treat the types of evidence separately, but I will review studies that have investigated how they relate to each other in their persuasive effectiveness.

3.3 A review of studies on the relative persuasiveness of evidence types

The persuasiveness of evidence has been empirically investigated for more than 60 years. In the first review of experimental research on the effects of evidence, McCroskey (1969, p. 170) concluded that "no firm generalization can be drawn concerning the effect of evidence on attitude change". He had found nine experiments, of which only two showed that the use of evidence was more effective than no use of evidence. Not surprisingly, some researchers were quite cynical towards the use of evidence, such as Gregg (1964):

> "the audience reaction to an argument may have little or nothing to do with whether the argument includes fully documented or completely undocumented evidence, relevant or irrelevant evidence, weak or strong evidence or any evidence at all" (cited in McCroskey, p. 169)

It was not until the 1980s that the persuasive impact of evidence was assessed in a review by Reinard (1988). On the basis of available experimental studies, he convincingly showed that inclusion of evidence enhances the persuasiveness of a message. If providing evidence is persuasively effective, the question that logically follows is whether some types of evidence are more persuasive than others. As a matter of fact, recent studies on the persuasiveness of evidence did not investigate whether evidence is persuasive or not, but compared the *relative* persuasiveness of different types of evidence. Before I provide a review of such studies (3.3.3), I will

first argue why a new review is desirable (3.3.1), and present the criteria used to select studies for this review (3.3.2).

3.3.1 Inconsistencies in existing studies

The first published article on the persuasiveness of evidence types was Cathcart (1955). Participants listened to a speech in favor of the abolition of capital punishment that contained one of four – as he called it – methods of presenting evidence. Current studies still resemble this Cathcart study, both with respect to their central research question (which type of evidence is the most persuasive?) and with respect to their research design (a speech or text containing a claim that is supported by different types of evidence presented to participants in a between-subject design). Despite these similarities, evidence studies have had quite inconsistent definitions and operationalizations of evidence and the evidence types. Kellermann (1980) wrote a critical review about the concept of evidence. She argued that the inconsistent results of empirical studies conducted before the 80s can be attributed to methodological weaknesses, and to the "inability to adequately isolate, limit and define 'evidence' as a construct" (p. 159). Her methodological remarks are not specific for evidence studies (e.g., about the lack of control for variations in participant's characteristics, and in message topics). Her commentary about the inconsistency in definitions of evidence is interesting. Although evidence is considered as a form of proof in all studies, the "definitions of 'evidence' are almost as numerous as the individuals who work in the area" (p. 162). The same goes for the operationalizations of evidence types. Kellermann gives the example of source evidence, which has been operationalized through a presence or absence of an authority attribution, a general or vague attribution, the amount of attribution (date, place, name, qualifications), or the source reliability.

Kellermann's (1980) analysis makes it clear that the inconsistencies in definitions and operationalizations of evidence fog our understanding of which type of evidence is the most persuasive. Reviews of evidence studies may provide more insight into the relative persuasiveness of evidence types. Three reviews have been presented over the years[2]: Reinard (1988), Baesler and Burgoon (1994), and Allen and Preiss (1997). The descriptive reviews of Reinard, and Baesler and Burgoon concluded that anecdotal evidence was more persuasive than statistical evidence. Allen and Preiss, however, answered the same question through a meta-analysis, and found that it was statistical evidence that was more persuasive. How can this difference be explained? A straightforward explanation is the research method: a meta-analytic review may lead to other conclusions than a descriptive

review. Another explanation is that the selection of studies in the reviews differs considerably. Only two experimental studies were included in all of the three reviews. As existing reviews have not taken into account differences in definitions and operationalizations of evidence types, and because a number of experiments have been published since 1997, a new review is desirable. For this review, more rigid criteria have been used to ensure a more homogeneous selection of studies, on the basis of which more reliable conclusions can be drawn. In the following section, I will present these criteria.

3.3.2 Selection of studies

In the ideal situation, only studies in which the evidence operationalizations meet the normative criteria of Chapter 2 should be included in the review. A first advantage of these operationalizations is that they allow a comparison between different studies. It is inconvenient to interpret the results of two studies if evidence was normatively stronger in the first than in the second study. A second advantage lies in the comparison of the types of evidence within each study. If the operationalization of one type of evidence is normatively stronger than a second type, a comparison of the persuasiveness of the two types is not appropriate. Unfortunately, as evidence studies have not taken into account these normative criteria, these criteria could not be a means to select studies to be reviewed. The criteria that I present, however, did result in a smaller selection of studies that are more similar with respect to a number of characteristics than in previous reviews. Published experiments about the relative persuasiveness of evidence types were selected employing the following five criteria:

1. Evidence (called information, data, or evidence) is used as proof to enhance the probability of a claim. The claim can be explicitly stated, but may also be inferred from the text message.

2. Type of evidence is an independent variable (statistical, anecdotal, causal, expert evidence, or combinations), and at least two types of evidence are compared.

3. The definitions of the types of evidence (explicitedly stated or inferred from the operationalizations) correspond to those of Rieke and Sillars (1984, pp. 74-94). Anecdotal evidence consists of a specific case, and statistical evidence is a numerical summary of a number of cases. Causal evidence consists of an explanation for the occurrence of the effect, and expert evidence consists of the testimony of an expert. With these definitions, the problem of different

operationalizations is partly covered. Different forms of statistical evidence were accepted. Statistical evidence can be divided into four types (e.g., Ehninger & Brockriede, 1963)[3]: a number (e.g., a study among 14 universities), a percentage (e.g., 90% of the universities), a comparison (e.g., students at such universities have four times more chance to get a high starting salary than students at universities without such an exam), or a mean (e.g., students at such universities have, on average, $150 more starting salary per month). A few studies were not selected, because their operationalizations were not mentioned (Harte, 1976), or because the operationalizations were different from the ones given above (Greene & Brinn, 2003; Iyengar & Kinder, 1987; Knouse, 1983; Rook, 1986; Spooren, Smith, & Renkema, 2000). In Greene and Brinn, for instance, statistical evidence apparently did not contain any numerical information. Iyengar and Kinder (study VI) investigated the impact of vivid (anecdotal) and pallid (statistical) information. Statistical evidence was not always statistical, for instance when a reporter discussed the link between a chemical dump and catastrophic illness as support for the message about an increase of childhood leukaemia[4]. Finally, differences in length and vividness of especially anecdotal and statistical evidence were accepted, as anecdotal evidence has traditionally been operationalized as a long, vivid story.

4. Participants are exposed to only one type of evidence for each claim. This research design was chosen in the majority of the persuasion studies on evidence. In a large number of studies in cognitive psychology and mass communication, however, the impact of information that corresponds to either statistical or anecdotal evidence was examined by providing participants both types of evidence at the same time (see Appendix B). Because of this criterion, a lot of studies were not used for the review (cf. Allen & Preiss, 1997). I did include studies in which participants were given two different claims with another type of evidence each time (e.g., Koballa, 1986). In such studies, the judgment of the first claim with evidence is expected not to have affected the judgment of the second claim with evidence, as the claims were different.

5. The acceptance of the claim supported by evidence is the dependent variable. Acceptance may be measured in terms of probability (e.g., Hoeken, 2001a) or prediction (e.g., Dickson, 1982), as an attitude toward an object or behavior (e.g., Koballa, 1986), or as an intention (e.g., Sherer & Rogers, 1984).

I looked for publications in various ways. First, I consulted previous reviews on the effectiveness of the use of evidence (McCroskey, 1969; O'Keefe, 1998; Reinard, 1998; Stiff, 1986), reviews on the persuasiveness of different types of evidence (Allen & Preiss, 1997; Baesler & Burgoon, 1994; Reinard, 1988), and non-empirical articles on evidence (e.g., Reynolds & Reynolds, 2002). Second, I searched through

databases and document-retrieval services (PsychINFO, Sociological Abstracts, Econlit, Current Contents Weekly, Current Contents Archives, and Social Sciences[5]). Finally, I used the snowball method on the basis of the publications that had been found to increase the number of publications.

This resulted in 14 studies on the relative persuasiveness of evidence types. To illustrate of what the five criteria do with the selection of studies, I will show which studies in the three previous reviews on statistical and anecdotal evidence meet these criteria. The review of Reinard (1988) consisted of 15 studies on statistical and anecdotal evidence. Of these 15 studies, 6 studies were not published, 6 studies presented anecdotal and statistical evidence together (criterion 4), and 3 studies did not meet my criteria for definitions and operationalizations (criterion 3). In the Baesler and Burgoon (1994) review, 19 evidence studies on the relative persuasiveness of statistical and anecdotal evidence were selected. Of these studies, 7 studies remained unpublished, 8 studies presented anecdotal and statistical evidence together (criterion 4), 1 study was Iyengar and Kinder (1987) (criterion 3), and the 3 remaining studies were included in my review (Dickson, 1982; Kazoleas, 1993; Koballa, 1986). The review of Allen and Preiss (1997) dealt with 16 studies, of which 5 studies were not published, 3 studies presented anecdotal and statistical evidence together (criterion 4), 1 study compared two forms of statistical evidence (criterion 2), and 1 study was Harte (1976) (criterion 3). In the end, 6 studies were included in my review (Baesler & Burgoon, 1994; Dickson, 1982; Kazoleas, 1993; Koballa, 1986; Sherer & Rogers, 1984; Slater & Rouner, 1996).

3.3.3 Results

The studies that met the five criteria are discussed on the basis of a comparison between two types of evidence at a time: statistical versus anecdotal, statistical versus causal, anecdotal versus causal, and expert evidence versus the three other types of evidence. Each paragraph contains a short presentation of the studies, an overview of the results, and a discussion. Table 3.1 lists the studies with their results.

Statistical – anecdotal

In Allen et al. (2000), participants received 1 of 15 texts (e.g., women's use of cosmetics) that contained either statistical or anecdotal evidence. The definitions of both types of evidence correspond to criterion 3, but no information about the operationalizations was provided. Statistical evidence proved to be more

persuasive than anecdotal evidence. Baesler and Burgoon (1994) investigated the relative persuasiveness of statistical and anecdotal evidence in support of the claim that most juvenile delinquents do not become adult criminals. Statistical evidence was more convincing than anecdotal evidence. The finding of Dickson (1982) can be interpreted as a persuasive advantage for statistical evidence. Participants were given a report about a breakdown rate of a refrigerator brand with either anecdotal or statistical evidence. Anecdotal evidence consisted of four housewives who had not had any problem with their refrigerator, and one housewife who had indeed encountered a problem.

Table 3.1 Results of the selected studies on the relative persuasiveness of evidence types (nc = not communicated)

study	type of claim	result
statistical – anecdotal evidence		
Allen et al. (2000)	nc	st > an
Baesler (1997)	general	st = an
Baesler & Burgoon (1994)	general	st > an
Cox & Cox (2001)	general	st = an
Dickson (1982)	general, specific	st > an
Hoeken (2001a)	specific	st > an
Hoeken (2001b)	specific	st = an
Hoeken & Hustinx (2003b)	general	st > an
Kazoleas (1993)	general	st = an
Koballa (1986)	general	an > st
Sherer & Rogers (1984)	general	st = an
Slater & Rouner (1996)	general	st > an
statistical – causal evidence		
Hoeken (2001a)	specific	st > ca
Hoeken & Hustinx (2003b)	general	st = ca
Slusher & Anderson (1996, study I)	general	ca > st
Slusher & Anderson (1996, study II)	general	ca > st
anecdotal – causal evidence		
Hoeken (2001a)	specific	an = ca
Hoeken & Hustinx (2003b)	general	ca > an
expert evidence – other types of evidence		
Hoeken & Hustinx (2003b)	general	ex > an, ex = st, ex = ca

In the statistical evidence, 395 housewives were reported to have had no problem, whereas 105 housewives said their refrigerator had broken down. Note that in both versions, around 80% of the sample did not have any problem with their refrigerator. Participants had to predict the breakdown rate in the hypothetical situation in which they owned this refrigerator themselves (specific claim), and had to predict the breakdown rate of 20 refrigerators (general claim). With both claims, the prediction of a breakdown was higher with anecdotal evidence. Although the judgments on the basis of statistical evidence were closer to the 20% breakdown rate, Dickson interpreted the higher prediction of anecdotal evidence as more influential than statistical evidence. However, I believe, and so do Allen and Preiss (1997), and Baesler and Burgoon (1994), that statistical evidence was more persuasive. If the message had had a persuasive character, its objective would have been to convince the readers that only 20% of the refrigerators of this brand break down. Participants with statistical evidence were more persuaded than those with anecdotal evidence, as their judgments were closer to the 20%.

In Hoeken (2001a), participants were given statistical, anecdotal, or causal evidence in support of the specific claim that a new cultural center would be successful. The anecdotal evidence stated that a similar center in another city had been successful. In the statistical evidence, the success of 27 such cultural centers was reported. Statistical evidence proved to be more convincing. In Hoeken and Hustinx (2003b), participants were given 20 general claims, which were accompanied by statistical, anecdotal, causal, expert, or no evidence. An example of a claim was 'Relaxation rooms in offices lead to a sharp decline of absence through illness'. Statistical evidence was found to be more persuasive than anecdotal evidence. In Slater and Rouner (1996), the claim was that alcohol is a harmful presence in society, because it leads to health risks, economic and career harm, and drunk driving[6]. Anecdotal evidence consisted of a person who had experienced these problems because of alcohol. An example of statistical evidence was that 25% of men who occasionally drink have these problems. Statistical evidence was found to be more convincing.

A few studies have reported no difference in the persuasiveness of statistical and anecdotal evidence. In Baesler (1997), these two types of evidence were used to support three messages about crime, internships, and birth control. Evidence definitions and operationalizations were similar to those of Baesler and Burgoon (1994). In Cox and Cox (2001), female participants were given information about the benefits of regular screening mammographs. In the statistical evidence, an early treatment was said to reduce the risk of dying of breast cancer by 30%. The

anecdotal evidence consisted of a report of a successful story of a woman. Statistical and anecdotal evidence were found to be equally persuasive. In Hoeken (2001b), the specific claim was that local taxes in a certain town should be raised in order to install extra streetlights on the sidewalks, because that would reduce the number of burglaries. The statistical evidence consisted of a study among 48 towns showing that installing extra streetlights decreased the number of burglaries with 42%. In the anecdotal evidence, the effect was shown to have occurred in another town. The two types of evidence were equally persuasive. The participants in Kazoleas (1993) received a message that advocated the use of safety belts. The message was supported by statistical evidence (people have a 50% smaller risk of getting injured when wearing a safety belt) or anecdotal evidence (an example of a person who used the safety belt). The attitude toward the use of safety belts was the same after statistical or anecdotal evidence. Sherer and Rogers (1984) constructed an essay in which the main claim was that less drinking avoids certain dramatic consequences. Anecdotal (2 drinkers) and statistical evidence (statistics about a group of 2 000 drinkers) were used to support this claim. The two types of evidence were equally effective in influencing the intention to limit alcohol use, and to abstain from it.

Koballa (1986) found that anecdotal evidence was more persuasive than statistical evidence in support of two general claims that the introduction of new science programs would be useful because it would lead to better results. Participants were given two messages, each time for a different science program with another type of evidence. The types of evidence did not compete with each other, as they each supported another message. Anecdotal evidence consisted of a report of a person who had experienced the program. In the statistical evidence, the usefulness of the programs was demonstrated in numerous studies.

In sum, 12 studies have compared the relative persuasiveness of anecdotal and statistical evidence. Statistical evidence proved to be more persuasive in 6 studies, anecdotal evidence in 1 study, and in the 5 remaining studies, no differences were found (see Table 3.1). The results of this review correspond with those of the meta-analysis of Allen and Preiss (1997), but not with Baesler and Burgoon (1994), or Reinard (1988). As I argued in 3.3.1, this discrepancy of conclusions can be explained by the selection of studies. Still, the present review shows conflicting results. This may partly be explained by the types of claims that were used: general or specific. With a general claim, statistical evidence is normatively stronger than anecdotal evidence because it has a large sample size. This does not hold under the condition of a specific claim, because there the criterion of a large sample size is no

longer relevant for anecdotal evidence. Three of the studies in the review used specific claims (see Table 3.1). For only one of these studies, anecdotal evidence was as persuasive as statistical evidence (Hoeken, 2001b). Although these opposing results can not be fully explained through the types of claims that were used, type of claim appears to be a factor to take into account[7].

Statistical – causal

In Hoeken (2001a), participants were given statistical, anecdotal, or causal evidence in support of the specific claim that a new cultural center would be successful. In the causal evidence, three reasons were provided why this center would be profitable. For instance, one of them was that a movie theatre in a nearby town had burnt down. Statistical evidence was found to be more convincing. In a follow-up study that used more claims (Hoeken & Hustinx, 2003b), there was no difference in the persuasiveness of statistical and causal evidence. Finally, Slusher and Anderson (1996) supported the general claim that Aids is not transmitted by personal contact or mosquitoes with causal and/or statistical evidence[8]. An example of the statistical evidence was that in rural villages, where mosquitoes are likely to spread the disease, only 0.8% of the people have the Aids virus. Causal evidence consisted of reasons why mosquitoes do not spread the disease, for instance, because they are too small to carry enough viruses to infect a person. Causal evidence was found to be more persuasive.

In sum, the results are mixed: sometimes statistical evidence was more persuasive (Hoeken, 2001a), sometimes causal evidence was more convincing (Slusher & Anderson, 1996; studies I and II), and sometimes there was no difference (Hoeken & Hustinx, 2003b). This means that there is no clear indication whether causal and statistical evidence are equally persuasive or whether one of them is more persuasive than the other.

Anecdotal – causal

The two studies that compared these types of evidence show different results: in Hoeken (2001a) the two types are equally persuasive, whereas in Hoeken and Hustinx (2003b) causal evidence is more convincing. As there are only two studies that directly compared the persuasiveness of anecdotal with causal evidence, it is impossible to draw reliable conclusions about which type of evidence is more persuasive. Also, the type of claim was different in the two studies. The claim was specific in Hoeken, and general in Hoeken and Hustinx. Anecdotal evidence appears to be less persuasive in support of a general claim (because of the criterion

of a large sample size), than in support of a specific claim. However, there are two reasons to suggest that causal evidence is more convincing than anecdotal evidence. First, this result was obtained in Hoeken and Hustinx who investigated the effect of evidence for 20 claims, which enhances the generality of its result. Second, although there was no difference in the actual persuasiveness of anecdotal and causal evidence in Hoeken, participants did perceive the persuasive power of causal evidence as higher than that of anecdotal evidence.

Expert – other types of evidence

In the only study that investigated expert evidence with other types of evidence (Hoeken & Hustinx, 2003b), this type of evidence was found to be as persuasive as statistical and causal evidence, and more persuasive than anecdotal evidence.

3.3.4 Conclusion

When the results of this review are taken together, it appears that statistical and causal evidence are more persuasive than anecdotal evidence. As one study shows that expert evidence too is more persuasive than anecdotal evidence, the overall, tentative conclusion would be that anecdotal evidence is the least persuasive type of evidence. The conflicting results that were sometimes found could be attributed to different factors, of which (a) the participants' elaboration, (b) the experimental design, and (c) the lack of control for evidence quality appear to be the most relevant. First, under the condition of high elaboration, normatively strong arguments are expected to be more persuasive than normatively weak arguments. Therefore, one could expect statistical evidence to be more persuasive than anecdotal evidence in this central route to persuasion. In cases where statistical evidence was not more persuasive than anecdotal evidence, it could be suggested that participants might not have scrutinized the persuasive message. As it is difficult to determine whether the participants in the 14 studies in the review were in the central (high elaboration) or peripheral route (low elaboration), this explanation can not be tested.

Second, most of the 14 studies used single-message designs, in which participants were given only one text with a type of evidence. Differences in the results between these studies are likely to be affected by the topic of the different messages that were used (e.g., refrigerators, alcohol, science programs). In order to be able to make more reliable generalizations about the effects of evidence, a multiple message design is more appropriate. This was used by Allen et al. (2000), Baesler (1997), and Hoeken and Hustinx (2003b). Finally, a factor that may explain

the opposing findings is that it is difficult to compare the persuasiveness of evidence types without controling for the quality of evidence. Some types of evidence were normatively stronger than others, and this also varied across the studies. In my empirical studies, I took into account the normative criteria presented in Chapter 2 in order to allow a fair comparison of the relative persuasiveness of evidence types. I did not only focus on *actual* persuasiveness – as was done in this chapter – but also considered *expected* persuasiveness. The next two chapters are concerned with this expected persuasiveness.

Notes

1. It should be noted that a few studies have shown that people also generalize on the basis of one single case when it is the only information that is given (e.g., Strange & Leung, 1999). This, however, does not disconfirm the conclusion of this section that evidence with larger sample sizes may be more persuasive than evidence with smaller sample sizes.
2. Four reviews have investigated the effectiveness of (one type of) evidence. McCroskey (1969) wrote a descriptive review that was aimed at indicating whether source evidence was effective or not. Reinard (1998) investigated the effect size of the inclusion of source evidence. In the two other reviews, the definition of evidence was generally too broad. Stiff (1986) selected quite heterogeneous studies that consisted of supporting information, such as evidence studies, but also studies on argument quality, and the number of arguments. Finally, the meta-analysis of O'Keefe (1998) focused on the effects of justification explicitness: source citation, completeness of arguments (premises, supporting information), and quantitative specificity.
3. Whereas it is easy to relate a number or a percentage to the argument by generalization, it requires more effort to interpret statistical evidence that consists of a comparison or a mean. For comparisons and means, however, one can suppose this numerical information to be based on numerous cases, even if this is not mentioned.
4. Knouse (1983) studied the effect of anecdotal and statistical evidence in a letter of recommendation. Statistical evidence consisted of a grade point average (a correct operationalization), but also of other numerical data, such as the number of years that the writer had known the student, and the number of surveys that the student had administered in a research project. In Rook (1986), the operationalization of statistical evidence was not clear. The claim about health risks for women was supported by "information with reference to women in general" (p. 528). Finally, in Spooren et al. (2000) statistical and anecdotal evidence supported the claim that a save-as-you-earn account is flexible. Statistical evidence was operationalized as a quantification of interest rates and amounts of money. This may be a correct operationalization, if one can assume that it is based on large number of cases (i.e., customers). However, this is not the case: the interest rate can be determined when the bank has no customers.
5. The following keywords were used: anecdotal evidence + empirical, base-rate + anecdote, base-rate + attitude, base-rate + case, base-rate + persuasion, causal evidence, evidence + documentation, evidence + persuasion, evidence + persuasive, evidence quality, evidence types, expert evidence, narrative evidence, statistical + anecdotal, statistical + anecdotical, statistical evidence + persuasion, and support + persuasi*.

6. This study partly does not meet criterion 4, because two different types of evidence were given together for one message. This study has been taken into account nevertheless, because in two of the four conditions, participants were only exposed to one of the two types of evidence.
7. Hoeken and Hustinx (2003a) report on an unpublished experiment that investigated whether anecdotal evidence was more persuasive with a specific than with a general claim. For the general claims, statistical evidence was more persuasive than anecdotal and causal evidence. However, with the specific claims, there was no significant difference in the persuasiveness of the three types. This study suggests that the difference in persuasiveness of statistical and anecdotal evidence decreases with a specific claim.
8. See footnote 6.

Expected persuasiveness: a corpus-analytic approach

Cultural differences in the *expected* persuasiveness of evidence types were investigated in two studies that aimed to answer the first general and specific research questions:

General RQ1 Are there cultural differences in the relative *expected* persuasiveness of anecdotal, statistical, causal, and expert evidence?

Specific RQ1 Is the *expected* persuasiveness of expert evidence higher for the French than for the Dutch?

Expected persuasiveness can be measured directly ("How persuasive do you expect this piece of statistical evidence to be?"), and indirectly by investigating the use of evidence types in a persuasive setting. Study 1 used an indirect measurement of expected persuasiveness, and Study 2 a direct measurement. These ways of measurement will be discussed in 4.1, where I will present a review of the few empirical studies on the relative expected persuasiveness of evidence types. Study 1 is presented in this chapter, Study 2 in Chapter 5.

4.1 Expected persuasiveness of evidence types

In contrast to the considerable number of studies on the *actual* persuasiveness of evidence types (see Chapter 3), there are only a few studies that have examined their *expected* persuasiveness. In some of these studies, expected persuasiveness was measured directly by asking participants to rank evidence types in terms of expected persuasiveness (Kline, 1971a). Other studies investigated expected persuasiveness indirectly on the basis of the actual use of evidence types (e.g., Kline, 1971b; Levasseur & Dean, 1996).

Kline (1971a) was probably the first to study the expected persuasiveness of evidence types. Participants, including college students and PhD candidates, were given a claim, either (a) 'Requirements for admission to college should be raised', or (b) 'The Federal government should exercise more control over primary and secondary education'. The claim was followed by 25 instantiations of evidence that differed on the dimensions of specificity and relevance, and on the attribution to different sources (differing in expertise and trustworthiness). For the college admission claim, three examples of evidence without an attribution to a source read as follows (p. 190):

(1) Although 65% of the students in the upper half of entering freshman classes will graduate, only 23% of those in the lower half will graduate (specific and relevant).

(2) Although a large percentage of the students who rank high in entering freshman classes will graduate, only a small percentage of those who rank low will graduate (non-specific and relevant).

(3) Nearly 40% of the college students come from high schools with an enrolment of less than 300 students (specific and irrelevant).

Participants ranked these instantiations according to the likelihood that they would use them to persuade either college students of the college admission claim, or the local Parents-Teachers Association of the government control claim. The main goal of Kline (1971a) was to investigate whether people differed in their ranking. Response patterns were analyzed in such a way that participants were clustered in different groups of encoders. Four groups were identified, some for both claims, others for only one of the claims. For the first group, the source credibility was essential (evidence with trustworthy and expert sources was preferred), and for the second group evidence attributed to a source – regardless of its trustworthiness or expertise – was preferred over evidence without a source. For the third group, the complexity of the evidence affected the ranking (evidence where two or fewer dimensions were manipulated had the highest expected persuasiveness), and the fourth group preferred specific and relevant evidence. Kline showed that people may differ in their expected persuasiveness of the evidence types, but – as the groups did not differ in initial attitude or on demographic characteristics (e.g., political preferences) – he was unable to explain why people differ.

Other studies have indirectly investigated the expected persuasiveness of evidence types. Instead of explicitly asking for an expectation of the persuasiveness of evidence, the use of different evidence types was analyzed. Kline (1971b) investigated the use of evidence types in students' speeches. He hypothesized that highly dogmatic participants (who are said to be closed-minded, intolerant, and deferential to authority; Rokeach, 1960) would use documented evidence in favor of undocumented evidence. Although Kline does not explain the difference between documented and undocumented evidence, I assume on the basis of Cathcart (1955) and Kline (1971a) that documented evidence is specific evidence that is attributed to sources. Participants were told that there are different types of evidence. The types of evidence were presented in such a way that nothing was said about possible differences in their persuasiveness. About two weeks later, each participant was asked to give a short speech, of which the use of documented and undocumented evidence was noted down. Highly dogmatic participants used more documented evidence than participants who scored low on the dogmatism scale, but these participants used more undocumented evidence than the highly dogmatic participants.

Levasseur and Dean (1996) analyzed the actual use of evidence types in debates between presidential candidates in the United States. They distinguished statistical, anecdotal, and source evidence. Anecdotal evidence consisted of specific historical instances, and statistical evidence of numerical instances (simple: 500 people; complex: 45%). Source evidence relied on experts or lay people. Based on the data they provide, I calculated that the presidential candidates more frequently used statistical (45%) and anecdotal evidence (44%) than source evidence (11%)[1].

Some argumentation scholars have been interested in the occurrence of different argumentation schemes in written discourse (e.g., Schellens, 1985; Schellens & De Jong, 2004). Schellens and De Jong investigated the use of argumentation schemes in persuasive texts. They analyzed 20 Dutch persuasive information brochures to determine how many used the different argumentation schemes. Pragmatic argumentation was employed in all the brochures. The two other most frequently used argumentation schemes were causal argumentation (in 75% of the brochures) and argumentation on the basis of examples (in 70% of the brochures). The argument by authority, which is related to expert evidence, occurred in 30% of the brochures. Schellens and De Jong considered authoritative institutions, (unnamed) scientists, celebrities, and the collective's opinion (e.g., 'more and more people') as authorities. The argument by authority only occurred

in relation to the probability and the desirability of consequences in pragmatic argumentation. It was not employed as support in another argumentation scheme.

This kind of corpus analysis can be used to investigate the expected persuasiveness of evidence types. Writers of persuasive texts are professionals whose job is to design texts that are highly convincing. These writers can be expected to specifically use the evidence types that they expect to be most effective for the target audience. Expected persuasiveness can therefore be measured on the basis of the use of evidence. If the audience has a large power distance, expert evidence could be more persuasive than when the audience has a small power distance (see Chapter 1). Therefore, text writers from large power distance cultures might use expert evidence more frequently than their colleagues from small power distance cultures. Study 1 was set up to investigate the use of evidence types of text writers of persuasive information brochures in France and the Netherlands.

Study 1

This first study was set up to answer the general and specific research questions about cultural differences in the relative expected persuasiveness of anecdotal, statistical, causal, and – in particular – expert evidence. As expected persuasiveness was measured on the basis of the actual *use* of evidence by professional text writers, the concrete research questions of Study 1 were the following:

RQ1 Are there cultural differences in the relative use of anecdotal, statistical, causal, and expert evidence in persuasive communication in France and the Netherlands?

RQ2 Is expert evidence more frequently used in persuasive communication in France than in the Netherlands?

An additional aim of this study was to assess the cross-cultural equivalence of the construct of evidence (Is evidence employed to support the probability within the scheme of pragmatic argumentation in both cultures?), and of the distinction between the four types of evidence (Do the four types of evidence occur in both cultures?). After a discussion of the material (4.2.1) and the procedure (4.2.2), I will comment on the statistical analysis (4.2.3).

4.2 Method

4.2.1 Material

As in Schellens and De Jong (2004), persuasive information brochures produced by government agencies (e.g., the city of Amsterdam) and non-profit organisators (e.g., the Animal Protection Society) were used as material[2]. These brochures were all written to convince readers to behave in a certain way. The effects of the different forms of behavior are not always directed to the readers of the brochures. Schellens and De Jong make a distinction between an interest for the readers themselves (e.g., avoiding excessive alcohol consumption), for society (e.g., the environment), or for others (e.g., fundraising). The brochures were gathered between October 2001 and August 2002. In the Netherlands, about 140 brochures were gathered in the public libraries and town halls of Amsterdam, Den Bosch, Nijmegen, and Veldhoven. A considerable number came from Postbus 51, the Dutch Government Information Service. A large part of the 100 French brochures in France were gathered through the French Committee for Health Education (CFES, 'le Comité français d'éducation pour la santé'), the French equivalent of Postbus 51. The remainder were received from the Local Hygiene and Health Service in Rennes ('Le service communal hygiène et santé') or were collected personally in a number of fundraising institutions in Paris (these institutions were found via *www.yeba.org/annuaire* and *www.aidez.org*).

Only a minority of these brochures was finally selected for analysis. A brochure was included in the final selection if it met two criteria: (a) it should have a persuasive character, and (b) it should have a counterpart brochure with the same topic but in the other culture. These criteria will now be discussed. First, the distinction between informative and persuasive texts is not a clear-cut case. From a theoretical point of view, the goals of both types of texts are different. The aim of informative brochures is to provide information to readers in order to help them form an opinion or make a decision. The goal of persuasive public information in the Netherlands is to change the readers' attitude or behavior in a direction proposed by the text writer (Koelen & Martijn, 1994). In France, similarly, the goal of persuasive public information is to propose a change of mentality or behavior that the citizen should adopt (Berthelot-Guiet & Ollivier-Yaniv, 2001). The distinction between information and persuasion is usually hard to make in reality, because persuasive brochures are often presented in an informative way (Schellens & De Jong, 2004), and because brochures may have multiple objectives, such as informing and persuading (De Jong & Schellens, 2000). Neither the apparent goal

of the brochure, nor the style in which it was written were considered as indications of the genre of a brochure, but the use of pragmatic argumentation was. Pragmatic argumentation is considered as the most popular argumentation scheme in persuasive communication (cf. Schellens & Verhoeven, 1994; Schellens & De Jong, 2004).

Brochures that contained pragmatic argumentation were only included if there was another brochure with the same topic in the other culture. In fact, the use of certain types of evidence may be related to the topic of a brochure. A text writer of a brochure about traffic accidents, for instance, might have much statistical information about accidents, whereas a fundraising organization may lack such information. Each brochure was attributed to a thematic category, such as alcohol, smoking, or support for Developing Countries. These categories were created on the basis of the available brochures. This matching procedure (cf. Neuendorf, 2002, p. 85) resulted in a corpus of 22 Dutch and 22 French brochures. In each corpus, there were 8 brochures with a receiver interest (e.g, alcohol, smoking), 1 brochure with a societal interest (environment), and 13 brochures with an interest for other people (Developing Countries, and fundraising).

The 44 brochures differed in format, text length, and number of pages. In the Dutch corpus, the number of pages ranged from 2 pages to 12 pages with a mean of 6.45 pages (SD = 2.89). The mean number of pages of the French brochures, 7.55 (SD = 2.69), did not differ significantly from the Dutch average ($t(42) = 1.30$, $p = .20$). The smallest French brochure had 4 pages, and the largest 16. There was no criterion for text length in the selection of brochures. It seems logical to suggest that larger brochures have more instantiations of evidence than smaller brochures, but this appeared not to be true. Although there was a positive correlation between the number of pages and the number of evidence instantiations in the Dutch corpus (r (22) = .67, $p < .01$), there was no significant correlation in the French corpus (r (22) = .22, $p = .33$). Regardless of the number of pages, brochures may differ in the number of actions (e.g., quit smoking, and reduce the number of cigarettes per day), and in the number of related positive and negative consequences (e.g., feeling better, getting positive reactions from friends). As a result, the number of evidence instantiations in support of the occurrence of consequences may also differ. These differences in the number of actions and consequences will be discussed below. Appendix C lists the 44 brochures with their title, sender, and number of pages. Note that the terms ‘sender’ and ‘text writer’ are used alternately in this chapter, although the sender of brochures usually is not the writer.

4.2.2 Procedure

The persuasive information brochures were analyzed with the grid of pragmatic argumentation. Evidence was counted if it was employed to support the probability of the occurrence of effects as a result of behavior. In each brochure, behavior was recommended or advised against by pointing out its (un)favorable consequences. The procedure consisted of three stages: (a) determining the intended behavior, (b) searching for the (un)favorable consequences, and (c) seeking evidence in support for the probability that the behavior will lead to the consequences. Below, I will discuss in detail the elements *behavior*, *consequences*, and *evidence*. Excerpts of the brochures used to exemplify the procedure are presented in translation, just as the titles of the brochures.

Behavior

A brochure may recommend one action or a series of actions. In the majority of the brochures in the corpus, there was just one action. In the Dutch brochure '*What does alcohol do with you?*', for instance, the behavior advised against was drinking alcohol. An example of a more complex brochure is '*If everybody helps a bit, we will save our lives*', a French brochure about the consequences of risky behavior on the road. In this brochure, four kinds of behavior were discussed: drinking and driving, driving too fast, wearing no seat belt, and driving when tired.

Consequences

Regardless of the number of actions in a brochure, each action can be recommended or advised against by presenting one consequence or multiple consequences. In its most simple form, one consequence was given for one action. In '*If everybody helps a bit, we will save our lives*', the consequence of not wearing a seat belt was death. In most of the brochures, however, multiple consequences were given (cf. Schellens & De Jong, 2004). There were two forms here. First, an action was related to multiple positive or multiple negative consequences. In the French brochure '*If everybody helps a bit, we will save our lives*' the negative consequences of driving too fast were a reduced field of vision, a larger braking distance, and death. Second, for one single action, both positive and negative consequences were presented. This was the case in the Dutch brochure about alcohol ('*What does alcohol do with you?*'). Drinking alcohol was said to have the positive consequence of a reduced risk of cardiovascular disease, but also, and above all, a number of negative consequences, such as a higher risk of breast cancer, and reduced fertility[3].

Evidence

After determining the action(s) and the consequence(s) in a brochure, evidence as support for the probability of the occurrence of this/these consequence(s) was analyzed. Evidence often referred to the past when it concerned anecdotes, expert opinions, and numerical data. Readers are expected to infer that if evidence shows that behavior has led to consequences in the past, this causal relation will also apply to the future. Only verbal evidence was taken into account. Purely visual evidence, such as an image of a person, was not analyzed, as the interpretation of such images is often unclear (see Oestermeier & Hesse, 2000). Nevertheless, just as in Oestermeier and Hesse's study of verbal and visual arguments, most images – such as graphs, diagrams, or photographs – were directly related to a text fragment. In an indirect way, therefore, these images were taken into account. An example of an image that was related to an instantiation of evidence was a series of pictures of Abdoulaye in the French brochure '*Daily hunger*'. The claim that a donation would combat hunger was supported by a series of pictures with anecdotal evidence. Text fragments of the anecdotal evidence are given in (4):

(4) "On the first day, Abdoulaye (3 years old) weighs 5.5 kilos, the weight of an average French baby of 4 months old. Two nutritionists of 'Action contre la faim' take charge of him and his mother in Chad. [...] Around 20 days later, Abdoulaye smiles. He has gained 3.5 kilos. He now manages to stand up, but his legs are still too thin to allow him to walk. For the first time since his arrival, one can hear him sing" (*Daily hunger*)

The way the occurrence of evidence in the brochures was counted was not the same for each type of evidence. In the case of expert evidence, each instantiation was counted as one instantiation of expert evidence. An expert has a high level of knowledge about the claim he underscores. There appeared to be two types of experts: persons or institutions designated by name, or anonymous experts such as 'Studies show that'. This last category was not taken into account, as no difference in power distance can occur: there is a reader on the one hand, but no other person or institute on the other hand. Moreover, explicit identification can be considered as an indicator for reliable and credible experts (see Toulmin, Rieke, & Janik, 1984). An example of an expert is the General Inspection of Social Affairs (IGAS), which was used as support in a French brochure about host families for children ('*A real family life*'). Senders of the brochures that refer to their own expertise were not considered as expert evidence (cf. Schellens & De Jong, 2004), although this

strategy was sometimes used. An example of an appeal to own expertise, is quote (5) from the French brochure '*Thanks to you*':

(5) "Its ideas about development, its experience, and the number of years on the spot have made [the organization] an important, respected association that is listened to"

When the person was not an expert, I labeled expert evidence source evidence. Source evidence sometimes occurred in combination with another type of evidence, such as anecdotal evidence. In this case, a person who talks about experiencing the consequences of behavior was quoted. Example (6) shows a quote of Rik Felderhof, a Dutch television host. Felderhof is a Dutch celebrity, but not an expert in foster parents programs. Contrary to Schellens and De Jong's (2004) analysis, in which celebrities were considered as authorities, celebrities were not regarded as experts in this study, unless they were used as experts in their own field.

(6) "With financial support for a child like Joli we do not only help Joli, but also her little brothers and sisters, and the other children from her village. Being a foster parent helps them to go to school and to benefit from medical assistance [...] [was signed] Rik Felderhof"

When expert evidence was used in combination with another type of evidence, or when the source was no expert, it was not considered as expert evidence. These separate categories were not taken into account in the analysis, as they were rare (combinations of expert evidence with the three other types of evidence per brochure: $M = 0.20$, $SD = 1.07$; source evidence alone or in combination: $M = 0.14$, $SD = 0.51$). An example of expert evidence and statistical evidence, taken from '*If everybody helps a bit, we will save our lives*' is (7), which was used as support for the claim that drinking leads to death:

(7) "Twenty-eight percent of drivers that are involved in fatal accidents have too high an alcohol level. [...] (Source: National Institute for Transport and Safety Research [...])"

Concerning causal evidence, there appeared to be three varieties. First, causal evidence could consist of two causal relations. The claim that drinking alcohol makes you gain weight in '*What does alcohol do with you?*, for instance, was

supported by information about two causal relations: if you drink alcohol (behavior), the breakdown of fat will be reduced; as a consequence, you will gain weight (consequence). Second, an explanation could consist of a causal chain. In the Dutch brochure '*Tobacco*', for example, there was this causal chain from smoking to negative effects for the vocal cords: if you smoke (behavior), there will be more mucus in your throat; in your attempt to reduce the mucus, you will have to clear your throat; regularly clearing your throat will damage your vocal cords (consequence). Third, causal evidence could consist of several independent causal relations. In the Dutch brochure '*Amsterdam has good relations with Managua*', different causes were given for the claim that by helping the organization SAM the social circumstances of the people in Managua would improve. Two examples of causes were (a) SAM invests in projects that were set up by the local inhabitants, and (b) SAM collaborates with the city of Amsterdam that supports Managua with knowledge about public transport, drinking water, and waste processing. When causal evidence consisted of several causal relations, these were together considered as one instantiation of causal evidence. The reason for this is that these relations often constituted an explanation in combination.

Concerning anecdotal evidence, each instantiation of anecdotal evidence was considered as an independent unit. As each single anecdote may be sufficient, the choice for more anecdotes was considered deliberate. In the French brochure '*Together, we combat indifference*' seven anecdotes were given to support the claim that supporting the ACAT organization would help to combat torture in the world. Two examples of these seven anecdotes are (8) and (9):

(8) "Peru: In January 2001, three police officers who were suspected of having committed torture on three occasions were found guilty after the intervention of ACAT [...]"

(9) "France: In 2001, the legal support of ACAT helped about 20 people who had requested asylum to finally get the status of refugee. If they had been sent back to their country, their life would have been in danger"

Finally, statistical evidence consisted of numbers, percentages, or other numerical information that is supposed to be based on a large dataset (cf. footnote 3 in Chapter 3). A simple example of statistical evidence is (10), which was used in the French brochure '*If everybody helps a bit, we will save our lives*' to support the claim that wearing a seat belt reduces the risk of dying:

(10) "By correctly wearing a seat belt, the risk of dying when you are involved in a car accident is halved"

Sometimes, there was more than one numerical piece of information, like in (11), which consists of two numerical elements. As statistical evidence is often the result of research in which several elements are investigated, examples such as (11) were regarded as one instantiation of statistical evidence.

(11) "250 dead and over 2 000 injured per year because of drink driving" (*'What does alcohol do with you?'*)

When two or more instantiations of statistical evidence were not presented together as support for the same claim, and when it was not likely that this information came from the same research, these instantiations were considered as independent instantiations of statistical evidence. This was the case in the French brochure *'What is your position concerning alcohol?'*, where three instantiations of statistical evidence were employed to support claims that drinking leads to dangerous risks on the road. Two examples are given, (12) and (13):

(12) "One out of three mortal car accidents is directly caused by alcohol"

(13) "The risk on an accident doubles with 0.5 grams of alcohol in your blood, which is equal to a maximum of two consumptions"

Intercoder reliability

The brochures were analyzed by four independent coders (see Appendix C). I analyzed all the brochures, and three other coders each analyzed a third of the brochures. In total there were three groups of coders. First, we agreed upon the analytic procedure, as described above. Next, each coder independently analyzed 5 brochures. As the intercoder reliability within each pair of coders was good for these brochures, the brochures were taken into account in the analysis that followed. After the independent analyses of the remaining 34 brochures, the reliability was assessed for each pair of coders on the basis of two decisions made in each brochure: (a) Is a text fragment indeed evidence as support for the probability of the occurrence of an effect?, and (b) If the fragment is evidence, which type of evidence is it? The measurement of the reliability of the analyses is not the same for both questions. Potter and Levine-Donnerstein (1999) claim that the determination of reliability in corpus research depends on the type of content

that is in play. In this study, the two decisions are related to another type of content. The type of content concerned with the question whether a text fragment is evidence or not, is called pattern by Potter and Levine-Donnerstein. The role of the coders of a pattern is as follows:

> "Every time the coders see an element on the surface of the content that may indicate the presence of a particular pattern [here: evidence], they search for other indicators. When there are enough indicators [here: behavior and its consequences] or when the right combination of indicators is present, coders conclude that the pattern exists, and they record its presence on their coding forms" (p. 265)

Both coders in each pair checked whether they had coded the same text fragment as evidence. If this was the case, it was an agreement. If one of the two had not coded that fragment as evidence, it was a disagreement. In case of a disagreement, the coders together decided as to whether the fragment had to be coded as evidence or not. For each brochure one extra agreement was added, because both coders implicitly agreed that no evidence occurred in the rest of the brochure. The reliability of the analysis of the occurrence of evidence was calculated with percent agreement[4]. The second question – which type of evidence is it? – concerns a manifest content (Potter & Levine-Donnerstein, 1999). Once a text fragment was identified as evidence, the determination of the type of evidence was a relatively easy step. This content is manifest, because it is "on the surface and easily observable" (p. 259). The intercoder reliability for the type of evidence was calculated with Cohen's κ, but only when both coders had independently coded the same text fragment as evidence in the first decision. Table 4.1 shows the intercoder reliability for both decisions. The reliability proved to be very adequate or good, as the percent agreements and κ's are near or above .80 (see Neuendorf, 2002).

Table 4.1 Intercoder reliability for decision (a) (in percent agreement) and decision (b) (in κ) for each pair of coders (with number of brochures)

pair of coders (brochures)	(a) evidence or not?	(b) type of evidence
1 (16 Dutch)	.76	.73
2 (11 French)	.77	.85
3 (6 Dutch, 11 French)	.84	.83

4.2.3 Analysis

In order to test cultural differences in the use of evidence types, the χ^2 test would be a straightforward test, as the the presence (or absence) of evidence is measured at the nominal level. The percentage of expert or causal evidence on the total of evidence instantiations in the Dutch and French corpora could thus be compared. The significance of that comparison could then be tested with χ^2. The assumption of this method is that the observations (of evidence) are independent. In this study, however, as in some other corpus studies in the field of linguistics (Rietveld, Van Hout, & Ernestus, 2004), the observations may be dependent. The occurrence of one type of evidence may affect the probability that the next occurrence of evidence is of the same type. Many brochures contained several actions and multiple consequences. As a result, more than one piece of evidence was used in most of the brochures. In the Dutch brochure '*What does alcohol do with you?*', for instance, 3 positive and 11 negative consequences of drinking alcohol were presented. This gives the text writer 14 opportunities to provide evidence. If he or she expects one type of evidence to be most persuasive, he or she may use it more regularly than the other types, and not just for one consequence, but for a whole series of consequences. As a matter of fact, in this particular brochure, statistical and causal evidence were each used five times.

A suitable test is a 2 (culture) x 4 (statistical, anecdotal, causal, expert evidence) analysis of variance, taking brochures as the unit over which to generalize. Culture was a between-subjects factor, and type of evidence a within-subjects factor. With this analysis, relations within the brochures were taken into account. Each brochure was considered as a subject that had a score for each type of evidence (e.g., no anecdotal evidence, one statistical evidence instantiation, two instantiations of causal evidence, and no expert evidence). The results of this analysis will be presented in the next section.

4.3 Results

The corpus study was set up to test whether there are cultural differences in the relative use of the four types of evidence, and of expert evidence in particular. On average, a French brochure contained 3.45 (*SD* = 2.32) instantiations of evidence, with a minimum of no evidence and a maximum of 8 instantiations. The Dutch brochures contained only 2.41 (*SD* = 2.74) evidence instantiations on average,

ranging from 0 to 10. This difference was not significant, as there was no main effect of culture on the occurrence of evidence (F (1, 42) = 1.87, p = .18). Table 4.2 shows the mean occurrence of the types of evidence per brochure in the Dutch and the French corpus.

Table 4.2 The mean occurrence of evidence types per Dutch and French brochure (a higher score means a higher occurrence; *SD* between parentheses; means with different superscripts significantly differ from each other within the French or Dutch corpus, alpha level of .05)

type of evidence	Dutch brochure	French brochure	total
causal	1.55^{a} (1.97)	$0.86^{a,b}$ (1.13)	1.20^{a} (1.62)
statistical	0.64^{b} (1.26)	1.18^{a} (1.37)	0.91^{a} (1.33)
anecdotal	$0.18^{b,c}$ (0.50)	$1.14^{a,b}$ (2.21)	$0.66^{a,b}$ (1.66)
expert	0.05^{c} (0.21)	0.27^{b} (0.55)	0.16^{b} (0.43)
total	2.41 (2.74)	3.45 (2.32)	2.93 (2.56)

There appeared to be a main effect of type of evidence (F (3, 40) = 7.74, p < .001, η^2 = .37): the four types of evidence were not employed to the same degree. Causal and statistical evidence were more frequently used than expert evidence, and anecdotal evidence held an intermediate position (see Table 4.2). This main effect was qualified by an interaction effect between culture and evidence type (F (3, 40) = 2.85, p < .05, η^2 = .18). This means that there are indeed cultural differences in the relative use of the four types of evidence in France and the Netherlands (RQ1). Figure 4.1 visualises the absolute number of occurrences of the types of evidence in the Dutch and the French corpus.

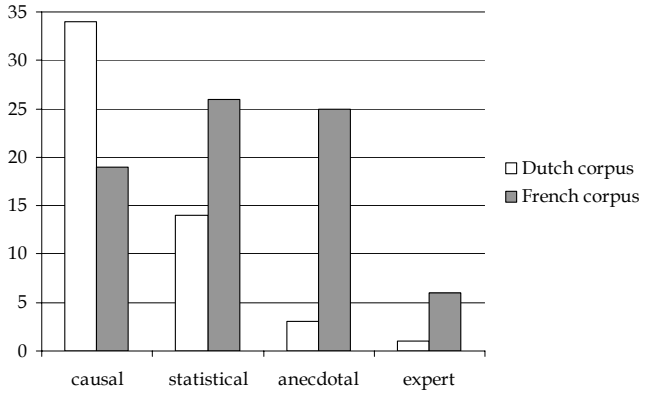

Figure 4.1 The frequency of use of the evidence types in function of culture

The second research question concerned expert evidence (RQ2): is it more frequently used in France than in the Netherlands? This was indeed the case ($t(27.16) = 1.81$, $p < .05$; one-tailed)[5]. Concerning the three other types of evidence, the Dutch and French brochures did not differ significantly in the occurrence of statistical ($t(42) = 1.38$, $p = .18$) or causal evidence ($t(42) = 1.41$, $p = .17$). There was a tendency for a higher occurrence of anecdotal evidence in the French corpus ($t(23.15) = 1.98$, $p = .06$). Within each culture, however, there were clear differences (see Table 4.2). The most frequently used type of evidence in the Dutch corpus was causal evidence. It occurred more often than statistical ($t(21) = 2.31$, $p < .05$), anecdotal ($t(21) = 3.04$, $p < .01$), and expert evidence ($t(21) = 3.49$, $p < .01$). Furthermore, statistical evidence was more frequently used than expert evidence ($t(21) = 2.27$, $p < .05$). In the French corpus, statistical evidence was more frequently used than expert evidence ($t(21) = 2.71$, $p < .05$), and there was a tendency for a higher occurrence of causal than expert evidence ($t(21) = 2.08$, $p = .05$).

As can be seen in Table 4.2, the standard deviations are mostly larger than the mean number of evidence instantiations per brochure. This means that there were large differences between the brochures. An analysis that circumvents this problem is the one Schellens and De Jong (2004) used. I calculated the percentage of brochures in which each type of evidence was used, regardless of the number of times they were used. Expert evidence was shown to occur in more French than Dutch brochures (χ^2 (1) = 12.18, $p < .001$). As shown in Table 4.3, expert evidence was used in 22.7% of the French brochures, and in only 4.5% of the Dutch brochures.

Table 4.3 The percentage of brochures in which the types of evidence occurred

type of evidence	Dutch brochure	French brochure	total
causal	72.7%	59.1%	65.9%
statistical	27.3%	59.1%***	43.2%
anecdotal	13.6%	27.3%*	20.5%
expert	4.5%	22.7%***	13.6%

* $p < .05$, *** $p < .001$

What did the expert evidence that was used look like? With one exception, it was always used in brochures in which the consequences were in the interest of other people, such as brochures about fundraising. Concerning its form, expert evidence was often implicit. The experts did not literally underscore that it was highly

probable that the consequences would occur, but they supported this causal relationship in a more indirect way, as the examples (14), (15), and (16) show.

(14) "Because of their effective method of working [...], the Liliane Fund received an award in 1990 from the United Nations" (*'Care about this child'*)

(15) "Receiving the Nobel Peace Prize in 1977, Amnesty International maintains close contacts with the UN Economic and Social Council [...]" (*'Amnesty International'*)

(16) "In a recent report the General Inspection of Social Affairs (IGAS) observes that the association [SOS Children's Villages] works coherently, rigorously, and with transparency" (*'SOS Villages d'Enfants'*)

The only expert evidence in the Dutch brochures was (14). This piece of evidence was given in support of the claim that a donation to the Liliane Fund would help poor, handicapped children. The fact that the United Nations honoured this fund underscores that it is a credible organisation, and that it is highly probable that the fund will succeed in helping these children with new donations. A similar use of expert evidence was (15) in a French brochure for Amnesty International. The fact that the Nobel Peace Prize committee awarded Amnesty International suggests that donations have led to positive effects, and that they will lead to these effects in the future. In the last example, (16), expert evidence was used to support the claim that a donation to SOS Children's Villages would result in host families for children.

4.4 Conclusion

In this study, the relative use of evidence types in Dutch and French public information brochures was analyzed. There was a cultural difference in the relative use of the four evidence types (RQ1), and expert evidence was more frequently used in the French brochures than in the Dutch brochures (RQ2). It should be noted, however, that expert evidence was the least used evidence type in both corpora. Therefore, a better way to formulate this result would be to say that expert evidence was hardly used in the Dutch corpus, and only moderately in the French corpus. Study 1 also showed cross-cultural equivalence in the way evidence

was used to support claims in pragmatic argumentation. The equivalence of the distinction between four types of evidence in both cultures was also assessed.

The relative use of the four types of evidence was analyzed in this study to investigate the relative expected persuasiveness of these evidence types. An important issue to address is the extent to which the use of evidence types is a good indicator of the persuasiveness the text writers assign these evidence types. The use of evidence may be the result of (implicit) genre conventions in each culture. But even if this were the case, conventions are likely to be related to text writers' preferences about what should be persuasive. Text writers in France and the Netherlands work on the same kinds of persuasive information brochures, and the same types of evidence are available to support their claims. It is highly plausible even that the same arguments are available: the consequences of drinking alcohol, for instance, are the same in France as in the Netherlands. If one assumes that text writers do their best to design the most persuasive brochures, this means that the choice for certain types of evidence is strategic. As a result, Dutch and French text writers can be said to have different expectations about which types of evidence are more persuasive than others. However, the four types of evidence may not be available in the same degree for each occasion where evidence can be used. In reaction to this limitation, an experiment was set up in which participants chose freely between the four types of evidence. Chapter 5 will report on this experiment. At the end of that chapter, Study 1 will be further discussed, and its results will be compared to those of Study 2.

Notes

1. Statistical evidence was more frequently used than expert evidence (χ^2 (1) = 21.49, p < .001), and anecdotal evidence was also more frequently used than expert evidence (χ^2 (1) = 19.80, p < .001).
2. These brochures are preferable to advertisements, which often lack explicit verbal argumentation. On the basis of a corpus of Dutch magazine advertisements, Van Gisbergen, Ketelaar, and Beentjes (2004), for example, showed that in 75% of the ads the size of verbal copy is less than a quarter of the total surface of the ad. Although a restricted space may still contain a large number of words printed in a small font, it is reasonable to assume that brochures contain more text than advertisements.
3. In a few cases, consequences were presented in the form of characteristics. In the Dutch brochure '*I'm looking for a home where I need to be before dark*', readers were asked to become foster parents. Besides the positive consequences of this action, the negative consequence of sometimes being troubled by these children who are regularly in a bad mood was mentioned. This consequence was formulated as a characteristic of the children: these children have their moods, just like any other child. This characteristic can easily be reformulated as a disadvantage: having foster children leads to the

consequence of being sometimes troubled by their bad moods. The distinction between consequences and characteristics therefore did not affect the analysis.

4. An appropriate statistical test to calculate intercoder reliability is Cohen's κ. For this first decision, the proportion of agreement that is expected by chance (PA_E in the κ formula) can not be calculated, as the chance that a coder categorizes a text fragment as evidence can not be predicted. Agreement was therefore computed by means of percent agreement (see Neuendorf, 2002): the number of agreements was divided by the total number of agreements and disagreements.
5. The results for the two research questions were corroborated with a χ^2 test. There was a cultural difference in the relative occurrence of evidence types (χ^2 (3) = 37.00, $p < .001$), and expert evidence was more frequently used in the French than in the Dutch brochures (χ^2 (1) = 3.42, $p < .05$; one-tailed). Next to the interaction effect between culture and the four types of evidence, there was also an interaction effect between culture and the kinds of evidence that Kuhn (1991) distinguishes (F (1, 42) = 14.18, $p < .01$, $\eta^2 = .25$). Anecdotal, statistical, and expert evidence are genuine evidence, and causal evidence is pseudoevidence (see Section 2.2.2). Genuine evidence was more frequently used in the French brochures (M = 2.59, SD = 1.94) than in the Dutch brochures (M = 0.86, SD = 1.42) (t(42) = 3.36, $p < .01$).

Expected persuasiveness: an experimental approach

In Study 1, the expected persuasiveness of evidences types in France and the Netherlands was investigated on the basis of professional text writers' use of evidence types in persuasive information brochures. Study 2 aimed to answer the same general and specific research question on the basis of an experiment. Before presenting this study (5.3 – 5.5), I will first discuss how the expected and actual persuasiveness of evidence types (5.1), and the effect of culture (5.2) was measured in the experiments that were conducted for this dissertation, including Study 2.

5.1 Outline of the experimental approach

In Chapter 3, I presented a review of experimental studies on the relative actual persuasiveness of evidence types. In most of these studies, participants read a long message in which evidence supported an often implicit claim. The persuasiveness of evidence was usually measured on the basis of the perceived probability of the claim supported by evidence. As long messages with evidence may differ in more aspects than only in the type of evidence, it is difficult to attribute the perceived probability to the evidence included. Therefore, I chose to reduce messages to simple claims, such as (1).

(1) Smoking cigarettes leads to a high risk of lung cancer.

Measuring the expected or actual persuasiveness of evidence types then consists of comparing the perceived probabilities of claims that are the same except for the type of evidence that supports them. If group A receives this claim with anecdotal evidence (2), and group B with statistical evidence (3), measuring the relative persuasiveness of anecdotal and statistical evidence consists of comparing the mean perceived probabilities of claim (1) in both groups.

(2) Smoking cigarettes leads to a high risk of lung cancer. John Smith from Rotterdam has smoked for more than 30 years and now suffers from lung cancer.

(3) Smoking cigarettes leads to a high risk of lung cancer. A British study has recently shown that 49% of the 254 smokers they investigated suffer from lung cancer.

Most of the evidence studies discussed in Chapter 3 used a single-message design: one message with one type of evidence. As the message topic may affect the persuasiveness of evidence, it is important to measure the effect of evidence across various topics. When participants are given short texts like (2) and (3), it is possible to provide them with more than one text. As such a multiple message design allows for more reliable generalizations than a single-message design (Jackson & Jacobs, 1983; O'Keefe, 2002), it was also used in the experimental studies in this dissertation. In the following sections, I will discuss the claims that were used (5.1.1) and the operationalizations of the evidence types (5.1.2), and I will explain the way in which the equivalence of the Dutch and French material was ascertained (5.1.3).

5.1.1 Selection of claims

A number of claims were needed for the multiple message design. Not every descriptive, causal claim could be used, since I only accepted claims that were perceived as moderately probable. Highly probable claims, such as (1), do not need evidence, because they are accepted without supporting evidence (thereby obscuring any differences in persuasive effects of evidence). Likewise, improbable claims are not accepted regardless of whether evidence supports these claims or not. In fact, studies on the so-called prior belief effect have demonstrated that the perceived probability of a claim may be influenced more by the prior acceptation of that claim than by the evidence that is provided (see, e.g., Edwards & Smith, 1996; Kunda, 1987; Lord, Ross, & Lepper, 1979). When people are faced with evidence for claims contrary to their initial beliefs, they try to undermine the evidence. In Lord et al., for instance, participants who either supported or were opposed to capital punishment were given the reports of two studies. One study was in favor of capital punishment, the other was against. The results confirmed the prior belief effect: participants rated the evidence that confirmed their own

initial beliefs as more convincing than evidence that disconfirmed their beliefs. This leads Edwards and Smith to conclude that "people are unable to judge the strength of an argument independently of their prior belief in the conclusion [the claim]" (p. 18). Therefore, moderately probable claims are best suited to detect evidence effects.

A pre-test was conducted to select 20 claims that were perceived as moderately probable, and that were rated equally probable in France and the Netherlands. I first created a pool of 40 moderately probable claims and 10 extremely probable or improbable claims with different topics (transportation, work, health, and education)[1]. The extreme claims were included because with 40 moderately probable claims it was thought likely that participants would try to differentiate between the claims, and judge some claims as relatively probable and others as relatively improbable. The 10 extreme claims served as anchor points, and allowed participants to give extreme opinions from time to time. The 50 claims, listed in Appendix D, all consisted of one phrase, indicating an action and a consequence, like example (1). The set of 50 claims was put in a random order, but the different topics were presented alternately, with an extreme claim in each group of 5 claims. Four versions that differed in terms of the order in which the claims were presented were created to neutralize possible order effects[2]. I translated the 50 Dutch claims into French and had this translation checked by the UTN Language Center at the Radboud University Nijmegen. As suggested by Brislin (1980), this French version was back translated into Dutch by another person. We compared the original Dutch version with the back-translated Dutch version, and agreed on four minor lexical changes (one in the Dutch version, three in the French version).

Participants received one of the versions, and were asked to rate the probability of each claim on a 5-point semantic differential (*very improbable – very probable*). In the Netherlands, the pre-test was held at the beginning of a course at the Faculty of Arts at the Radboud University Nijmegen in September 2003. Thirty Dutch students participated, of which 93.3% were female. The mean age was 22.37 (*SD* = 1.22) and ranged from 21 to 25. In France, 28 French students participated at the end of a course at the Faculty of Arts at the University of Montpellier in October 2003. The mean age was 21.71 (*SD* = 1.36), ranging from 20 to 25; 64.3% of them were female. Appendix D lists the 20 claims for which the mean French/Dutch probability rating ranged from 2.15 to 3.50 on a 5-point scale[3], and for which the French and Dutch probability ratings did not significantly differ (alpha level of .05).

5.1.2 Operationalization of evidence types

As was mentioned in Chapter 3, studies on the relative actual persuasiveness of evidence types did not take into account the quality of their evidence instantiations, although the importance of evidence quality has regularly been emphasized (see, e.g., Luchok & McCroskey, 1978; Reinard, 1988; Reynolds & Reynolds, 2002). In my experiments, I used the normative criteria presented in Chapter 2 to create normatively strong instantiations of statistical and expert evidence. As we will see below, it was difficult to create normatively strong anecdotal and causal evidence. For each of the 20 claims, a set of anecdotal, statistical, causal, and expert evidence was created.

Normatively strong statistical evidence should consist of a large sample of cases that is representative for the population in the claim that it supports. The statistical evidence instantiations had a large sample size. Ideally, I should have searched for the Dutch and French population sizes relevant to the 20 claims before determining each normatively correct sample size. Because of the unavailability of this information, however, I made up different sample sizes such as 246, 315, and 429. The statistical evidence instantiations also contained a high percentage of cases in the sample for which the claim holds (e.g, 71%, 77%, and 82%). The representativeness of the sample (the second criterion) could have been manipulated by indicating that the sample was randomly selected (cf. Kassin, 1979). However, as participants received a number of statistical evidence instantiations, the indication of random selection in each instantiation could have made the material unrealistic. Therefore, representativeness was not manipulated.

The operationalization of anecdotal evidence was normatively weak. As it consists of only one case to support the general claim about a whole population, a normatively strong operationalization was impossible. Anecdotal evidence consisted of a short sentence about a person who had experienced the consequence (e.g., better sporting performances) as a result of behavior (e.g., eating basil).

From a normative point of view, causal evidence should provide a mechanism that explains the causal relation between the cause and the effect in the claim. For causal evidence, a strict template was impossible. It consisted either of two causal relations (which is typical of everyday explanations; Hesslow, 1988), or of a causal chain (see Section 4.2.2).

Finally, expert evidence is normatively strong if the expert is credible and reliable, and when the expert's field of expertise corresponds to the field of the claim. Credibility and reliability were operationalized together through the titles of the expert. For the Dutch version, I used the expression 'Prof. dr. [name]'. This

would not have evoked enough credibility and reliability in France, since teachers are also called 'professeur'. I therefore employed the expression that Breton (2003) used when he gave a prototypical French example of a competent expert: 'Prof. [name], director of a research institute affiliated to the CNRS [Centre National de la Recherche Scientifique / French National Center for Scientific Research]'. It is important to stress that the terms 'expert' and 'claim' were never mentioned literally in the Dutch or French material. Relevant fields of expertise for the experts in the 20 claims were chosen in mutual agreement in a session with 17 Dutch students (Radboud University Nijmegen). Dietetics, for instance, was considered as a relevant field of expertise for this claim: 'The consumption of basil in tomato pasta sauce improves sporting performance'.

The templates for the operationalizations of each type of evidence are given in Table 5.1. For purposes of variation, two formulations for statistical and expert evidence were used. The original Dutch and French templates are given in Appendix E.

5.1.3 Cross-cultural equivalence

One of the problems in cultural persuasive effects research is the equivalence of constructs, material, and measurement instruments (e.g., Harkness, Van de Vijver, & Mohler, 2003; Hoeken & Korzilius, 2003). Equivalence of the construct of evidence, and of the distinction into four types of evidence was assured in Study 1. A solution to the problem of the material's equivalence was not only found by back translating the claims and the evidence, but also by searching for Dutch and French equivalents of first names, last names, places, and universities used in the material. In the discussion of the material in each experiment, I will refer to the concrete information that was used. Here, I will discuss how the equivalence was ascertained. I searched for last names that were neither very common, nor very rare, as they both may be considered unrealistic. For France, I randomly picked names from an online database (*www.cartedefrance.tm.fr*) containing the most frequent names between 1966 and 1990. As I did not find any such database for the Netherlands, I selected Dutch names from the material of Hoeken and Hustinx (2003b). These names were neither peculiar nor too ordinary.

Table 5.1 The templates of the four types of evidence, and an example

evidence	template	example
anecdotal	Since [first name + last name] from [place] experienced the cause, the consequence has taken place	Since Peter Johnson from Amsterdam regularly eats tomato pasta sauce with basil, his sporting performance has improved
statistical	A [Dutch/French] study among N [Dutch/French] participants showed that the consequence occurred with X % of them as a result of the cause	A study among 315 Dutch people showed that the sporting performance of 74% of them has improved after having regularly eaten tomato pasta sauce with basil
	The results of a [Dutch/French] study among N [Dutch/French] participants showed that the consequence occurred with X % of them as a result of the cause	The results of a French study among 315 participants showed that the sporting performance of 74% of them has improved after having regularly eaten tomato pasta sauce with basil
causal	-	Basil improves the production of adrenaline, the substance that is necessary for better sporting performance
expert	According to Prof. dr. [name] from the university of [place], an expert / a specialist in the field of [research area], [the claim]	According to Prof. dr. Timmermans from the University of Amsterdam, a specialist in dietetics, the consumption of basil in tomato pasta sauce improves sporting performance
	Prof. dr. [name], an expert / a specialist in the field of [research area] at the university of [place], underscores [the claim]	Professor dr. Timmermans, a specialist in the field of dietetics at the University of Amsterdam, underscores that the consumption of basil in tomato pasta sauce improves sporting performance

Concerning the first names, I used online databases containing the most popular names in the Netherlands (*www.voornamelijk.nl*), and in France (*www.voila.fr*). I selected names from the 1980 list, supposing that popular names from that year are still familiar, yet less frequently given. Julien and Peter, for instance, were the third most popular names in France and the Netherlands respectively (see Appendix F). The German online database *www.gazetteer.de*, which gives current population figures of cities was used to pick cities in France and the Netherlands in the top 20, but outside the top 10. Dutch and French universities comparable in size were selected for the expert evidence. For the Netherlands, size was defined through the number of students of each Dutch university in 2003 (*www.vsnu.nl*). As I could not find a similar French list, I used the number of faculties of the universities in one city as an indicator of size (*www.excite.fr*). The following university cities were not taken into account: cities mentioned in anecdotal evidence, cities where the questionnaires for the experiment were to be filled in, and three universities that have a strong connotation with a particular research area (Eindhoven, Delft, and Toulouse). The choice for large universities had a practical reason: the larger the university, the higher the chances that the 17 research areas of the experts in the experimental material would all be present. I checked the occurrence of these research areas in the universities selected, using the websites of these universities. There was a fit in 90% of the combinations of research area and university. The Dutch evidence operationalizations were translated into French by the UTN Language Center (Radboud University Nijmegen), and back translated by the same person who had done this for the claims. Three minor lexical changes were made in the Dutch version.

5.2 Measuring culture

The material that the Dutch and French participants received in the experiments has just been discussed. If these cultural groups differ on a given dependent variable, no indications are present why the factor culture should account for that result. As cultures differ in many ways (e.g., religion, economy, politics), there are a number of alternative explanations. Since cultural preferences for values were taken as a point of departure that culture may affect the relative persuasiveness of evidence types, a straightforward solution lies in attributing differences to these values. In Section 5.2.1, I will explain how this can be done. Subsequently, I will

present the instrument that was used to measure the effect of culture on the relative persuasiveness of evidence types (5.2.2).

5.2.1 Values and context variables

If expert evidence should appear to be more persuasive to the French than to the Dutch participants, there is no proof that this effect is caused by a difference in power distance. If people live in a culture where the value of a large power distance is important, it will also be important for individuals in this cultures on average, but this does not have to be the case for every individual (e.g., Hofstede, 1980; Smith & Schwartz, 1997; Van de Vijver & Leung, 1997). This leads to the invalid inference problem (see Hoeken & Korzilius, 2003). It is invalid to explain cultural differences on a dependent variable by a supposed cultural difference in the importance that is attached to a value without any knowledge about the value preferences of the participants of the study. Hoeken and Korzilius argue that a straightforward solution consists in measuring cultural values at the level of the individual participants. There is a set of individual values that has been used in experiments on cultural persuasive effects research, namely the Schwartz (1994) value list. Schwartz classified and cross-culturally validated more than 50 values. People's preference for these values is measured by asking to what extent the values are a guiding principle in their lives. A minor problem of measuring values at an individual level is that "there is no apparent direct relationship" between Hofstede's cultural values, and Schwartz' individual values (Hoeken, et al., 2003, p. 203). Researchers have to discover reasonable combinations of values at the two levels on their own.

Peng, Nisbett, and Wong (1997) discuss three major problems of the measurement of values at an individual level. First, the problem of meaning equivalence holds that the meaning of a value, such as equality, may differ from culture to culture. Second, people often assess the importance of a value on the basis of the importance that people around them attach to this value. A Dutch participant may think he finds authorities important because his friends do not. A French participant may think that she does not attach great value to authorities compared to her family. In reality, however, the French participant may attach more importance to authorities than the Dutch participant. Finally, people often attach more importance to the values that are not important to them, and relatively less importance to their own values because these are self-evident. This phenomenon appeared to occur in a study by Hornikx and Starren (2004). Whereas the Dutch culture is characterized by a weak uncertainty avoidance, and France by

a strong uncertainty avoidance (Hofstede, 1980, 2001), the Dutch participants in Hornikx and Starren attached more importance to the Schwartz (1994) values related to strong uncertainty avoidance, and the French participants attached more importance to the Schwartz values related to weak uncertainty avoidance. Peng et al. further emphasized the difficulties of measuring values by demonstrating that different methods of assessing participants' values correlated weakly or not at all. If values are problematic to measure at an individual level, what else can be measured at that level? A solution is provided by the concept of so-called context variables.

The idea of context variables, which originates from cross-cultural psychology (Poortinga & Van de Vijver, 1987), is similar to that of measuring values at an individual level as a solution for the problem of how cultural differences can be explained. Appropriate context variables have to be chosen with respect to the context where an effect of culture is expected, for instance the persuasiveness of expert evidence. The researcher's task therefore consists of carefully selecting or constructing a context variable that is expected to determine the persuasiveness of expert evidence. In contrast, in the value approach the choice of values at an individual level is determined by the expected difference in values at a cultural level, and not so much on the dependent variable under investigation. In conclusion, values are essential for inspiring research questions, and for providing a social-cultural framework of cultural differences and similarities. As the assessment of values at an individual level has methodological problems, context variables other than individual values constitute a better alternative. The next question is to ask what context variable may serve as a predictor for the persuasiveness of expert evidence. This question will be answered below.

5.2.2 The Preference for Expert Information (PEI)

A context variable that is related to the persuasiveness of expert evidence may be the degree to which people accept and appreciate information given by experts. The only scales close to this concept are the social power scales of Swasy (1979), which regroup six interpersonal power types based on French and Raven (1959). Two of these types are important here: informational power (the importance of the information given by a high power person), and expert power (attribution of superior knowledge of a high power person by a low power person). However, the formulations of these items were too abstract (e.g., 'The information provided by A about this situation makes sense'). Inspired by Swasy, I therefore developed six new items for a Preference for Expert Information scale (PEI). These items are:

1. If an expert says it is right, then it is right.
2. Judgments of experts are very important to me.
3. I prefer to base my decisions on the opinion of an expert.
4. The judgment of an expert needs to be believed.
5. I accept that there are people who know more about a topic than me.
6. Knowledge makes experts powerful.

The Dutch Preference for Expert Information scale was translated into French, and back translated by a native French researcher. French participants are expected to have a higher preference for expert information than Dutch participants, and the score on this scale is expected to correlate positively with the persuasiveness of expert evidence.

Study 2

Study 1 demonstrated that Dutch and French text writers differ in their use of evidence types in persuasive information brochures. A limitation of this study was that these writers may not always have had the four types of evidence at their disposal when they wanted to use evidence. Therefore, a second study was conducted, in which participants chose explicitely between the four types of evidence. The research questions of Study 2 are those that Studies 1 and 2 both aimed at:

General RQ1 Are there cultural differences in the relative *expected* persuasiveness of anecdotal, statistical, causal, and expert evidence?

Specific RQ1 Is the *expected* persuasiveness of expert evidence higher for the French than for the Dutch?

In order to answer these questions, an experiment was conducted in which French and Dutch participants were given a series of claims, each supported by four types of evidence. Where possible, these evidence instantiations were normatively strong. The participants were asked to rank these evidence types according to their expected persuasiveness to convince another person. An advantage of this ranking method (cf. Kline, 1971a) is that participants are forced to judge the relative expected persuasiveness of the four types of evidence.

5.3 Method

5.3.1 Material

Eight claims were randomly selected from the pool of 20 claims presented in Section 5.1.1 (see Appendix D). For each of the 8 claims, four types of evidence were constructed in accordance with the templates in Section 5.1.2. In order to avoid having four completely identical anecdotal evidence operationalizations, the order of the main and subordinate clause was sometimes reversed, and instead of *since*, the words *after*, *and*, and *through* were used from time to time. The two operationalizations of statistical and expert evidence were used alternately. The percentages in the statistical evidence ranged from 71% to 84%, and the sample sizes from 138 (for relatively small populations, such as young criminals in prison) to 429 (for relatively large populations, such as boys at secondary school). The names, cities, and universities mentioned in the material are listed in Appendix F.

5.3.2 Participants

In line with the pre-test and the majority of the experimental studies on the persuasiveness of evidence types, students were chosen as participants for this and the following experiments. No student participated in more than one experiment presented in this dissertation. The Dutch participants (n = 88) studied Business Communication Studies at the Radboud University Nijmegen. The French participants studied Applied Foreign Languages (University of Montpellier-II, n = 56) or Linguistics (University of Paris-VIII, n = 30). The programs are very similar. Questionnaires filled in by non-natives were not used for further analyses. The percentage of male participants was higher in France (45.3%) than in the Netherlands (14.8%) (χ^2 (1) = 19.41, p < .001). On average, the Dutch participants were 19.48 (SD = 1.95) years old (range: 18 – 26 years old), and the French participants 22.05 (SD = 2.22) (range: 18 – 30 years old). This difference was significant: t(172) = 8.12, p < .001. Participants' sex and age did not affect the expected persuasiveness of expert evidence, and hardly affected the other types of evidence[4].

5.3.3 Design

Four versions of the material were created, in which the order of the eight claims was identical, but in which the order of the types of evidence was different. A

balanced Latin square design was employed to assign the different orders of the types of evidence to the eight claims, and to the four versions. The scores were pooled across the four versions, as there were no effects of version on the mean ranking of anecdotal evidence (Kruskal-Wallis χ^2 (3) = 6.12, p = .11), statistical evidence (Kruskal-Wallis χ^2 (3) = 3.81, p = .28), causal evidence (Kruskal-Wallis χ^2 (3) = 1.05, p = .79), or expert evidence (Kruskal-Wallis χ^2 (3) = 5.15, p = .16).

5.3.4 Instrumentation

The booklet that participants received was titled 'Convincing another person with arguments'. Participants were asked to rate the four types of evidence in terms of their persuasive power to convince another person of the believability of each of the eight claims. After the eight rankings of the evidence types, participants were asked to fill in on a 5-point Likert scale the degree to which they agreed with six items of the Preference for Expert Information (PEI) scale, and with seven items of the Need for Cognition scale (Cacioppo, Petty, & Kao, 1984). Concerning the PEI scale, it was hypothesized that (a) French participants would score higher than the Dutch participants, and that (b) the higher participants scored on PEI, the higher their expected persuasiveness of expert evidence would be. The first four items of the PEI scale proved to be reliable, both for the French participants (α = .73), and for the Dutch participants (α = .75). Van de Vijver and Leung (1997, p. 60) suggested a formula to test the equality of reliability coefficients in two groups. There proved not to be a cultural difference in the coefficients in the present study (p > .05), suggesting that the scale is not culturally unreliable.

The second scale, the Need for Cognition (NFC), measures the degree to which people spontaneously engage in thinking and enjoy it (Cacioppo, et al., 1984). In the ELM, people who score high on NFC are said to be more inclined to take the central route to persuasion, and to differentiate between strong and weak arguments (Petty & Cacioppo, 1986). The NFC scale was included in the questionnaire as a check on the route that Dutch and French participants took (cf. Section 3.3.4). I used the seven items of the French version from Falcy (1997), and the corresponding items of the Dutch version as used by Pieters, Verplanken, and Modde (1987). The scale was reliable for the French participants (α = .74), but not for the Dutch (α = .58). Therefore, I will not refer to this scale in the presentation of the results. After these items, there were eight control questions about the expertise of the eight experts that each participant read about in the claims. For each of the eight experts, participants had to indicate the degree to which they agreed with a standpoint on a 5-point Likert scale, such as: "Professor Timmermans is a

researcher in the field of retail marketing at the University of Rotterdam. In that capacity, he has enough expertise to make a judgment about the relation between slow music in supermarkets and their turnover". The questionnaire ended with questions about participants' age, sex, nationality, and current education.

5.3.5 Procedure

The questionnaires were filled in at one Dutch university and two French universities in January – March 2004. As for the other experiments in this dissertation, questionnaires were not filled in at the universities mentioned in the material. The questionnaires were distributed at the beginning of a lecture. The students received no reward for their participation. After the questionnaires had been collected, the real research purpose was revealed, and participants were thanked for their cooperation. The whole procedure took about 15 minutes. There were no disturbances during the experiment.

5.3.6 Statistical tests

A Friedman test was used to investigate whether the mean ranking of the four types of evidence differed. Next, a Wilcoxon signed rank test was employed to test for each culture which types of evidences differed from each other in their mean ranking. Cultural differences in the mean rankings of the evidence types were measured with a nonparametric Mann-Whitney test, and cultural differences in the number of first rankings with an independent t-test.

5.4 Results

Cross-cultural methodology suggests checking whether participants with different cultural backgrounds have the same scale use (Van de Vijver & Leung, 1997). One of the types of scale bias where European differences have been reported is extreme response style, "the tendency to endorse extreme response categories on a rating scale (e.g., the 1 and/or the 5 on a 5-point scale), regardless of content" (Van Herk, Poortinga, & Verhallen, 2004, p. 347). The study of Johnson, Kulesa, Cho, and Shavitt (2005) suggests that extreme response bias is likely for the French culture, and not for the Dutch culture. In fact, they demonstrated for a large number of cultures that there are positive correlations between the level of power distance of a culture, and the number of extreme responses on 5-point Likert

scales. I therefore tested whether the Dutch and French differed in their use of the two endpoints 1 and 5 on the context variables (11 items), and on the perceived expertise of the experts (8 items), using the Bachman and O'Malley (1984) index (number of items with extreme responses divided by the total number of items). The French participants scored more extremely than the Dutch (see Table 5.2).

Table 5.2 Extreme response indices for Dutch and French participants and t-test for cultural differences (*SD* between parentheses)

variable	Dutch index	French index	t-test
context variables	.13 (0.16)	.24 (0.20)	$t(172) = 4.04$, $p < .001$
expertise of experts	.04 (0.15)	.10 (0.20)	$t(154.89) = 2.37$, $p < .05$

I therefore standardized the scores on the context variables, and on the perceived expertise of the experts[5]. The analyses below of the PEI scale and the perceived expertise of the experts will concern only the standardized data. I checked whether there were cultural differences on the PEI scale, and on the expertise of the experts. First, it was found that, contrary to expectation, the Dutch ($M = 2.51$, $SD = 0.69$) and French participants ($M = 2.52$, $SD = 0.77$) scored equally on the PEI scale ($t(172) = 0.95$, $p = .35$). Next, I tested whether the experts were indeed considered as persons with high expertise. This was the case, as the mean perceived expertise for the French ($M = 3.25$, $SD = 0.60$) and the Dutch participants ($M = 3.52$, $SD = 0.54$) scored above the midpoint (3.00) of the scale (French: $t(85) = 3.82$, $p < .001$; Dutch: $t(87) = 9.17$, $p < .001$). For each expert there was no cultural difference in his or her perceived expertise (p's > .14).

This experiment was conducted to investigate whether there are cultural differences in the expected persuasiveness of the four types of evidence, and of expert evidence in particular. The types of evidence were ranked from lowest to highest expected persuasiveness. There was a difference in the mean rankings of the four types of evidence (Friedman χ^2 (3) = 284.63, $p < .001$). The mean rankings were then compared for each individual culture. The Dutch participants had the highest preference for statistical evidence, followed by expert, causal, and anecdotal evidence (see Table 5.3).

Table 5.3 Mean ranking in function of evidence type and culture (*SD* between parentheses; a smaller number indicates higher expected persuasiveness; different superscripts refer to significant differences within-cultures, alpha level of .001)

evidence type	Dutch participants (n = 88)	French participants (n = 86)
statistical	1.25[a] (0.44)	1.55[a] (0.67)
expert	2.33[b] (0.62)	2.33[b] (0.69)
causal	2.68[c] (0.67)	2.71[c] (0.78)
anecdotal	3.73[d] (0.49)	3.40[d] (0.70)

The same pattern of rankings of the four types of evidence was found for the French participants. This means that there was no cultural difference in the relative expected persuasiveness of the four types of evidence (general RQ1). Still, there may be differences in the Dutch and French mean rankings of the individual types of evidence. Statistical evidence had a higher ranking for the Dutch participants (z = 3.26, p < .01), whereas the French ranking of anecdotal evidence was higher than the Dutch (z = 4.05, p < .001). Causal evidence (z = 0.48, p = .63) and expert evidence (z = 0.26, p = .79) were not significantly differently ranked by the two cultural groups. Concerning the specific RQ1 about expert evidence, this means that there was no cultural difference in the expected persuasiveness of expert evidence.

In order to explore in more detail participants' rankings of expert evidence for each of the eight claims, I tested whether there were cultural differences in the number of times that the participants put expert evidence in first position in the ranking (that is, expert evidence is expected to be most persuasive). As Table 5.4 shows, statistical evidence was put in first position in the majority of the eight claims.

Table 5.4 Mean number of times that a type of evidence was put in first position in the ranking (*SD* between parentheses)

evidence type	Dutch participants (n = 88)	French participants (n = 86)
statistical	6.47** (2.52)	5.02 (3.01)
causal	0.75 (1.81)	1.24 (2.06)
expert	0.70 (1.73)	1.26* (2.14)
anecdotal	0.04 (0.28)	0.47** (1.50)

* p < .05, ** p < .01

French participants appeared to put expert evidence in first position more often than the Dutch participants ($t(163.28) = 1.87$, $p < .05$, one-tailed). Each French participant put it on average 1.26 times in first position, whereas the Dutch participants did this only 0.70 times. The number of top rankings for expert evidence was not affected by participants' score on the PEI scale (for both cultures: $F < 1$).

Finally, the effect of PEI on the expected persuasiveness of expert evidence was tested. A positive correlation was hypothesized between a person's score on the PEI scale and the mean ranking of expert evidence. For the French participants, this was the case, but the correlation was not particularly strong (r (86) = .26, $p <$ .05). For the Dutch participants, surprisingly, the correlation was not significant (r (88) = .16, p = .13).

5.5 Conclusion

Study 2 investigated cultural differences in the expected persuasiveness of the four types of evidence, and of expert evidence in particular, on the basis of the participants' preference for evidence types to persuade other people. Concerning the general RQ1, there was no cultural difference in the relative rankings of the four types of evidence. Two more subtle results were indications for cultural differences: the cultural background of the participants affected the mean rankings of statistical and anecdotal evidence, and the number of times that three of the four types of evidence were put in first position in the ranking. Concerning the specific RQ1, expert evidence was not ranked higher in France than in the Netherlands. There was one indication, however, for a cultural difference in the expected persuasiveness of expert evidence: French participants put expert evidence in first position more frequently than the Dutch participants.

The scores on the PEI scale were expected to differ between the two cultural groups, and were expected to correlate with the expected persuasiveness of expert evidence. Contrary to these expectations, the two cultural groups scored equally on this scale, and the scale scores correlated positively with the expected persuasiveness of expert evidence for the French participants only, but not for the Dutch participants.

5.6 Conclusion and discussion of Studies 1 and 2

Studies 1 and 2 addressed the research questions whether there are cultural differences in the relative *expected* persuasiveness of anecdotal, statistical, causal, and expert evidence (general RQ1), and of expert evidence in particular (specific RQ1). In Study 1, expected persuasiveness was measured on the basis of the *use* of evidence in persuasive information brochures, because – as I argued in Section 4.4 – this use is an indicator for expected persuasiveness. Studies 1 and 2 both demonstrated with different methodologies that there are cultural differences between France and the Netherlands in the relative expected persuasiveness of the evidence types, and of expert evidence in particular. However, the differences were rather small. Before I will discuss how Studies 1 and 2 relate to each other, I will first discuss them separately.

In Study 1, there was only a moderate use of expert evidence in the French corpus. A first explanation for this is that expert evidence may be considered less relevant than the other types of evidence, as it usually does not provide new information. A second explanation may lie in the suggestion that the brochures' senders do not like to use expert evidence, because they do not feel the need for support of other experts to defend their claims. French text writers in particular may feel that their own authority makes support from other experts redundant. Conversely, senders who are not sure whether the readers will accept their claims, such as fundraising companies, may be more inclined to use expert evidence. Study 1 supported this: all expert evidence instantiations except one came from fundraising brochures. The low use of expert evidence corresponds with the finding of Levasseur and Dean's (1996) study on the use of evidence in presidential debates. In Schellens and De Jong (2004), argumentation from authority, which is related to expert evidence, occurred in 30% of the brochures. This percentage is much higher than the 4.5% of the brochures in which expert evidence was used in Study 1, but the 30% would have been lower with a more narrow definition of authority. In Schellens and De Jong, the argument from authority also included sources such as celebrities, and unnamed scientists. Causal argumentation was very frequently used in the corpus in Schellens and De Jong. It occurred in 75% of brochures, similar to the 72.7% in which causal evidence was used in the brochures in Study 1. These high percentages are quite remarkable, as causal evidence is not genuine, but pseudoevidence (Kuhn, 1991).

In Study 2, students were asked to rank evidence types in terms of their expected persuasiveness. With this design, the limitation of the possible

unavailability of the four evidence types in Study 1 was combatted. A drawback of the ranking method is that differences in relative expected persuasiveness may sometimes be artificial, because participants are not able to rank two types of evidence as equally persuasive. Study 2 can not be related to similar studies, as it is the first to investigate the expected persuasiveness of four types of evidence on the basis of a ranking technique.

Finally, the number of differences in the approaches used in Studies 1 and 2 limit to a certain extent the comparability of the studies' findings: the concrete brochures versus the more abstract claims with evidence, the text writers versus the students as participants, the use of evidence types versus a ranking of evidence types, and the target audiences of the brochures versus the other persons that the students had to imagine. Such a comparison nevertheless provides some interesting insight: the persuasiveness of expert evidence as expected by students appears relatively higher than the actual use of expert evidence in persuasive brochures, and the use of anecdotal evidence appears relatively higher than the degree to which students expect it to be persuasive. A problem that the two studies share is that it is difficult to empirically explain the (small) cultural differences that were found. In Study 1, the power distance of the text writers was not measured, and in Study 2, the context variable PEI could not account for the cultural differences. I will come back to the issue of the measurement of cultural effects in the final chapter of this dissertation.

Notes

1. The 40 moderate claims consisted of 4 claims from Hornikx and Hoeken (2003) that were judged between 2.76 and 3.46 on a 5-point scale, 4 claims from Hoeken and Hustinx (2003b) that were judged between 2.90 and 3.40, 6 claims from Hoeken and Hustinx (2003a) that were judged between 2.81 and 3.40, and 26 newly developed claims. Next, I searched for claims that were rated as extremely (im)probable by both the Dutch and the French participants in Hornikx and Hoeken. Finally, I added claims from Hoeken and Hustinx (2003a, 2003b) that were judged as extreme by their Dutch participants, to come to a total of 10 claims. The 5 extremely improbable claims scored on average 1.94, and the 5 extremely probable claims 4.05.
2. Version 1: claims 1 to 50; version 2: claims 50 to 1; version 3: claims 1 to 25, and 50 to 26; version 4: claims 50 to 26, and 1 to 25. There was no effect of version on the acceptability of the claims for the Dutch participants, or for the French participants. For the comparison of the scores on the 50 claims in each version, the use of a multivariate F test was not appropriate, as the claims hardly correlated. Instead, univariate F tests were performed for each culture with a corrected alpha level (Bonferroni correction: p = 0.05 / 50 comparisons = .001). With this alpha level, none of the 50 claims was rated differently in the four versions in both cultures.

3. As the range 2.50 – 3.50 did not include 20 claims, the range was expanded at the lower bound, where effects of evidence are more likely to occur than for claims scoring over 3.50.
4. Participants' sex did not affect the mean ranking of statistical (z = 0.75, p = .45), causal (z = 1.67, p = .10), or expert evidence (z = 0.03, p = .98), but it did have an effect on anecdotal evidence (z = 3.78, p < .001). Men (M = 3.32, SD = 0.75) ranked anecdotal evidence higher than women (M = 3.67, SD = 0.53). The sex of the participants did not affect the number of times that anecdotal (t(62.12) = 1.70, p = .09), statistical (t(172) = 0.30, p = .77), causal (t(172) = 0.74, p = .46), and expert evidence (t(172) = 0.02, p = .99) were ranked as best, nor the score on PEI (t(172) = 0.30, p = .76). Age did not significantly correlate with the mean ranking of anecdotal (r (174) = .15, p = .05), causal (r (174) = .03, p = .74), or expert evidence (r (174) = .01, p = .88), but it did with statistical evidence (r (174) = -.21, p < .01). Age (as covariate) did not affect the number of times that anecdotal (F (1, 172) = 2.13, p = .15) causal (F (1, 172) = 2.98, p = .09), or expert evidence (F (1, 172) = 3.09, p = .08) were put on the first place, but it did affect the number of first positions of statistical evidence (F (1, 172) = 9.26, p < .01). On the basis of the median-split technique, two groups were created. The participants aged 20 years and younger (M = 6.55, SD = 2.41) put statistical evidence in first position more frequently than the participants aged 21 years and older (M = 4.95, SD = 3.06). Finally, age had no effect on the score on PEI (F (1, 172) = 2.72, p = .10)
5. For each culture, the scores on the context variables were standardized by subtracting from each original score the mean score of the 11 items of the context variables, and by dividing the remaining number by the standard deviation of the mean of these 11 items. The same procedure was followed for the expertise of the eight experts.

Actual persuasiveness: an experimental approach

Studies 1 and 2 showed cultural differences in the relative expected persuasiveness of evidence types in general, and of expert evidence in particular. If culture plays a role in the relative *expected* persuasiveness of evidence types, it is possible that it also affects the relative *actual* persuasiveness of these evidence types. Therefore, the second general and specific research questions of this dissertation were:

General RQ2 Are there cultural differences in the relative *actual* persuasiveness of anecdotal, statistical, causal, and expert evidence?

Specific RQ2 Is the *actual* persuasiveness of expert evidence higher for the French than for the Dutch?

In this chapter, I will present three studies. Study 3 aimed to answer both research questions with (normatively strong instantiations of) the four types of evidence. Study 4 focused on the specific research question by examining normatively strong and normatively weak statistical and expert evidence. The chapter will end with Study 5, which was conducted to test an explanation given for a finding in Study 4.

Study 3

If French and Dutch people differ in how persuasive they expect evidence types to be (Studies 1 and 2), the *actual* persuasiveness of these evidence types may also differ culturally. Therefore, Study 3 was set up to answer the general and specific RQ2 mentioned above. An experiment was conducted in which Dutch and French participants received 20 claims supported by different types evidence. Where possible, these evidence instantiations were normatively strong. Actual persuasiveness of evidence was measured on the basis of the perceived probability of the claims supported by evidence.

6.1 Method

6.1.1 Material

The 20 claims selected in the pre-test (Section 5.1.1) were all used in this experiment. Their order in the questionnaire was random, but also respected a high variation in the claims' probability as measured in the pre-test, and the claims' content. Each participant received four instantiations of each type of evidence. The order of the evidence types differed between the five versions of the material, but participants received the same names, cities, and universities in exactly the same order (see Appendix G). For anecdotal evidence, a maximum of two instantiations started with *Since*. There were four fixed combinations of a first name, last name, and city (e.g., Marcel Elling from Haarlemmermeer). For statistical evidence, the two different templates were each used twice. The sample sizes[1] (138, 246, 315, and 429) were assigned to the claims in accordance with the assumed population size, and the percentages (71%, 74%, 77%, and 82%) were randomly assigned to the claims. There was no fixed relation between the numbers and the percentages. For expert evidence too, the two templates were each used twice. One of the four combinations of an expert name and a university was, for instance, Professor Giraud from the University of Bordeaux.

The length of the operationalizations of the four types of evidence was compared. There was a main effect of culture on evidence length (F (1, 152) = 30.14, $p < .001$, $\eta^2 = .17$): French evidence (28.81 words) was longer than Dutch evidence (24.94 words). Table 6.1 shows the mean length of the types of evidence.

Table 6.1 Evidence length in number of words in function of language and type (*SD* between parentheses; different superscripts refer to significant differences within-cultures, alpha level of .001)

type of evidence	Dutch evidence	French evidence
statistical	28.85[a] (3.87)	33.35[a] (5.43)
expert	28.70[a] (3.84)	35.50[a] (5.54)
causal	22.65[b] (4.51)	24.90[b] (4.19)
anecdotal	19.55[c] (3.38)	21.50[b] (4.49)
total	24.94 (5.56)	28.81 (7.58)

There was also a main effect of evidence type (F (3, 152) = 63.63, p < .001, η^2 = .56). As this main effect was nearly qualified by an interaction effect with culture (F (3, 152) = 50.98, p = .06), analyses were run for each culture. These analyses showed for the French material that anecdotal and causal evidence were shorter than statistical and expert evidence, and for the Dutch material that anecdotal evidence was shorter than causal evidence, which was shorter than expert and statistical evidence.

6.1.2 Participants

The participants were students from universities with a large geographical variation in each country in order to better represent the Dutch and the French culture (cf. Smith, 2004). The Dutch cities were Amsterdam (n = 42; two groups), Delft (n = 23), Groningen (n = 46), Nijmegen (n = 108; three groups), and Tilburg (n = 86; two groups). The French cities were Angers (n = 30), Paris (n = 107; two groups), Roubaix (n = 34), Toulouse (n = 26), and Tours (n = 98; two groups). Questionnaires were excluded from the analyses if they had been filled in by non-natives (about 40 in France, and 20 in the Netherlands) and/or had not been filled in seriously (e.g., all the 20 claims were rated as extremely improbable). Only students older than 16 and younger than 31 were taken into account. In this dataset (France: n = 333; Netherlands: n = 310) there were differences in the number of participants in each version. As the lowest number of participants in each version was 59 in France and 61 in the Netherlands, participants were randomly rejected in versions that contained more than 59 or 61 participants respectively. In total, 305 Dutch students (61 in each version) and 295 French students (59 in each version) participated. Appendix G gives an overview of the number of participants for each version, and for each group in France and in the Netherlands[2].

The Dutch participants studied at the Faculty of Arts (89%; Business Communication Studies, Communication Sciences, Dutch language and culture), Social Sciences (4%), or a technical discipline (7%). The French participants studied Arts (63%; Applied Foreign Languages, English language and culture), or Social Sciences (36%; Sociology, Psychology). The percentage of female participants was higher in France (86.8%) than in the Netherlands (77.4%) (χ^2 (1) = 8.98, p < .01), but there was no interaction effect between participants' sex and the relative persuasiveness of the types of evidence (F < 1). On average, the Dutch participants were 20.98 (SD = 2.36) years old, and the French participants 20.75 (SD = 1.96). This difference was not significant: t(585.16) = 1.33, p = .19.

6.1.3 Design

In each of the five versions, four claims were supported by statistical evidence, four by anecdotal evidence, four by causal evidence, four by expert evidence, and four were not supported by evidence (cf. Hoeken & Hustinx, 2003b). A balanced Latin square design was used to determine the order in which the types of evidence were used as support for the claims. Each claim was supported by another type of evidence in each of the versions. Claim 8, for instance, was supported by statistical evidence in version 2, by anecdotal evidence in version 3, by causal evidence in version 4, and by expert evidence in version 5, but was not supported by evidence in version 1.

6.1.4 Instrumentation

The booklet that participants received was titled 'Opinions on social issues'. A short instruction explained how to fill in the questionnaire, which was said to have been constructed by Dutch and French universities. After the instruction, participants read 20 pairs of claims with a type of evidence. Each pair was followed by a repetition of the claim, and by a 5-point semantic differential (*very improbable* – *very probable*) on which participants were asked to judge the probability of each claim. After these 20 judgments, participants received a number of items of three context variables for which they had to indicate their agreement on a 5-point Likert scale. These variables were included in order to explain possible cultural effects. Participants were given seven items of the Need for Cognition (Cacioppo, Petty, & Kao, 1984), eight items of the Preference for Numerical Information (Viswanathan, 1993), and six items of the Preference for Expert Information scale. The three scales were found not culturally unreliable with Van de Vijver and Leung's (1997) formula to test the equality of reliability coefficients in two groups (p's > .05).

For the PEI scale, the first four items proved to be reliable (cf. Study 2), both for the French participants (α = .77), and for the Dutch (α = .75). Based on the findings in Study 2, the French participants were not expected to score higher on this scale than the Dutch. However, I hypothesized that the score on PEI would correlate positively with the actual persuasiveness of expert evidence. For the three other types of evidence, I did not have any indications for cultural differences. Nevertheless, context variables were still selected, as they might provide insight into personal characteristics that affect the persuasiveness of those types of evidence. I hypothesized that the persuasiveness of causal evidence was related to

the score on NFC (Cacioppo, et al., 1984)[3]. People who score high (as opposed to low) on NFC could engage more in spontaneously thinking about the causal relations in the claim, and could therefore find causal evidence more persuasive. NFC was also included in the questionnaire as a check on the route to persuasion that Dutch and French participants take (central or peripheral). Given that the four types of evidence are not equally persuasive (see review in Chapter 3), it was hypothesized that people who scored high on NFC would perceive larger differences between the persuasiveness of evidence types than people who scored low on NFC. Finally, NFC may be related to anecdotal evidence. People who engage in thinking to a lesser degree may be less sensitive to the normative weakness of anecdotal evidence, and may – as a result – be more persuaded by anecdotal evidence than people who score high on NFC. The NFC scale was reliable: $\alpha = .74$ for the French participants, and $\alpha = .67$ for the Dutch.

The persuasive power of statistical evidence may be related to the degree to which people appreciate numerical information. The Preference for Numerical Information (PNI) scale measures the "proclivity toward using numerical information and engaging in thinking involving numerical information" (Viswanathan, 1993, p. 741). This scale measures people's attitude toward statistical information. It was hypothesized that the score on this scale would affect the persuasiveness of statistical evidence. I used the validated small 8-item version of PNI, which proved to be reliable ($\alpha = .86$ for the French participants, and $\alpha = .85$ for the Dutch). The original English items were translated into French by a French psychologist, and into Dutch by a Dutch communication scholar. A French native and I compared the three versions on equivalence. After these context variables, there were four questions about the perceived expertise of the four experts that each participant read about (cf. Study 2), and questions about participants' age, sex, nationality, and current education.

6.1.5 Procedure

The questionnaires were filled in at several Dutch and French universities in October – November 2003. The questionnaire was introduced and distributed at the beginning or at the end of tutorials or lectures[4]. After the questionnaires had been collected, the real research purpose was revealed, and participants were thanked for their cooperation. The students received no reward for their participation. The whole procedure took about 10 to 12 minutes in the smaller groups, and 13 to 15 minutes in the larger groups. There were no disturbances during the experiment.

6.1.6 Statistical tests

The persuasiveness of evidence was computed as a difference score: the judgment of a claim with evidence minus the judgment of the same claim without evidence. Thus, claims without evidence served as a baseline. Each participant judged four instantiations of, for instance, anecdotal evidence. The four difference scores for anecdotal evidence were averaged into a score indicating the persuasiveness of anecdotal evidence for that participant. As the procedure for statistical, causal, and expert evidence was the same, each participant finally had a difference score for each type of evidence. The research question about cultural differences in the relative persuasiveness of evidence types was evaluated by means of a 2 (culture) x 4 (type of evidence) analysis of variance, where culture was a between-subjects factor and type of evidence a within-subjects factor. Concerning the other research question, cultural differences in the persuasiveness of expert evidence were investigated in two ways. First, the persuasiveness of expert evidence in the two cultural groups was directly compared with an independent t-test. Second, the relative persuasiveness of expert evidence was investigated by comparing its persuasiveness with that of the other types of evidence. In addition to these analyses by participants, analyses by stimuli were also conducted in order to test whether the effects of culture and evidence type could be generalized over the different claims.

A within-subjects design carries the risk of a carry over effect (see Elmes, Kantowitz, & Roediger, 1992): the participants' judgments of claims in the second part of the booklet may be influenced by their judgments of claims in the first part. Therefore, a 2 (first judgment, last judgment) x 4 (type of evidence) analysis of variance with repeated measures was conducted. If participants had learned to perceive differences between the types of evidence, there should have been a significant interaction effect between the two factors. The interaction, however, was not significant: F (3, 594) = 1.03, p = .38.

6.2 Results

As in Study 2, I tested whether the Dutch and French participants differed in their use of scale extremities. Table 6.2 shows the Bachman and O'Malley (1984) indices for the judgments of the 20 claims[5] with evidence, the three context variables (19 items), and the expertise of the four experts. As the French participants always

scored more extremely than the Dutch, the scores on these variables were standardized (see, for the procedure, footnote 5 in Chapter 5). The analyses below will concern only the standardized data, unless indicated otherwise.

Table 6.2 Extreme response indices for Dutch and French participants and t-test for cultural differences (*SD* between parentheses)

variable	Dutch index	French index	t-test
claims	.22 (0.16)	.31 (0.16)	$t(598) = 6.33$, $p < .001$
context variables	.22 (0.16)	.32 (0.18)	$t(598) = 7.22$, $p < .001$
expertise of experts	.05 (0.16)	.12 (0.23)	$t(521.11) = 4.10$, $p < .001$

I tested whether there were cultural differences in the perception of the expertise of the 20 experts used in the experimental material (with a corrected alpha level, see footnote 5 in this chapter). The perceived expertise of the 20 experts did not differ between the two cultural groups. The mean perceived expertise was 3.29 (*SD* = 0.64) for the Dutch participants, and 3.22 (*SD* = 0.74) for the French (non-significant difference: $t(598) = 1.36$, $p = .17$; t-test with raw data). There was one expert for which the expertise was significantly lower than the midpoint of the scale (the fourth Dutch expert in version 1: $M = 2.61$, $SD = 0.90$). Below, the results will be presented that are relevant to the research questions (6.2.1), and that concern the context variables (6.2.2).

6.2.1 Research questions

Table 6.3 shows the persuasiveness of the types of evidence for the Dutch and the French participants. As these scores are based on standardized data, they do not reflect the absolute difference scores on a 5-point scale. However, a higher score still means higher persuasiveness, and a negative score means that using evidence is less persuasive than no evidence.

First of all, there was a main effect of type of evidence on persuasiveness (F_1 (3, 596) = 35.05, $p < .001$, $\eta^2 = .15$; F_2 (3, 17) = 18.37, $p < .001$, $\eta^2 = .76$). This means that the four types of evidence were not equally persuasive: in fact, statistical evidence was more persuasive than causal evidence, with expert evidence taking an intermediate position, and anecdotal evidence was the least persuasive (see Table 6.3). Next, there was a main effect of culture with an analysis by participants (F_1 (1, 598) = 6.28, $p < .05$, $\eta^2 = .01$) – suggesting more impact of evidence for the Dutch (*M*

= 0.57, *SD* = 1.07) than for the French participants (*M* = 0.35, *SD* = 1.09) – but only a tendency for this main effect with an analysis by stimuli (F_2 (1, 19) = 3.38, p = .08).

Table 6.3 Persuasiveness of evidence in function of culture and type (standardized data; *SD* between parentheses; different superscripts refer to significant differences within-cultures, alpha level of .01)

type of evidence	Dutch (*n* = 305)	French (*n* = 295)	total (*N* = 600)
statistical	1.04^{a} (1.75)	$0.48^{a,b}$ (1.53)	0.77^{a} (1.67)
expert	0.66^{b} (1.62)	0.60^{a} (1.64)	$0.63^{a,b}$ (1.63)
causal	0.64^{b} (1.61)	0.31^{b} (1.69)	0.48^{b} (1.66)
anecdotal	-0.08^{c} (1.68)	0.00^{c} (1.54)	-0.04^{c} (1.61)

These main effects were qualified by an interaction effect between type of evidence and culture (F_1 (3, 596) = 6.33, p < .001, η^2 = .03; F_2 (3, 17) = 4.56, p < .05, η^2 = .45)[6]. This means that there are cultural differences in the relative actual persuasiveness of the four types of evidence (general RQ2). In particular, statistical evidence was more persuasive to the Dutch participants (*M* = 1.04) than to the French participants (*M* = 0.48) (t_1 (598) = 4.18, p < .001; t_2 (38) = 3.24, p < .01), and causal evidence too was more persuasive to the Dutch participants (*M* = 0.64) than to the French participants (*M* = 0.31) with an analysis by participants (t_1 (598) = 2.48, p < .05), but not with an analysis by stimuli (t_2 (38) = 1.54, p = .13).

In an absolute way, expert evidence was not significantly more persuasive to the French participants (*M* = 0.60) than to the Dutch participants (*M* = 0.66) (t_1 (598) = 0.50, p = .62; t_2 (38) = 0.34, p = .74) (specific RQ2). However, when the persuasiveness of expert evidence was compared to that of the other types of evidence, expert evidence appeared to be relatively more persuasive in France than in the Netherlands (see Figure 6.1). More specifically, there was an interaction effect between culture and statistical / expert evidence (F_1 (1, 598) = 9.92, p < .01, η^2 = .02; F_2 (1, 38) = 5.38, p < .05, η^2 = .12). Concerning the Dutch participants, anecdotal evidence (*M* = -0.08) was the least persuasive, and statistical evidence (*M* = 1.04) was the most persuasive, with expert (*M* = 0.66) and causal evidence (*M* = 0.64) taking an intermediate position. For the French participants too, anecdotal evidence was the least convincing type of evidence (*M* = 0.00). Expert evidence (*M* = 0.60) was more persuasive than causal evidence (*M* = 0.31), and statistical evidence (*M* = 0.48) took an intermediate position.

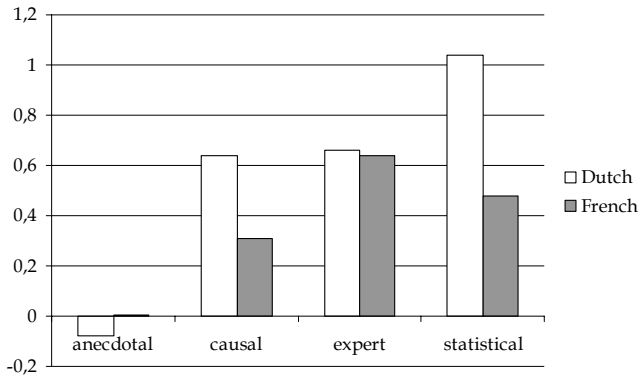

Figure 6.1 The actual persuasiveness of evidence types in the Dutch and the French culture (based on Table 6.3)

The four types of evidence differed in their word length (see Section 6.1.1). This length operationalization may have contributed to the difference in persuasiveness of the types of evidence. Statistical evidence might, for instance, be persuasive because of its length, and not because of its content. In order to determine the effect of word length, I ran repeated measures ANCOVA's with length as covariate (unstandardized data). For the Dutch material, evidence length did not affect the relative persuasiveness of evidence types ($F < 1$), and the main effect of type of evidence was still present (F (3, 75) = 4.70, $p < .01$, $\eta^2 = .16$). For the French material, there was no effect for length as covariate ($F < 1$), but there was no longer a main effect of type of evidence (F (3, 75) = 1.76, $p = .16$). Therefore, evidence length seems to have played a role for the French participants, but – in the absence of an effect for length – it is not a strong alternative explanation for the difference in persuasiveness of the types of evidence.

6.2.2 Context variables

Context variables were selected in order to explain possible cultural differences in the persuasiveness of the types of evidence. As Table 6.4 shows, there were no cultural differences in the scores on the three context variables. As in Study 2, French and Dutch participants scored equally on the PEI scale.

Table 6.4 Scores on the context variables and t-test for cultural differences (*SD* between parentheses; means with raw data, test with standardized data)

variable	Dutch participants	French participants	t-test
NFC	3.16 (0.61)	3.35 (0.72)	$t(597) = 0.87$, $p = .38$
PEI	2.41 (0.74)	2.40 (0.77)	$t(597) = 0.87$, $p = .38$
PNI	2.82 (0.77)	2.90 (0.90)	$t(597) = 0.22$, $p = .83$

The expectation concerning PEI was that the higher participants scored, the more they would be persuaded by expert evidence. PEI as covariate had a significant effect on the persuasiveness of expert evidence for the Dutch participants (F (1, 300) = 20.00, $p < .001$, $\eta^2 = .06$). There was a significant, positive correlation between the two variables: r (304) = .24, $p < .001$. For the French participants the PEI score as covariate had no effect on the persuasiveness of expert evidence (F (1, 291) = 1.83, $p = .18$). The degree to which French participants found expert evidence persuasive was independent of their personal preference for expert information.

The scores on NFC, PNI, or PEI could not account for differences that were found, because Dutch and French participants did not differ in their scores on these scales. Still, the context variables may provide insight into the way personal characteristics affect the persuasiveness of the types of evidence. As hypothesized in Section 6.1.4, the score on NFC was related to the overall difference in persuasiveness of the types of evidence. On the basis of the median-split technique, two groups were created: high and low NFC. Participants who scored high on NFC perceived larger differences in the persuasiveness of statistical, causal, and expert evidence (F (2, 271) = 7.78, $p < .01$, $\eta^2 = .05$) than participants who scored low on NFC (F (2, 324) = 1.34, $p = .26$). This only applied to these three types of evidence. With four types of evidence, participants who scored low on NFC also perceived differences, but only between anecdotal evidence (lower persuasiveness) and the three other types of evidence (higher persuasiveness). Furthermore, NFC was related to the persuasiveness of anecdotal evidence for the Dutch participants (F (1, 300) = 6.38, $p < .05$, $\eta^2 = .02$), but not for the French participants ($F < 1$). For the Dutch participants, their NFC correlated positively with the persuasiveness of anecdotal evidence (r (304) = .10, $p = .07$), but this correlation was only marginally significant. Other hypotheses were not confirmed: NFC was not related to the persuasiveness of causal evidence (for both cultures: $F < 1$), and PNI was not related to the persuasiveness of statistical evidence (for both cultures: $F < 1$)[7].

6.3 Conclusion

Study 3 was set up to answer the general RQ2 about whether there are cultural differences in the relative *actual* persuasiveness of anecdotal, statistical, causal, and expert evidence. The results of this study confirm such cultural differences. The specific research question RQ2 was whether expert evidence is more persuasive in France than in the Netherlands. In an absolute way, normatively strong expert evidence was not more persuasive to the French participants, but it was in a relative way. This cultural difference was less pronounced than could be expected on the basis of what is known about the power distances in the Dutch and the French culture. Larger differences could be suggested with normatively weak expert evidence. The French communication scholar Breton (2003) argued that experts can influence people's opinions about an issue that is far from their own field of expertise. This suggests that – under conditions of a large power distance – normatively weak expert evidence with an irrelevant field of expertise may be as persuasive as normatively strong expert evidence. People from a large power distance culture (e.g, the French) more easily accept differences in power distance (Hofstede, 1980, 2001), and may therefore be less affected by the relevance of the experts' field of expertise, provided that these experts have a high status (e.g., because of titles or affiliation to a research center). People from small power distance cultures such as the Dutch culture could be said to take into account the relevance of the field of expertise, as predicted by normative rules from argumentation theory. Study 4 therefore investigated whether there are cultural differences in the persuasiveness of normatively strong and normatively weak expert evidence in France and the Netherlands.

Study 4

This study focused on the specific research question RQ2 about whether the actual persuasiveness of expert evidence is higher for the French than for the Dutch. In Study 3, I demonstrated that normatively strong expert evidence was only relatively more persuasive in France, and I suggested that larger cultural differences could be in play with normatively weak expert evidence instantiations. Based on Breton (2003), and Hofstede (1980, 2001), the following research questions were formulated:

RQ3 Is there a cultural difference in the relative actual persuasiveness of normatively strong and normatively weak expert evidence in France and the Netherlands?

RQ4 Is normatively weak expert evidence more persuasive in France than in the Netherlands?

I also included statistical evidence for three reasons. First, if French participants in this study turn out not to differentiate between normatively strong and normatively weak expert evidence, this could also be attributed to their insensitivity to evidence quality. This attribution would find support, if this insensitivity also occurs with statistical evidence. Second, statistical evidence was chosen because it was possible to create normatively strong and weak instantiations of it. In contrast, this was impossible for anecdotal and causal evidence (see Section 5.1.2). Third, the inclusion of statistical evidence allowed me to examine whether the interaction effect between culture and statistical / expert evidence in Study 3 could be replicated. Study 4 was set up to answer these research questions.

6.4 Method

6.4.1 Material

Participants received 20 claims: 10 claims with fillers, and 10 claims with manipulated evidence that were the least probable claims in Study 3, and that did not differ significantly between the two cultures. In Studies 2 and 3, expert evidence was normatively strong, because the experts were credible and reliable ('Prof. dr.' in the Dutch material, and also the affiliation with the National Research Center in the French material), and because their field of expertise was relevant to the field of the claim. Normatively weak expert evidence was created by changing the relevant field of expertise into an irrelevant field of expertise. Note that it is likely that experts with an irrelevant field of expertise may also be considered less credible and reliable (cf. Section 3.1.1).

Each field of expertise in Study 3 was used again for strong expert evidence, but also for weak expert evidence. For instance, the field of dietetics was linked to a claim about basil in pasta sauce leading to better sporting performance in the strong version, and was also linked to a claim about an improvement in boys'

school performance as a result of putting them next to girls in class in the weak version. The distribution of the fields of expertise to the weak expert evidence instantiations followed two criteria: (a) there should be a strong differentiation between the relevant and the irrelevant field related to a claim (e.g., behavioral therapy versus eye diseases, and not versus social psychology), and (b) participants should not receive the same field of expertise twice[8]. The two templates for expert evidence were used, each for normatively strong and for normatively weak expert evidence.

The statistical evidence instantiations in Studies 2 and 3 were normatively strong, because they had large sample sizes, and high percentages of cases in the sample. In this experiment, two sets of normatively strong and normatively weak statistical evidence were created: '78% of 314 persons' and '74% of 381 persons' for the strong instantiations, and '35% of 46 persons' and '38% of 53 persons' for the weak instantiations. The two templates for statistical evidence were used, each template for normatively strong and for normatively weak evidence. Finally, some claims were supported by causal or anecdotal evidence. These claims were taken from Study 3, and were used as fillers between the normatively strong and normatively weak statistical and expert evidence instantiations. The names, cities, and universities mentioned in the material are listed in Appendix H.

6.4.2 Participants

The Dutch participants in this study were students from universities in Amsterdam (n = 73; five groups), Delft (n = 21; two groups), Enschede (n = 28; three groups), Nijmegen (n = 77), and Tilburg (n = 101; three groups). The French participants studied in Besançon (n = 49), Paris (n = 56; two groups), Roubaix (n = 58), Strasbourg (n = 65; six groups), and Tours (n = 72). Questionnaires from participants who were non-natives (54 in France, and 7 in the Netherlands), and/or who were younger than 17 or older than 30 were excluded from analyses. In the dataset that resulted from this selection (France: n = 313; Netherlands: n = 315) there were differences in the number of participants in each version. As the lowest number of participants in each version was 60 in both countries, participants were randomly rejected in versions that contained more than 60 participants. In the end, there were 300 French participants (60 in each version), and 300 Dutch participants (60 in each version). Appendix H gives an overview of the number of participants for each version, and for each group in France and in the Netherlands[9].

Of the French participants, 81.3% was female, whereas this percentage was only 70.0% for the Dutch participants. This difference was significant (χ^2 (1) = 10.32, p < .01). Participants' sex did not affect the relative persuasiveness of the types of evidence (F < 1), but did affect the relative persuasiveness of strong and weak evidence (F (1, 597) = 4.71, p < .05, η^2 = .01). In fact, strong evidence was more persuasive to the male participants (M = 0.42, SD = 0.73) than to the female participants (M = 0.27, SD = 0.68) (t(597) = 2.25, p < .05). However, more importantly, for both the male participants (t(145) = 3.41, p < .01) and the female participants (t(452) = 2.55, p < .05) strong evidence was more persuasive than weak evidence. The age of the French participants ranged from 17 to 30, with a mean of 20.19 (SD = 1.81). The Dutch participants were 20.64 years old on average (SD = 1.91), with ages from 17 to 26. The Dutch participants were significantly older than the French participants (t(596.21) = 2.97, p < .01). This difference did not affect the persuasiveness of evidence, as age did not interact with evidence type (F (1, 598) = 1.35, p = .25), or evidence quality (F < 1). The French participants studied at the Faculty of Arts (98.18%; Applied Foreign Languages, Communication, Languages), or Social Sciences (1.84%). The Dutch participants studied at the Faculty of Arts (92.33%; Business Communication Studies, Communication Sciences, Languages), Social Sciences (0.67%), or a technical discipline (7.00%).

6.4.3 Design

The design of Study 3 was used, but there were only 10 claims with manipulated evidence; the other 10 claims were fillers (5 with anecdotal, and 5 with causal evidence)[10]. A design with 20 manipulated claims could have led participants to discover the goal of Study 4, as they would encounter only statistical and expert evidence. All participants received the 20 claims in exactly the same order in each version, but the distribution of the five types of evidence over the 10 experimental claims and the five versions followed a balanced Latin square.

6.4.4 Instrumentation

Again, the booklet that participants received was titled 'Opinions on social issues' and started with an instruction. After the instruction, 20 pairs of claims with different types of evidence followed. For each of the claims, participants judged the probability on 5-point semantic differentials (*very improbable – very probable*). After these judgments, participants judged the degree to which they agreed with items of three context variables on a 5-point Likert scale: seven items of the NFC

scale (Cacioppo, et al., 1984), four items of the PEI scale (Studies 2 and 3), and 10 items of the RWA scale (Altemeyer, 1988).

The RWA scale was included, because the context variable PEI could not account for cultural differences in Studies 2 and 3. The Right-Wing Authoritarianism scale (Altemeyer, 1981, 1988) is considered the most elaborate and widely accepted theory of authoritarianism (Kemmelmeier, et al., 2003). Right-wing authoritarianism is composed of three related concepts (authoritarian submission, authoritarian aggression, and conventionalism) of which authoritarian submission appears relevant for expert evidence. It is "submission to the authorities who are perceived to be established and legitimate in the society in which one lives" (Altemeyer, 1981, p. 148). A number of empirical studies have shown that power distance is related to this scale. Feather (1998), and Rohan and Zanna (1996), for instance, showed that the Schwartz (1994) power value (related to power distance) was more important for people who score high on RWA than for people who score low on RWA. The selection of the 10 items used in the present study is explained in Appendix I. With respect to RWA, it was hypothesized that French participants would score higher than Dutch participants, and that the higher participants scored on RWA, the more they would be persuaded by strong expert evidence. The reliability of the 10 items of this scale proved to be adequate for the French participants ($\alpha = .71$), but only marginally so for the Dutch participants ($\alpha = .60$).

The NFC scale was included in the questionnaire as a control with Study 3, where the score on this scale proved to affect participants' ability to perceive differences in the relative persuasiveness of three evidence types. I hypothesized that people who scored high on NFC would perceive the difference between normatively strong and normatively weak evidence better than participants who scored low on NFC. The seven items of this scale were reliable (Dutch participants: $\alpha = .72$; French participants: $\alpha = .78$). The PEI scale too served as a control with Study 3, where it affected the persuasiveness of expert evidence for the Dutch, but not for the French participants. The PEI scale was reliable (Dutch participants: $\alpha = .75$; French participants: $\alpha = .79$). NFC and PEI were found not to be culturally unreliable (p's > .05; Van de Vijver & Leung, 1997).

After these items, four control questions followed about the expertise of the four experts that each participant read about (cf. Studies 2 and 3). One control question about the perceived quality of statistical evidence followed. On a 5-point semantic differential, participants were asked to indicate which of the two examples they would choose as proof for the generality of the occurrence of an

effect: "the effect occurs in 35% of 46 persons" or "the effect occurs in 78% of 314 persons". The questionnaire ended with questions about participants' age, sex, nationality, and current education.

6.4.5 Procedure

The procedure was the same as for Study 3. Students of several universities in the Netherlands and France filled in the questionnaire in October – December 2004. The study was introduced as being about social issues. The students were not rewarded for their participation, which took about 13 to 18 minutes. After the questionnaires had been collected, the real research purpose was revealed, and participants were thanked for their cooperation. There were no disturbances during the experiment.

6.4.6 Statistical tests

The persuasiveness of statistical and expert evidence was computed as a difference score between the judgment of each claim with evidence and that of a claim without evidence. The research question about cultural differences in the persuasiveness of normatively strong and normatively weak expert evidence was evaluated through a 2 (culture) x 2 (quality) x 2 (type) analysis of variance, where culture was a between-subjects factor, and type of evidence and evidence quality within-subjects factors. The research question about cultural differences in the persuasiveness of normatively weak expert evidence was investigated in two ways (cf. Study 3). The persuasiveness of weak expert evidence in the two cultural groups was directly compared with an independent t-test, and indirectly by comparing it with the persuasiveness of strong expert evidence. Next to these analyses by participants, analyses by stimuli were also conducted.

The occurrence of a carry over effect was tested with a 2 (first judgment, last judgment) x 2 (statistical, expert) analysis of variance with repeated measures, and a 2 (first judgment, last judgment) x 2 (strong, weak) analysis of variance with repeated measures. If participants had learned to perceive differences between the type and the quality of evidence, there should have been significant interaction effects. However, the interaction effect was not significant for time of judgment and type of evidence (F (1, 599) = 2.33, p = .13), or for time of judgment and quality of evidence (F (1, 599) = 1.63, p = .20).

6.5 Results

As in Studies 2 and 3, I used the index of Bachman and O'Malley (1984) to test whether the Dutch and French participants differed in their use of the scale endpoints. Table 6.5 shows the indices for the judgments of the 10 claims[11], the three context variables (21 items), and the expertise of the four experts. Because of cultural differences in response extremity on the claims and the context variables, the scores on these items were standardized (for procedure, see Study 2). The analyses below will concern only the standardized data, unless indicated otherwise.

Table 6.5 Extreme response indices for Dutch and French participants and t-test for cultural differences (*SD* between parentheses)

variable	Dutch index	French index	t-test
claims	.22 (0.19)	.31 (0.19)	$t(598) = 5.98$, $p < .001$
context variables	.20 (0.17)	.31 (0.17)	$t(598) = 7.97$, $p < .001$
expertise of experts	.19 (0.24)	.20 (0.27)	$t(598) = 0.67$, $p = .50$

I tested whether the manipulations of strong and weak statistical and expert were successful. As cultures may differ in their sensitivity to evidence quality, these tests were run per culture. I first checked whether strong statistical evidence was indeed perceived as stronger than weak statistical evidence (t-tests with raw data). This was the case for both the French participants ($M = 3.71$, $SD = 1.28$; $t(290) = 9.52$, $p < .001$), and the Dutch participants ($M = 4.39$, $SD = 0.99$; $t(298) = 24.12$, $p < .001$), as each group of participants scored above 3.00 on the manipulation check. However, the manipulation was more successful for the Dutch participants than for the French participants ($t(547.15) = 7.12$, $p < .001$).

Next, I checked whether the normatively strong experts were considered as having more expertise than the normatively weak experts. The French participants perceived the strong experts ($M = 3.02$, $SD = 0.86$) as more expert than the weak experts ($M = 2.61$, $SD = 0.92$), $F(1, 299) = 46.48$, $p < .001$, $\eta^2 = .14$. Similarly, the Dutch participants considered the strong experts ($M = 3.30$, $SD = 0.83$) had more expertise than the weak experts ($M = 2.33$, $SD = 0.85$), $F(1, 299) = 255.81$, $p < .001$, $\eta^2 = .46$. The operationalization of weak and strong expert evidence was successful, but this main effect of quality was qualified by culture ($F(1, 598) = 43.93$, $p < .001$, $\eta^2 = .07$). This interaction effect is shown in Figure 6.2.

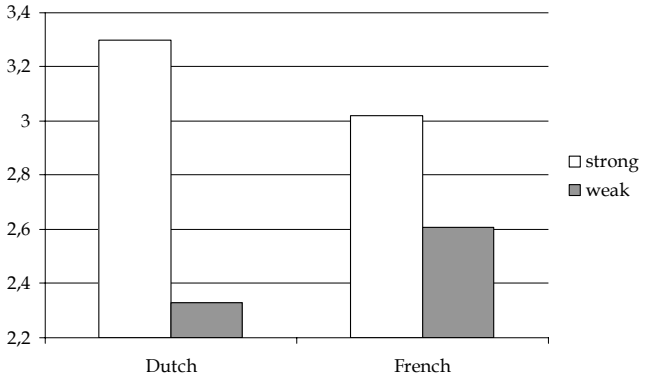

Figure 6.2 The perceived expertise of strong and weak expert in function of culture

As Figure 6.2 suggests, the difference in expertise between strong and weak experts was more pronounced for the Dutch participants (η^2 = .46) than for the French participants (η^2 = .14). The Dutch participants perceived the strong experts as *more* expert than the French participants ($t(598) = 3.99$, $p < .001$), and the weak experts as *less* expert than the French participants ($t(598) = 3.94$, $p < .001$). In sum, the manipulations of strong and weak statistical and expert evidence were successful, but to a larger extent for the Dutch participants than for the French participants. Whether these cultural differences affected the sensitivity to evidence quality will be shown below, where the results relevant to the research questions (6.5.1), and relevant to the context variables (6.5.2) are presented.

6.5.1 Research questions

The persuasiveness of the strong and the weak instantiations of statistical and expert evidence for the Dutch and the French participants is shown in Table 6.6.

Table 6.6 Persuasiveness of evidence in function of culture, type and quality (standardized data; *SD* between parentheses; different superscripts refer to significant differences within-cultures, alpha level of .05)

type of evidence	Dutch (n = 300)	French (n = 300)	total (N = 600)
statistical evidence			
strong	1.04[a] (1.77)	0.46[a] (1.72)	0.75[a] (1.77)
weak	0.42[c] (1.72)	0.35[a] (1.62)	0.39[b,c] (1.67)
expert evidence			
strong	0.73[b] (1.73)	0.25[a] (1.66)	0.49[b] (1.71)
weak	0.25[c] (1.74)	0.36[a] (1.59)	0.31[c] (1.67)

In order to answer the research questions, an analysis of variance with the factors culture (Dutch, French), quality (high, low), and evidence type (statistical, expert) was carried out. First, there was a main effect of type of evidence on persuasiveness with an analysis by participants (F_1 (1, 598) = 8.22, p < .01, η^2 = .01) – suggesting that statistical evidence (M = 0.57, SD = 1.30) was more persuasive than expert evidence (M = 0.40, SD = 1.32) – but there was only a tendency for such a main effect with an analysis by stimuli (F_2 (1, 9) = 4.01, p = .08). There was also a main effect of quality (F_1 (1, 598) = 18.53, p < .001, η^2 = .03; F_2 (1, 9) = 9.26, p < .05, η^2 = .51): high quality evidence (M = 0.62, SD = 1.39) was more persuasive than low quality evidence (M = 0.35, SD = 1.29). A main effect of culture – suggesting more impact of evidence for the Dutch (M = 0.61, SD = 1.09) than for the French participants (M = 0.36, SD = 1.06) – occurred with an analysis by participants (F_1 (1, 598) = 8.43, p < .01, η^2 = .01), but not with an analysis by stimuli (F_2 (1, 9) = 3.29, p = .10).

There was no interaction effect between evidence type and evidence quality (F_1 (1, 598) = 2.16, p = .14; F_2 (1, 9) = 1.48, p = .25). Another interaction effect, however, did occur, namely between culture and evidence quality (F_1 (1, 598) = 17.91, p < .001, η^2 = .03; F_2 (1, 9) = 25.61, p < .01, η^2 = .74). For the Dutch participants, there was a main effect of quality (F_1 (1, 299) = 35.47, p < .001, η^2 = .11; F_2 (1, 9) = 16.47, p < .01, η^2 = .65): strong evidence (M = 0.89, SD = 1.41) was more persuasive than weak evidence (M = 0.34, SD = 1.28). For the French participants, however, there was no main effect of evidence quality (F_1 < 1; F_2 < 1). Normatively strong evidence (M = 0.36, SD = 1.33) was as persuasive as normatively weak evidence (M = 0.35, SD = 1.30).

With respect to the research question about the persuasiveness of normatively strong and normatively weak expert evidence in France and the Netherlands (RQ3), the interaction effect between culture and quality had to be investigated for expert evidence alone. Although there was no three-way interaction effect between the three factors (F_1 < 1; F_2 < 1) – implying that the interaction effect did not differ between statistical and expert evidence – this interaction effect was still analyzed for each type of evidence separately. Figure 6.3 visualizes the interaction effects between culture and quality for expert and statistical evidence.

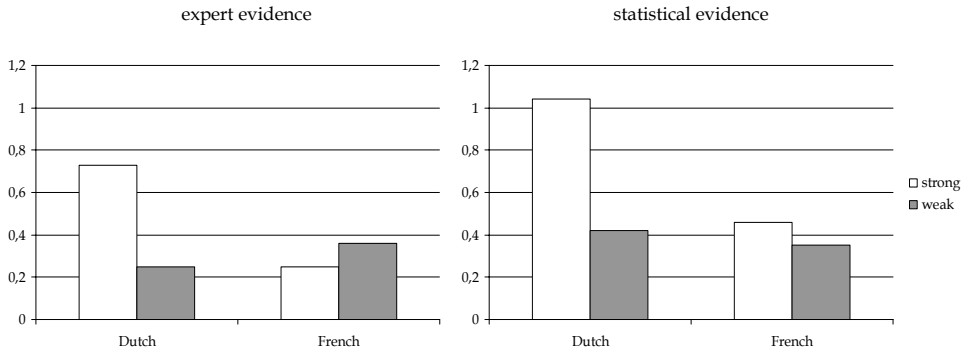

Figure 6.3 The persuasiveness of strong and weak evidence in function of type and culture (based on Table 6.6)

First, the interaction effect between culture and quality on the persuasiveness of expert evidence was significant (RQ3): F_1 (1, 598) = 11.43, p < .01, η^2 = .02; $F2$ (1, 9) = 14.05, p < .01, η^2 = .61. For the French participants, there was no difference in the persuasiveness of strong and weak expert evidence (t_1 (299) = 0.89, p = .37; t_2 (9) = 1.03, p = .33), whereas strong expert evidence was more persuasive than weak expert evidence for the Dutch participants (t_1 (299) = 3.77, p < .001; t_2 (9) = 2.37, p < .05). A similar interaction effect was found for statistical evidence: F_1 (1, 598) = 7.62, p < .01, η^2 = .01; F_2 (1, 9) = 20.47, p < .01, η^2 = .70. For the French participants, strong statistical evidence was as persuasive as weak statistical evidence (t_1 (299) = 0.90, p = .37; t_2 (9) = 1.65, p = .13), but for the Dutch participants strong statistical evidence was more persuasive than weak statistical evidence (t_1 (299) = 4.65, p < .001; t_2 (9) = 4.63, p < .01).

The last possible interaction effect was that between culture and evidence type. This interaction effect was not significant (F_1 (1, 598) = 1.37, p = .24; F_2 (1, 9) = 1.34, p = .28)[12]. In relation to this, the specific research question (RQ4) focused on the persuasiveness of normatively weak expert evidence in France and in the Netherlands. In an absolute way, weak expert evidence was equally persuasive in both cultures (t_1 (598) = 0.77, p = .44; t_2 (9) = 0.61, p = .56). In a relative way, as was shown above, weak expert evidence was more persuasive in France, as it was as persuasive as strong expert evidence. Finally, it was tested whether the interaction effect between culture and evidence type (strong statistical, strong expert) in Study 3 was replicated. This interaction effect did not occur (F_1 < 1; F_2 < 1).

6.5.2 Context variables

Context variables were selected in order to explain possible cultural differences in the persuasiveness of evidence types. The French and Dutch participants did not differ with respect to PEI, NFC (cf. Studies 2 and 3), and RWA (see Table 6.7).

Table 6.7 Scores on the context variables and t-test for cultural differences (*SD* between parentheses; means with raw data, test with standardized data)

variable	Dutch participants	French participants	t-test
NFC	3.13 (0.61)	3.25 (0.74)	$t(573.79) = 1.65$, $p = .10$
PEI	2.44 (0.67)	2.38 (0.77)	$t(585.51) = 1.65$, $p = .10$
RWA	2.72 (0.46)	2.73 (0.55)	$t(581.81) = 0.61$, $p = .54$

The persuasiveness of strong expert evidence was expected to be affected by RWA and, for the Dutch participants, by PEI. For the Dutch participants, both variables had a marginal effect on the persuasiveness of strong expert evidence (RWA: F (1, 296) = 2.44, $p = .06$, one-tailed; PEI: F (1, 296) = 3.67, $p < .05$, one-tailed), but the correlations were not significant (RWA: r (300) = -.07, $p = .25$; PEI: r (300) = .09, $p = .11$). For the French participants, RWA and PEI did not affect the persuasiveness of strong expert evidence (both F's < 1).

A participant's score on NFC was expected to affect the perception of the difference in evidence quality. As the Dutch and French participants reacted differently to evidence quality, the interaction effect between NFC and evidence quality was computed for each culture. For the French participants, there was no effect for NFC on the difference between low and high evidence quality ($F < 1$). For the Dutch participants, however, there was an interaction effect between NFC and evidence quality (F (1, 298) = 5.95, $p < .05$, $\eta^2 = .02$). More precisely, NFC affected the quality of expert evidence. Contrary to what was hypothesized, normatively strong expert evidence was more persuasive than the weak counterpart for the participants with low NFC (F (1, 163) = 12.42, $p < .01$, $\eta^2 = .07$), whereas there was only a tendency for this difference for the participants that scored high on NFC (F (1, 135) = 2.94, $p = .09$)[13].

Finally, I examined the relationships between RWA and PEI on the one hand, and the perceived expertise of the strong and weak experts for each culture on the other (raw data). For the French participants, the perception of the strong experts' expertise correlated positively with the score on PEI (r (300) = .24, $p < .001$), and that of the weak experts correlated positively with the score on RWA (r (300) = .14,

$p < .05$). For the Dutch participants, PEI correlated positively with the perception of the expertise of strong (r (300) = .18, $p < .01$) and weak experts (r (300) = .18, $p < .01$), whereas RWA correlated positively with the perception of the expertise of weak experts (r (300) = .11, $p < .05$)[14].

6.6 Conclusion

Based on Breton (2003), and Hofstede (1980, 2001), it was suggested that there could be cultural differences in the persuasiveness of normatively weak expert evidence. Study 4 showed that there was a cultural difference in the relative actual persuasiveness of normatively strong and weak expert evidence in France and the Netherlands (RQ3). Strong expert evidence was more persuasive than weak expert evidence for the Dutch participants, but both types of expert evidence were equally convincing for the French participants. In a relative way, then, weak expert evidence was more persuasive in France than in the Netherlands. In an absolute way, this was not the case (RQ4).

The fact that the two types of expert evidence were equally persuasive to French participants can be explained in two directions. A first, more general, explanation comes from the ELM. The French participants were not sensitive to the quality of statistical and expert evidence, because they may have taken the peripheral route to persuasion by using heuristics: 'There is numerical information or an expert source, so the claim must be probable'. There was also an indication, however, that the French participants were following the central route to persuasion. In fact, their scores on NFC, a marker for the central route, were moderate, and similar to those of the Dutch participants.

A second, more specific, explanation deals with expert and statistical evidence separately. In fact, the reason why French participants were insensitive to the strength of expert evidence is not necessarily related to the reason why they were insensitive to the strength of statistical evidence. As the focus in this dissertation lies on expert evidence, and not on statistical evidence, I will only explore an explanation for expert evidence. Power distance suggests that people in large power distance cultures are insensitive to the experts' field of expertise when these experts have a high status (e.g., professors). This explanation was tested by having participants judge expert sources with a lower status than professors. This was done in Study 5.

Study 5

In Study 4, strong and weak expert evidence were equally persuasive to the French participants. Similarly, there was a relatively small difference in the perception of the expertise of strong and weak experts. This can be explained by the experts' status. Because of their professorship, these experts may not need to be specialized in a certain field in order to influence people of claims within this field. This explanation holds, if field of expertise has a greater influence on perceived expertise when the experts have a lower status (i.e., when they are not professors):

RQ5 In France, does the influence of the field of expertise on the perception of experts' expertise depend on their status?

If this field of expertise indeed matters more with low status experts, the finding in Study 4 can be explained by the status of the professors. For comparison reasons, these relationships will also be tested with Dutch participants. For them, field of expertise mattered greatly for the perceived expertise of professors in Study 4, and there is no reason to believe this will change with an expert with a lower status:

RQ6 In the Netherlands, is the influence of the field of expertise on the perception of experts' expertise independent on their status?

Study 5 was conducted to answer these two research questions.

6.7 Method

6.7.1 Material

Four types of expert evidence were created on the basis of status (professor – researcher), and field of expertise (high – low). Two of these four expert types were also used in Study 4: (1) a professor with a relevant field of expertise, and (2) a professor with an irrelevant field of expertise. These experts are supposed to have a relatively high status. Experts with lower status were created by leaving out references to Professor and to the CNRS: (3) a researcher with a relevant field of expertise, and (4) a researcher with an irrelevant field of expertise.

(1) According to Prof. dr. Timmermans from the University of Amsterdam, a specialist in *dietetics*, the consumption of basil in tomato pasta sauce improves sporting performance.

(2) According to Prof. dr. Timmermans from the University of Amsterdam, a specialist in *didactics*, the consumption of basil in tomato pasta sauce improves sporting performance.

(3) According to Timmermans from the University of Amsterdam, a specialist in *dietetics*, the consumption of basil in tomato pasta sauce improves sporting performance.

(4) According to Timmermans from the University of Amsterdam, a specialist in *didactics*, the consumption of basil in tomato pasta sauce improves sporting performance.

The two templates for expert evidence were used (see Table 5.1), and presented alternately. The combinations of names and universities made in Study 2 were used in this experiment, but with one change. The city of Paris was mentioned instead of Strasbourg, because the French participants came from this city (see Appendix J).

The 8 claims in the material were selected from the 10 claims of Study 4 for which there was an effect for expert evidence quality for the Dutch participants, and not for the French participants. In this way, claims (22) and (3) from Study 4 were not taken into account. The eight claims were presented in the same order as in Study 4[15].

6.7.2 Participants

Questionnaires from students who were non-natives (34 in France, and 5 in the Netherlands), and/or who were younger than 17 or older than 30 were excluded from analyses. The Dutch participants (n = 106) studied Business Communication Studies (44.3%; n = 47) or History (55.7%; n = 59) at the Radboud University Nijmegen. There was no main effect of study on the perception of the expertise of expert type (1) ($t(104) = .09$, $p = .93$), (2) ($t(104) = .02$, $p = .99$), (3) ($t(104) = 1.18$, $p = .24$), or (4) ($t(104) = .64$, $p = .52$). The French participants (n = 112) studied Applied Foreign Languages at the University Marc-Bloch in Strasbourg (two groups; n = 59, n = 53). The Dutch participants were 19.69 years old on average (SD = 1.60), with ages from 18 to 25. The age of the French participants ranged from 18 to 26, with a

mean of 19.67 (*SD* = 1.37). This difference in ages in the two cultures was not significant ($t(207.09) = 0.00$, $p = .93$). The percentage of male participants was higher in the Dutch group (41.5%) than in the French group (14.3%) (χ^2 (1) = 20.23, $p < .001$), but participants' sex hardly affected the perceived expertise of the experts[16].

6.7.3 Design

All participants received the eight claims in exactly the same order, but in each of the four versions of the material the distribution of the four types of experts differed across claims. The types of experts were assigned to different claims in each version through a balanced Latin square.

6.7.4 Instrumentation

In the booklet called 'Social issues', it was explained that participants would receive a number of social issues highlighted by researchers. There were eight claims supported by different types of expert evidence. The participants indicated their perception of the experts' expertise by indicating on a 5-point Likert scale the degree to which they found that each expert had enough expertise to make a judgment about a claim (similar to the control questions in Studies 2, 3, and 4). Subsequently, participants indicated on a 5-point Likert scale the degree to which they agreed with 10 items of the RWA scale (Altemeyer, 1988), seven items of the NFC scale (Cacioppo, et al., 1984), and four items of the PEI scale (Studies 2, 3, and 4). These items were included in order to control with the other experiments, and to examine their effects on perceived expertise. After leaving out item (10), the RWA scale was reliable for the Dutch participants ($\alpha = .75$), but not for the French ($\alpha = .58$). For comparison with Study 4, this scale was still taken into account. The NFC scale was reliable for both cultural groups (Dutch: $\alpha = .81$; French: $\alpha = .78$), and the same was true for the PEI scale (Dutch: $\alpha = .79$; French: $\alpha = .82$). Moreover, these two scales were not culturally unreliable (p's > .05; Van de Vijver & Leung, 1997).

A new question about expert evidence was added. Participants had to indicate on three 5-point semantic differentials the degree to which they found the use of an expert to persuade other people: *unwise* – *sensible*, *definitively a success* – *no guarantee for success*, and *to be advised against* – *recommendable*. The scale was found not to be reliable (Dutch: $\alpha = .30$; French: $\alpha = -.03$). Participants may have given responses in the same direction on the questionnaire, but they appear to have overlooked the

endpoints of the second item, which were reversely coded. As the scale reliability of the two other items was too low (Dutch: $\alpha = .52$; French: $\alpha = .34$), these items will be analyzed separately. The questionnaire ended with questions about participants' age, sex, nationality, and current education.

6.7.5 Procedure

The procedure was the same as in the other studies. Questionnaires were filled in at universities in Nijmegen and Strasbourg in the months March and April 2005. The study was introduced as being about social issues. The students were not rewarded for their participation, which took about 8 to 10 minutes. After the questionnaires had been collected, the real research purpose was revealed, and participants were thanked for their cooperation. There were no disturbances during the experiment.

6.7.6 Statistical tests

First, a 2 (culture) x 2 (status) x 2 (field of expertise) analysis of variance was run, in which culture was a between-subjects factor, and status and field of expertise within-subjects factors. Next, similar analyses of variance were run for each cultural group in order to test the research questions. Analyses by stimuli were also conducted. The occurrence of a carry over effect was tested with a 2 (first judgment, last judgment) x 2 (professor, researcher) analysis of variance with repeated measures, and a 2 (first judgment, last judgment) x 2 (relevant, irrelevant field of expertise) analysis of variance with repeated measures. If participants had learned to perceive differences between the types of experts, there should have been significant interaction effects. The interaction effect was not significant for time of judgment and status (F (1, 217) = 1.50, p = .22), but it was for time of judgment and field of expertise (F (1, 217) = 5.15, $p < .05$, $\eta^2 = .02$). More precisely, it was only significant for the Dutch participants (F (1, 105) = 15.83, $p < .001$, $\eta^2 = .13$): the difference in field of expertise of the first pair of experts was larger than for the second pair. As the strong and weak fields of expertise were still significantly different for the second pair of experts, this interaction effect has no consequences.

6.8 Results

Using the index of Bachman and O'Malley (1984), I tested whether the Dutch and French participants differed in their use of the scale endpoints (cf. Studies 2 – 4). Table 6.8 shows the indices for the judgments for the eight experts, and the four context variables (22 items, as one RWA item and one new question were not taken into account).

Table 6.8 Extreme response indices for Dutch and French participants and t-test for cultural differences (*SD* between parentheses)

variable	Dutch index	French index	t-test
expertise of experts	.31 (0.22)	.32 (0.22)	$t(216) = 0.31$, $p = .76$
context variables	.22 (0.18)	.32 (0.18)	$t(216) = 4.00$, $p < .001$

Because of cultural differences in response extremity on the context variables, the scores on these items were standardized (for procedure, see Study 2). These standardized data were used for the analyses. I first tested whether there was a cultural difference in the interaction effect under investigation, namely between status and field of expertise. Although the three-way interaction effect was not significant ($F_1 < 1$; $F_2 < 1$), I did run analyses of variance for each culture in order to answer the research questions. The Dutch and French perceptions of the expertise of the four types of experts are shown in Table 6.9.

Table 6.9 The perception of expertise in function of culture and type of expert (*SD* between parentheses; different superscripts refer to significant differences within-cultures, alpha level of .05)

type of expert	Dutch (n = 106)	French (n = 112)
professor		
relevant field of expertise	3.77^{a} (0.79)	3.01^{a} (0.96)
irrelevant field of expertise	2.28^{b} (0.84)	$2.75^{b,c}$ (0.95)
researcher		
relevant field of expertise	3.84^{a} (0.80)	$2.93^{a,c}$ (0.94)
irrelevant field of expertise	2.12^{b} (0.77)	2.54^{b} (0.91)

For the French participants, there was a main effect of field of expertise on perceived expertise (F_1 (1, 111) = 12.54, p < .01, η^2 = .10; F_2 (1, 7) = 5.68, p < .05, η^2 = .45): experts with a relevant field of expertise (M = 2.97, SD = 0.73) were considered as having more expertise than experts with an irrelevant field of expertise (M = 2.65, SD = 0.71), but the effect size was moderate. Next, there was a tendency for a main effect of status with an analysis by participants (F_1 (1, 111) = 3.15, p = .08) – suggesting that high status experts (M = 2.88, SD = 0.72) were perceived as more expert than low status experts (M = 2.74, SD = 0.64) – but not with an analysis by stimuli (F_2 (1, 7) = 1.92, p = .21). Finally, there was no interaction effect between status and field of expertise (F_1 < 1; F_2 < 1). This means that the relationship between field of expertise and perceived expertise was not moderated by the status of the expert (RQ5).

For the Dutch participants, there was a very strong main effect of field of expertise on perceived expertise (F_1 (1, 105) = 357.70, p < .001, η^2 = .77; F_2 (1, 7) = 33.31, p < .01, η^2 = .83): experts with a relevant field of expertise (M = 3.81, SD = 0.68) were considered as having much more expertise than experts with an irrelevant field of expertise (M = 2.20, SD = 0.57). Next, there was no main effect of status (F_1 < 1; F_2 < 1), nor an interaction effect between status and field of expertise (F_1 (1, 105) = 2.96, p = .09; F_2 < 1). Just as for the French participants, the relationship between field of expertise and perceived expertise was not affected by the expert's status (RQ6).

Finally, there were strong cultural differences in the perceptions of the expertise of the four experts. There was a large interaction effect between culture and field of expertise on the expertise perception (F_1 (1, 216) = 105.344, p < .001, η^2 = .33; F_2 (1, 14) = 17.14, p < .01, η^2 = .55). In fact, the experts with a relevant field of expertise were perceived as *more* expert by the Dutch participants than by the French participants (professor: t_1 (216) = 6.38, p < .001; t_2 (7) = 5.07, p < .01; researcher: t_1 (216) = 7.70, p < .001; t_2 (7) = 6.09, p < .001). Next, the experts with an irrelevant field of expertise were perceived as *less* expert by the Dutch participants than by the French participants (professor: t_1 (216) = 3.87, p < .001; t_2 (7) = 2.03, p = .08; researcher: t_1 (216) = 3.71, p < .001; t_2 (7) = 2.13, p = .07). Simply stated, the effects of field of expertise on the expertise perception were much more pronounced for the Dutch than for the French participants. This is visualized in Figure 6.4.

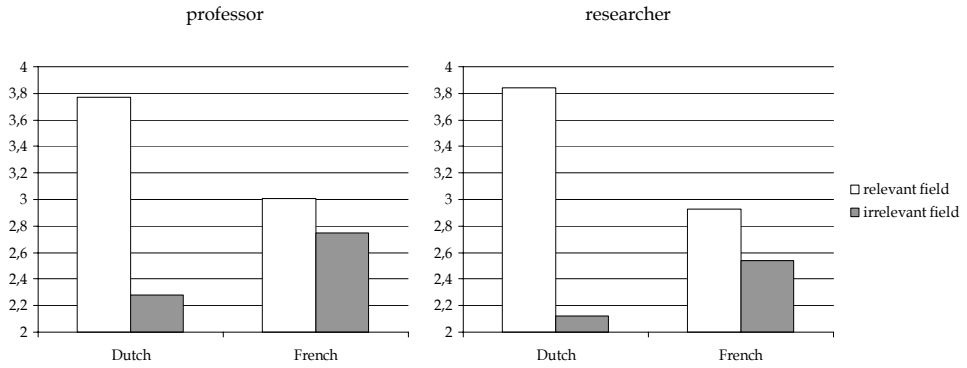

Figure 6.4 The perceived expertise of experts with a relevant and an irrelevant field of expertise in function of expert type and culture (based on Table 6.9)

A number of context variables were included in the questionnaire in order to explain possible cultural differences, and to control with the other experiments. The French and Dutch participants did not differ in their scores on NFC (cf. Studies 2, 3, and 4), and RWA (cf. Study 4), but did differ in PEI, and on the two new questions about the use of experts. The Dutch participants scored higher on PEI, and the French participants scored higher on the two questions about the use of experts (see Table 6.10).

Table 6.10 Scores on the context variables and t-test for cultural differences (*SD* between parentheses; means with raw data, test with standardized data)

variable	Dutch	French	t-test
NFC	3.13 (0.75)	3.28 (0.75)	$t(216) = 0.09$, $p = .93$
PEI	2.66 (0.76)	2.40 (0.87)	$t(216) = 2.93$, $p < .01$
RWA	2.68 (0.59)	2.90 (0.52)	$t(198.96) = 1.59$, $p = .11$
the use of experts is:			
a sensible thing to do	3.77 (0.80)	4.15 (0.80)	$t(214) = 3.50$, $p < .01$
recommendable	3.45 (0.81)	3.65 (0.98)	$t(214) = 1.65$, $p = .05$*

* one-tailed test

I therefore examined the effects of these two context variables on the perceived expertise of the four experts in each culture. For the French participants there were only two significant effects. First, the answer to the question about how sensible it is to use experts had an effect on the perception of the expertise of the researcher with an irrelevant field of expertise (F (1, 104) = 5.41, $p < .05$, $\eta^2 = .05$). These variables correlated positively, but weakly: r (110) = .24, $p < .05$. Second, the

answer to the question about whether the use of experts is recommendable affected the perception of the expertise of the researcher with a relevant field of expertise (F (1, 104) = 4.65, $p < .05$, $\eta^2 = .05$). These variables too were correlated positively, but weakly: r (110) = .24, $p < .05$. For the Dutch participants, this same question about how recommendable it is to use experts affected the perception of the expertise of the researcher with an irrelevant field of expertise (F (1, 100) = 5.74, $p < .01$, $\eta^2 = .09$). There was a moderate negative correlation between the two variables: r (106) = -.31, $p < .01$.

6.9 Conclusion

The fact that French participants judged strong and weak expert evidence as equally persuasive in Study 4 was attributed to the relatively small difference in the perceived expertise of professors with a relevant and an irrelevant field of expertise. Study 5 tested whether this small difference could be explained by the professors' status by having participants indicate the perceived expertise of professors (relatively high status) and researchers (relatively low status) with a relevant or an irrelevant field of expertise. The results showed that the influence of field of expertise on the perception of the expert's expertise did not depend on status in France (RQ5). The same small difference in perception of strong and weak professors also applied to researchers. This means that (a) French participants again did not greatly differentiate between relevant and irrelevant fields of expertise, and that (b) this small difference can not be explained by an expert's high status. Finally, for the Dutch participants the large difference in perception of the expertise of strong and weak professors in Study 4 was replicated, both for professors and researchers (RQ6).

6.10 Conclusion and discussion of Studies 3, 4, and 5

The studies presented in this chapter demonstrated that there are cultural differences in the relative *actual* persuasiveness of evidence types (general RQ2). The relative actual persuasiveness of anecdotal, statistical, causal, and expert evidence was different in France and in the Netherlands (Study 3). Normatively strong causal and statistical evidence were more persuasive to the Dutch participants than to the French participants (Study 3). The specific RQ2 focused on

the persuasiveness of expert evidence. Strong expert evidence was not more persuasive in France in an absolute way, but it was in a relative way (Study 3): it was among the most persuasive types of evidence in France, but it took an intermediate position in the relative persuasiveness of the types of evidence in the Netherlands. Next, it was suggested that normatively weak expert evidence – consisting of experts with an irrelevant field of expertise – could be more persuasive in France than in the Netherlands. Similar to the finding in Study 3, normatively weak expert evidence was not more persuasive in France in an absolute way, but in a relative way: strong and weak expert evidence were equally convincing for the French participants, but strong expert evidence was more persuasive than weak expert evidence for the Dutch participants. The equal persuasiveness of strong and weak expert evidence for the French participants was attributed to the small difference in the perception of the expertise of strong and weak experts. Study 5 further demonstrated that this small difference in perception can not be explained by the high status of the professors of strong expert evidence, because this small difference also applied to researchers with a lower status. Below, I will discuss the three studies separately.

Study 3

Of the three studies in this chapter, Study 3 can be compared most easily with previous studies. Comparing the conclusion of the review in Chapter 3 (studies with mostly American participants) with the results of Study 3 would be inappropriate after having shown that culture affects the relative persuasiveness of evidence types. It is therefore more suitable to compare the results for the Dutch participants in Study 3 with a similar experiment with Dutch participants, namely Hoeken and Hustinx (2003b). In Hoeken and Hustinx, statistical, causal, and expert evidence were more persuasive than anecdotal evidence. In Study 3, statistical evidence was also more persuasive than expert and causal evidence, whereas there was only a tendency for such an effect in Hoeken and Hustinx with an analysis by participants. This difference could be attributed to the material. Contrary to Hoeken and Hustinx, the material in Study 3 consisted of moderately probable claims supported by (mostly) normatively strong evidence. Because of these improvements in the material, it is likely that statistical evidence is indeed more persuasive than causal and expert evidence for Dutch participants. Also, it should be noted that anecdotal evidence was not persuasive in both studies, especially in Study 3, where it had a very simple operationalization. Anecdotal evidence was as persuasive as when no evidence was provided. An explanation lies in the

moderately probable claims. In Section 3.2, anecdotal evidence was said to be persuasive, because of the similar cases that it may evoke. With moderately probable claims, participants are likely not to have such similar cases, so that anecdotal evidence can not benefit from this. The low persuasiveness of anecdotal evidence is interesting, since anecdotal evidence has regularly been found to be persuasive when it was operationalized in a more elaborate, vivid way (see Chapter 3).

In Study 3, context variables were included in order to explain possible cultural differences in the actual persuasiveness of evidence types. The two cultural groups did not differ in their PEI, but the role of this personality trait was not the same in both cultures. For the Dutch participants, the score on PEI correlated positively with the persuasiveness of expert evidence, but this correlation was absent for the French participants (cf. Study 4). Expert evidence was persuasive to them, but regardless of their own preference for expert information. As with PEI, the Dutch and French participants did not differ in their scores on NFC and PNI. Moreover, NFC was not related to the persuasiveness of causal evidence, and PNI was not related to the persuasiveness of statistical or anecdotal evidence. The absence of an effect of PNI on the persuasiveness of statistical evidence has parallels with Kazoleas (1993), where the participants' level of experience in statistics and methodology did not affect the persuasiveness of statistical evidence, and with Parrott, Silk, Dorgan, Condit, and Harris (2005), where the level of numeracy skills did not affect the perceived persuasiveness of statistical evidence. The score on NFC, however, affected the overall difference in persuasiveness of the types of evidence: the difference was larger for participants with a relatively high NFC. This corresponds with the ELM (Petty & Cacioppo, 1986), which postulates that NFC is related to the degree to which people are sensitive to differences in argument quality.

Finally, there are two alternative explanations for the actual persuasiveness of statistical evidence. First, the (long) statistical evidence was more persuasive than the (short) anecdotal evidence. Other evidence studies that controlled for evidence length (Allen, et al., 2000; Hoeken & Hustinx, 2003b), however, also showed that statistical evidence was more persuasive than anecdotal evidence. If evidence length affected the relative persuasiveness of evidence types, this effect appears to have been small. Second, it could be suggested that statistical evidence was more persuasive to the Dutch than to the French participants, because the samples sizes were relatively larger for the Dutch populations than for the French populations. I could have multiplied the French sample sizes by three, since France has three

times more inhabitants than the Netherlands. This, however, would have resulted in large sample sizes (e.g., 900 people) that are unrealistic for claims about new information, such as 'The consumption of basil in tomato pasta sauce improves sporting performance'. The use of the same sample sizes has not affected the persuasiveness of statistical evidence for the French participants, as these were found to be insensitive to sample sizes used in statistical evidence (Study 4).

Study 4

This study showed that culture can have an impact on the importance of normative criteria for the persuasiveness of strong evidence. After all, the degree to which expert and statistical evidence met the criteria of a relevant field of expertise and a large sample size respectively did not influence the persuasiveness of these evidence types for the French participants. These findings are interesting, since other experimental studies on experts have shown that sources with a relevant field of expertise were more persuasive than sources with an irrelevant field of expertise (e.g., Luchok & McCroskey, 1978; Maddux & Rogers, 1980). The French insensitivity to evidence quality can be attributed to the small differences perceived in the instantiations of strong and weak (expert and statistical) evidence. The insensitivity of the quality of expert evidence was of particular interest. RWA, the context variable introduced in this study, did not account for this result, since it was not reliable (cf. Study 5), and since the Dutch and French participants scored similarly on it. Another explanation – the high status of the professors used in normatively strong expert evidence – was tested in Study 5.

Finally, an interaction effect in Study 3 – strong statistical evidence was more persuasive than strong expert evidence for the Dutch participants, but the two types of evidence were equally persuasive to the French participants – was not replicated in Study 4. This may be explained by the design of Study 4: participants received both strong and weak instantiations of evidence. Whereas the Dutch participants appeared to polarize between strong and weak evidence, the French participants had difficulties in differentiating between strong and weak instantiations. This polarization effect may also explain why strong expert evidence tended to be more persuasive to the Dutch participants than to the French participants (a result contradictory to Study 3).

Study 5

As mentioned above, the equal persuasiveness of strong and weak expert evidence for French participants in Study 4 was attritubed to the small difference in the

perceived expertise of experts with a relevant and an irrelevant field of expertise. The field of expertise of an expert apparently is not important for French participants (cf. Breton, 2003). Power distance suggests that this effect can be explained by the high status of the experts, who were professors. However, Study 5 showed that the same small difference between the perceived expertise of strong and weak experts also applied to researchers without a professor's title, who have a lower status. Explanations for this result will be given in Chapter 7, where the relationships between power distance and the cultural differences found in the present dissertation will be discussed.

Notes

1. I did not use sample sizes that were all larger than 300, because for claims that participants are unfamiliar with, such as basil consumption leading to better sporting performance, a study conducted with a large group of participants may be perceived as unrealistic.
2. As different groups were used in each culture, I tested whether there was an interaction effect between group and evidence type on the persuasiveness. For France, there was no interaction effect ($F < 1$), but for the Netherlands, the interaction effect was significant (F (24, 888) = 1.85, $p < .01$, $\eta^2 = .05$). Therefore, the relative persuasiveness of evidence types was analyzed for each Dutch group. In four of five groups with a sufficient number of participants, the order from the least persuasive type of evidence to the most persuasive type of evidence was the same: statistical evidence was more persuasive than expert and causal evidence, which both were more persuasive than anecdotal evidence. Therefore, all the groups were taken into account in the final analysis.
3. Another predictor of the persuasiveness of causal evidence is the degree to which people engage in 'causal thinking' (Anderson, 1983; see Section 3.2). Anderson used a number of measures for causal thinking that are appropriate for a single text, but inconvenient for 20 different claims.
4. I did not personally distribute the questionnaires in 3 of the 16 groups: Angers, Nijmegen (first group), and Toulouse.
5. I tested whether there were cultural differences for each of the claims without evidence. As four claims without evidence were presented in each version, the alpha level was corrected: $p = .05 / 4 = .0125$. (see footnote 2 in Chapter 5). The independent t-tests were computed with the standardized data, but the means below are from the raw data. The scores on the claims without evidence ranged from 1.84 to 3.84 for the Dutch participants, and from 1.76 to 3.75 for the French participants. There were two significant differences: claim (2) was more probable for the Dutch ($M = 3.44$, $SD = 0.98$) than for the French participants ($M = 2.88$, $SD = 0.99$), and claim (12) was more probable for the French ($M = 3.75$, $SD = 1.08$) than for the Dutch ($M = 3.07$, $SD = 1.28$).
6. The same effects were found with the raw data: (a) a main effect of type of evidence: F_1 (3, 596) = 34.18, $p < .001$, $\eta^2 = .15$; F_2 (3, 17) = 19.15, $p < .001$, $\eta^2 = .77$, (b) a tendency for a main effect of culture: F_1 (1, 598) = 3.48, $p = .06$; F_2 (1, 19) = 1.84, $p = .19$, and (c) an interaction effect between type of evidence and culture: F_1 (3, 596) = 5.13, $p < .01$, $\eta^2 = .03$; F_2 (3, 17) = 3.62, $p < .05$, $\eta^2 = .39$.

7. On the other hand, the context variables did have unhypothesized effects on the persuasiveness of certain types of evidence. For the French participants, the effect of PEI on statistical evidence was significant (F (1, 291) = 4.31, p < .05, η^2 = .02). PEI was correlated positively, but weakly with the persuasiveness of statistical evidence: r (295) = .11, p = .05. For the Dutch participants, there was an effect of NFC on statistical evidence (F (1, 300) = 9.27, p < .01, η^2 = .03) – they were weakly correlated in a positive way (r (304) = .15, p < .01) – and on expert evidence (F (1, 300) = 4.05, p < .05, η^2 = .01), but this correlation was not significant (r (304) = .09, p = .14). In addition, there were effects of PEI on anecdotal evidence (F (1, 300) = 17.01, p < .001, η^2 = .05) – resulting in a positive correlation (r (304) = .22, p < .001) – and on statistical evidence (F (1, 300) = 9.08, p < .01, η^2 = .03), leading to a weak, positive correlation: r (304) = .16, p < .01.
8. For claim (47), the original field of expertise 'clinical psychology' was changed into 'psychiatry', as 'clinical psychology' resembles too much other fields of psychology used in the material. As two pairs of claims, (22) – (25) and (37) – (19), were related to the same field, (19) was replaced by claim (24), and (25) by claim (10).
9. Dutch and French participants came from groups from different universities. I tested whether there was an interaction effect between group and evidence type or quality on the persuasiveness. For France, there was no interaction effect between group and evidence type, nor between group and evidence quality (both F's < 1). For the Netherlands, there was no interaction effect between group and evidence type (F < 1), nor between group and evidence quality (F (12, 287) = 1.24, p = .26). This means that, within each culture, the different groups of participants did not differ in their judgment of the relative persuasiveness of evidence types.
10. The positions of the fillers in the 20 claims were positions 1, 2, 5, 7, 10, 12, 13, 15, 17, and 20. Participants sometimes received consecutive evidence of the same type but with another quality, but there was always a filler between two such instantiations. The persuasiveness of these fillers was not computed, because they were linked to the same claims in each version, and because these claims were more probable than those with statistical and expert evidence, so that the persuasive effect of the fillers would be lower.
11. I again tested whether the scores on these claims without evidence were comparable between the two cultures. The scores on these claims ranged from 1.88 to 3.07 for the French participants, and from 1.83 to 2.98 for the Dutch participants (raw data). As two claims without evidence were presented in each version, the alpha level was corrected: p = .05 / 2 = .025 (see footnote 5 above). Independent t-tests computed with the standardized data showed that claim (11) was more probable for the French participants (M = 2.50, SD = 1.20) than for the Dutch participants (M = 2.02, SD = 0.98) (means in raw data).
12. The same effects were found with the raw data: (a) a main effect of type of evidence with an analysis by participants (F_1 (1, 598) = 7.91, p < .01, η^2 = .01), but a tendency for this effect with an analysis by stimuli (F_2 (1, 9) = 3.98, p = .08), (b) a main effect of quality: F_1 (1, 598) = 16.64, p < .001, η^2 = .03; F_2 (1, 9) = 8.71, p < .05, η^2 = .49, (c) a main effect of culture with an analysis by participants (F_1 (1, 598) = 5.37, p < .05, η^2 = .01), but not with an analysis by stimuli (F_2 (1, 9) = 2.19, p = .17), (d) no interaction effect between culture and evidence type: F_1 (1, 598) = 1.04, p = .31; F_2 (1, 9) = 1.09, p = .32, (e) no interaction effect between evidence type and quality: F_1 (1, 598) = 2.23, p = .14; F_2 (1, 9) = 1.57, p = .24, (f) an interaction effect between culture and evidence quality: F_1 (1, 598) = 16.02, p < .001, η^2 = .03; F_2 (1, 9) = 26.13, p < .01, η^2 = .74, and (g) no interaction effect between culture, evidence type, and evidence quality: F_1 < 1; F_2 < 1.

13. With the median-split technique Dutch participants with high NFC were distinguished from those with low NFC. NFC did not affect the quality of statistical evidence, as normatively strong statistical evidence was more persuasive than the normatively weak variant for the each of the two groups (low NFC: F (1, 163) = 13.52, $p < .001$, $\eta^2 = .08$; high NFC: F (1, 135) = 8.13, $p < .01$, $\eta^2 = .06$).
14. The context variables did have unhypothesized effects on the persuasiveness of certain types of evidence. For the French participants, there was an effect of NFC on the persuasiveness of strong expert evidence (F (1, 296) = 5.36, $p < .05$, $\eta^2 = .02$), and on weak expert evidence (F (1, 296) = 6.12, $p < .05$, $\eta^2 = .02$). The correlations between NFC and strong expert evidence, and NFC and weak expert evidence were positive, but weak (strong: r (300) = .13, $p < .05$; weak: r (300) = .12, $p < .05$). For the Dutch participants, there was an effect of NFC on the persuasiveness of weak expert evidence (F (1, 296) = 6.44, $p < .05$, $\eta^2 = .02$) – the variables were weakly, positively correlated (r (300) = .15, $p < .01$) – and an effect of PEI on weak statistical evidence (F (1, 296) = 7.00, $p < .01$, $\eta^2 = .02$), with a similar weak, positive correlation: r (300) = .17, $p < .01$.
15. There was one exception, however. In each version, participants received six unique fields, and one field twice. In order to ensure that these two fields never succeeded each other, the first and the seventh claim were exchanged.
16. The Dutch participants' sex did not affect the perceived expertise of expert types (1) (t(102.49) = 1.48, $p = .14$), (2) (t(104) = .06, $p = .95$), (3) (t(104) = 0.21, $p = .84$), or (4) (t(104) = .56, $p = .58$). Next, the French participants' sex did not influence the perceived expertise of expert types (2) (t(110) = .28, $p = .78$), (3) (t(110) = 1.76, $p = .08$), or (4) (t(110) = 1.12, $p = .27$), but it did affect that of expert type (1) (t(110) = 2.25, $p < .05$). Male participants considered that this expert type had more expertise (M = 3.50, SD = 0.77) than the female participants (M = 2.93, SD = 0.97). As the percentage of French male participants was low (14.3%), the consequences of this effect were not considered as important.

Conclusion and discussion

In this chapter, I will recapitulate and discuss the findings of the studies presented in this dissertation. I will first answer the general and the specific research questions (7.1). In Section 7.2, I will discuss how the differences can be explained by culture, and what kinds of limitations pertain to the studies. On the basis of this discussion I will make suggestions for further research (7.3).

7.1 Conclusions

Although the need for research on the effect of culture on the persuasiveness of evidence has been underscored repeatedly (Allen & Preiss, 1997; Bradac, Sandell, & Wenner, 1979; Greene & Brinn, 2003; Reynolds & Reynolds, 2002), this research was never conducted. In this dissertation, therefore, I investigated whether culture affects the relative persuasiveness of anecdotal, statistical, causal, and expert evidence. This question was addressed for two cultures: the Dutch and the French culture. In this investigation, I specifically focused on expert evidence. The answers to the question as to whether expert evidence is more persuasive in France will be given in Section 7.1.2. The first section is concerned with cultural differences in the relative persuasiveness of the four types of evidence.

7.1.1 General research questions

I first investigated how persuasive people from different cultures expect the four types of evidence to be:

General RQ1 Are there cultural differences in the relative *expected* persuasiveness of anecdotal, statistical, causal, and expert evidence?

Studies 1 and 2 showed that there indeed were cultural differences in expected persuasiveness. In Study 1, expected persuasiveness was measured on the basis of

Dutch and French text writers' use of the four types of evidence in persuasive public information brochures. There were cultural differences in the relative use of these evidence types. In particular, expert evidence was more frequently employed in the French corpus than in the Dutch corpus, and there was a tendency for a similar effect for anecdotal evidence. In Study 2, Dutch and French participants ranked the four types of evidence in relation to their expected persuasiveness as support for a number of claims. There was no cultural difference in the pattern of rankings of the four types of evidence. However, there were two more subtle results that indicated cultural differences. First, statistical evidence was ranked higher by the Dutch participants than by the French participants, and anecdotal evidence was ranked higher by the French participants than by the Dutch participants. Second, the mean number of times that anecdotal and expert evidence were put in first position was higher for the French than for the Dutch participants, whereas statistical evidence was put in first position more frequently by the French than by the Dutch participants.

If people from different cultures have other expectations about the persuasiveness of different types of evidence, the question arises whether there are also cultural differences in the relative *actual* persuasiveness of these evidence types:

General RQ2 Are there cultural differences in the relative *actual* persuasiveness of anecdotal, statistical, causal, and expert evidence?

In Study 3, actual persuasiveness was measured on the basis of the probability assessment of a number of claims supported by different evidence types. Where possible, these evidence instantiations were normatively strong. The relative persuasiveness of the four evidence types was different in the Dutch and the French culture. In particular, statistical and causal evidence were more persuasive to the Dutch than to the French participants. A comparison of the expected and actual persuasiveness of evidence types in Studies 2 and 3 shows that the expectations that the Dutch and French participants had about the persuasiveness of the evidence types were quite accurate. The ranking of the expected persuasiveness of evidence types in Study 2 corresponded fully to the relative persuasiveness of evidence types for the Dutch participants, and to a high degree for the French participants in Study 3[1].

Taken together, the results from these studies demonstrate that cultures may differ in the relative expected and actual persuasiveness of the four types of evidence. It should be noted that the differences were not particularly large,

especially with regard to expected persuasiveness. In Section 7.2.1, I will give a tentative explanation for these cultural differences. Parallel to the two general research questions, two specific research questions about expert evidence were formulated. The answers to these questions will be discussed in the next section.

7.1.2 Specific research questions

In Chapter 1, it was suggested that the persuasiveness of expert evidence could be related to the extent that receivers of expert evidence accept that the expert possesses more knowledge about the topic in question than they do. Therefore, expert evidence could be more persuasive in large power distance cultures than in small power distance cultures, and thus, more persuasive in the French culture than in the Dutch culture (Hofstede, 1980, 2001; Koopman, et al., 1999). Similar to the distinction between expected and actual persuasiveness in the general research questions, the first specific research question was:

Specific RQ1 Is the *expected* persuasiveness of expert evidence higher for the French than for the Dutch?

Study 1 showed that expert evidence was hardly used in the Dutch corpus, and more frequently – but still only moderately – in the French corpus. The more frequent use of expert evidence is an indicator that the expected persuasiveness of expert evidence is higher in the French culture. Study 2 demonstrated that Dutch and French students differed in the expected persuasiveness they assigned expert evidence, but only in a subtle way. There was no cultural difference in the mean ranking of expert evidence, but the French participants more frequently expected expert evidence to be the most persuasive type of evidence than the Dutch participants. In addition to *expected* persuasiveness, other studies were conducted to test this specific research question about *actual* persuasiveness:

Specific RQ2 Is the *actual* persuasiveness of expert evidence higher for the French than for the Dutch?

In Study 3, normatively strong expert evidence was not more persuasive to the French than to the Dutch participants in an absolute way, but it was in a relative way. In fact, it was among the most persuasive types of evidence for the French participants, whereas it only took an intermediate position in the relative persuasiveness of the types of evidence for the Dutch participants. The actual

persuasiveness of normatively strong and normatively weak expert evidence was investigated in Study 4. Normatively weak expert evidence, consisting of professors with an irrelevant field of expertise, was not more persuasive in the French culture than in the Dutch culture in an absolute way. However, it was more persuasive to the French participants in a relative way: both types of expert evidence were equally convincing for the French participants, whereas strong expert evidence was more persuasive than weak expert evidence for the Dutch participants. In sum, the differences concerning expert evidence can be summarized as follows:

1. French people expect expert evidence to be somewhat more persuasive than Dutch people (Studies 1 and 2).

2. Expert evidence is more persuasive in the French culture than in the Dutch culture in a relative way (Studies 3 and 4).

3. Normatively strong expert evidence is more persuasive than normatively weak expert evidence for the Dutch, but the two types of evidence are equally persuasive for the French (Study 4).

In conclusion, the four studies demonstrate at a general level that there are cultural differences in the relative expected and actual persuasiveness of evidence types in the Dutch and the French culture. At a specific level, they show that the relative expected and actual persuasiveness of expert evidence is different in the Dutch than in the French culture. In the next section, these studies will be discussed further.

7.2 Discussion

The discussion addresses the following two questions: how can the cultural differences be explained (7.2.1), what are the limitations of the studies (7.2.2)?

7.2.1 Effects of culture

Besides the differences concerning expert evidence, two other results were found. First, the actual persuasiveness of evidence, regardless of its type, was higher for the Dutch than for the French participants. Second, the expected and actual

persuasiveness of statistical evidence, and the actual persuasiveness of causal evidence were higher in the Dutch than in the French culture. These results could not be attributed to context variables measured at the level of individual participants. A tentative explanation for the pattern of results regarding the persuasiveness of the four types of evidence lies in the research traditions in the Dutch and the French culture. Each type of evidence can be related to a form of research. In fact, Hoeken and Hustinx (2003b) pointed out that there is an analogy between basic forms of scientific research on the one hand (the experiment, the survey, and the case study), and evidence types on the other. The experiment is conducted to determine possible causal relations between factors, and thus leads to causal evidence. The survey method is used to question a great number of subjects, and therefore results in statistical evidence. The last method, the case study, is qualitative in a sense that it aims to describe as detailed as possible a case or phenomenon (anecdotal evidence). Although there is no research method that leads to expert evidence, source citation in articles and books can be considered as a form of expert evidence.

A parallel between the research traditions in the Dutch and the French culture, and the relative persuasiveness of the evidence types is particularly appropriate, given that the participants in the experiments were students, who are exposed to research traditions intensively. Very few studies have been conducted on research traditions in various cultures. One of these is the essay that Galtung (1981) wrote about intellectual styles. He distinguished the saxonic (e.g., US, UK), teutonic (e.g., German), gallic (e.g., France), and nipponic (e.g., Japan) style. Although Galtung speculates that the Netherlands is influenced by the first three styles, it appears to me that the saxonic style is dominant. This intellectual style is characterized by weak theory formation, and by a strong focus on data collection in order to investigate smaller parts of a theory that in the end may lead to a more comprehensive theory. In this tradition, experiments and surveys are essential. It comes as no surprise, then, that causal and statistical evidence were among the most persuasive types of evidence for the Dutch participants in the present studies. Galtung argues that the gallic style, typical for France, is characterized by very strong theory formation, for which data are not necessary as a back up. French scholars develop and refine theories that are never tested empirically. At their best, data serve as illustration (cf. anecdotal evidence). Perhaps correspondingly, the actual persuasiveness of causal and statistical evidence was lower in the French culture than in the Dutch culture in the present studies. In the French intellectual style, the relationships between the scholar's theory and those of others are

important. References to other scholars (cf. expert evidence) may be frequently used either to emphasize the relevance of own ideas or to disconfirm competing theories and ideas. In conclusion, even if the links between the types of evidence and the styles distinguished in Galtung can not be documented, this parallel provides a possible account for the relative persuasiveness of the types of evidence in the Dutch and the French culture.

As far as expert evidence is concerned, it was suggested in Chapter 1 that this could be more persuasive to French than to Dutch people because of a difference in power distance between the two cultures (Hofstede, 1980, 2001; Koopman, et al., 1999). Below, I will discuss five possible explanations – other than Galtung's (1981) intellectual styles – for the cultural differences in the persuasiveness of expert evidence. The overall pattern of results on expert evidence may be explained by differences in power distance and uncertainty avoidance. Some specific results may be explained by the importance of cultural versus personal norms, by differences in educational systems in the two cultures, and by the route to persuasion (central or peripheral).

Power distance

As I argued in Chapter 4, in order to determine the effect of power distance, this value dimension should be measured at the level of the participants in the experiments. Personal values related to power distance were not measured, but a context variable was constructed, the Preference for Expert Information scale (PEI). It was expected that the French participants would score higher on this scale than the Dutch participants. No cultural differences in PEI were found in Studies 2, 3, and 4, but the Dutch participants scored higher on the PEI scale than the French participants in Study 5. Also, it was expected that the score on this scale would affect the actual persuasiveness of expert evidence. The relationship between the two variables was assessed in Studies 3 and 4. For the Dutch participants, there was a significant, positive correlation between their PEI and the actual persuasiveness of expert evidence. For the French participants, this correlation was not significant. In Study 2, however, PEI correlated positively with the expected persuasiveness of expert evidence for the French participants, but not for the Dutch participants. Another scale, the Right-Wing Authoritarianism (RWA) scale, was administered in Studies 4 and 5. There were no cultural differences in the scores on this scale in either study. In Study 4, RWA had a marginal effect on the persuasiveness of strong expert evidence for the Dutch participants, but not for the French participants.

Taken together, there were more cultural similarities than differences in the scores on the scales related to power distance. Cultural differences found in the persuasiveness of expert evidence were smaller than the differences in power distance, as reported in Hofstede (1980, 2001), suggest. Also, the status of the expert did not affect the perception of the expertise of strong and weak experts for the French participants (Studies 4 and 5) – a finding in apparent contradiction with power distance. All these findings may lead one to believe that power distance does not provide a solid, plausible explanation for the cultural differences in the persuasiveness of expert evidence.

Uncertainty avoidance

The Dutch and the French culture also differ in their scores on Hofstede's (2001) dimension of uncertainty avoidance, which is "The extent to which the members of a culture feel threatened by uncertain and unknown situations" (p. 161). Uncertainty avoidance is higher in the French culture than the Dutch culture. Differences observed in the relative persuasiveness of expert evidence may therefore be explained by uncertainty avoidance. In fact, people from high uncertainty avoidance cultures are said to rely more on experts than people from low uncertainty avoidance cultures, because experts can be used to reduce uncertainty (Hofstede, 1980, 2001). As people from high uncertainty avoidance cultures rely more on experts, they may therefore be more persuaded by expert evidence than people from low uncertainty avoidance cultures. French participants could therefore be expected to have a higher preference for expert information (PEI) than the Dutch participants, but this was not the case. Similar to the value dimension of power distance, no empirical support for an effect of uncertainty avoidance was found in the present study. There are three other possible explanations for more specific results. These explanations are discussed below.

A personal or a cultural norm

As mentioned above, the actual persuasiveness of normatively strong expert evidence was related to the Dutch participants' preference for expert information, but not to that of the French participants (Studies 3 and 4). This suggests that the persuasiveness of expert evidence is a personality related norm for the Dutch participants, and a cultural norm for the French participants. A study by Kagitçibasi (1970), where the determinants of respect for authority of participants from the US (small power distance) and Turkey (large power distance) were compared, supports this suggestion. The Turkish participants had a great respect

for authorities regardless of their score on 'core authoritarianism', a measure that is related to power distance. The American participants, on the other hand, only had a great respect for authorities if these participants scored high on core authoritarianism. Kagitçibasi came up with a social norm explanation. Authoritarian attitudes are a matter of personal preference for people in small power distance cultures, whereas they are a social norm for people in large power distance cultures. In terms of Studies 3 and 4, the persuasiveness of expert evidence might be a personal matter for the Dutch participants, and a cultural matter for the French participants. In the French culture, expert evidence may be persuasive because it is a cultural norm.

Educational system

The fact that strong and weak expert evidence were equally persuasive to the French participants in Study 4 was attributed to the small difference in the perceived expertise of professors with a relevant and an irrelevant field of expertise. This small difference could not be explained by the experts' status (Study 5). A plausible explanation lies in the educational systems in France and the Netherlands. The Dutch findings mirrorred normative considerations: a relevant field of expertise is important for the persuasiveness of expert evidence. The French researcher De Bony (2003) notes that Dutch children on elementary schools are already incited to adopt critical attitudes towards the texts that they are learning to read. In this light, it is not surprising to find that the Dutch participants were not convinced by experts with an irrelevant field of expertise. Such experts simply are not considered competent. In the French educational system, on the other hand, children are not educated to develop such critical attitudes. French pupils are not supposed to learn themselves, they can only learn because of teachers, who are considered omniscient (e.g., Gruère & Morel, 1991; Planel, 1997). Since the 15th century, French philosophers such as Rabelais have insisted on the need for omniscient and erudite instructors. Today, French teachers at high school still are such instructors who are expected to spread a *culture générale* (Blom, 1995). In such a system of erudite teachers and dependent pupils, it is understandable that the French students in Studies 4 and 5 accorded the professors and researchers quite a high level of expertise on a domain that is not their field of expertise. Professors and researchers are not easily attacked on the irrelevance of their fields of expertise in relation to a given topic.

Peripheral or central route to persuasion

Finally, the lack of an effect of evidence quality on the persuasiveness of expert evidence for the French participants in Study 4 could also be explained through the peripheral route that these participants may have taken. People that use this route are said not to take into account the quality of arguments (Petty & Cacioppo, 1986). In fact, strong and weak expert evidence were equally persuasive for the French participants (and the same was true for strong and weak statistical evidence). As suggested in Chapter 6, the French participants could have used the heuristic 'There is an expert source, so the claim must be probable'. In contrast, however, the moderate scores on NFC suggest that these participants may have taken the *central* route. Also, the French participants in Study 3 did differentiate between the persuasiveness of different types of evidence. It could therefore be concluded that there is no strong support for the use of either the central or the peripheral route.

Conclusion

The goal of this section was to discuss how the cultural differences in the relative expected and actual persuasiveness of expert evidence in the Dutch and French culture could be explained. As the scales constructed to measure power distance did not differ between the two cultures, it is impossible to empirically attribute these differences to power distance. However, it should be noted that if context variables had not been taken into account in the present study, power distance would have been a plausible explanation. This stresses the importance of measuring context variables or values at the level of participants in studies on cultural effects. Without context variables, power distance would have been more plausible, because all explanations (except the route to persuasion) are related to this value dimension. As Turkey and the US in the Kagitçibasi (1970) study are said to differ in power distance in a similar way as France and the Netherlands, this study suggests that power distance may be related to the finding that the persuasiveness of expert evidence was related to PEI for Dutch participants, but not for the French participants. Next, educational systems are a natural part of culture as an overall system (Hofstede, 1980, 2001). In his discussion of the influence of culture on teacher – student interactions, Hofstede (1986) argued that teachers are not contradicted or criticized in large power distance cultures, but that students are allowed to contradict and criticize teachers in small power distance cultures. Finally, Hofstede (2001) showed that there is a very strong, positive correlation between power distance and uncertainty avoidance in the European region.

In sum, there are a number of possible explanations for the subtle cultural differences found in the present study. Whereas most of the explanations (indirectly) appear to point to power distance as an underlying dimension to which the results may be attributed, empirical evidence suggests that power distance is not an explanation. The concept of power distance itself may account for the difficulties in measuring cultural differences in power distance at an individual level. People in large power distance cultures deal with inequality of power in an ambiguous way. Hofstede (2001, p. 96), for instance, remarked that "in high-PDI [power distance] countries where [...] [inequality] is, in practice, expected and desired by the less powerful, it remains problematic at the same time". In the French culture, there is an ambiguous attitude towards authorities, which is recognized by the French themselves, and by outsiders studying the French culture. The French anthropologist d'Iribarne (1989, p. 77) notes that "The often strongly emotional character of hierarchical relationships in France is intriguing. There is an extreme diversity of feelings towards superiors: They may be either adored or despised with equal intensity" (translation from Hofstede, 2001). Demangeat and Molz (2003) – in their comparison of the French and German culture – reach the same conclusion: the French attitude towards authorities varies from great appreciation to total denial. In conclusion, future research is needed to solve these complexities of unraveling the impact of power distance and alternative explanations of the differences in persuasiveness of (expert) evidence found in the present study.

7.2.2 Limitations

The experimental studies reported in this dissertation have several limitations that restrict the generalizability of their results. These limitations are related to the material, to the selection of the participants, and to the context in which the experiments were conducted.

Concerning the material, the claims were moderately probable in order to rule out the negative influence of extreme (im)probability on the effect of evidence. As a consequence, the results are limited to these kinds of claims. In reality, however, persuaders also aim to convince people of claims that are perceived as highly probable (e.g., smoking leads to a higher risk of getting lung cancer). Next, the choice for claims and evidence without a context also has a drawback with respect to generalizability. The claims with evidence were presented in two short sentences. An advantage of this operationalization is that differences in persuasiveness can be better attributed to evidence, and that differences in

evidence quality can be perceived more easily by the participants. A disadvantage of this operationalization is that claims and supporting evidence are usually part of a longer text in real life argumentation. People may react differently to evidence when it is presented in a richer context than when it is presented in isolation with a claim.

A second series of limitations is related to the selection of the participants. The conclusions now only hold for students. It could be argued that the persuasiveness of expert evidence in particular was affected by the fact that the participants were students who are regularly exposed to the experts used in the experiments: professors and researchers. Moreover, they were mostly students of Arts, who are more likely to have had a course on argumentation than other kinds of students. The persuasiveness of expert evidence may indeed be different when other participants are involved. However, it should be noted that the type of participants used in the present studies has not influenced the cultural *comparison* of the persuasiveness of expert evidence. In fact, the internal validity was high because the Dutch and French participants were highly similar in education, age, and sex.

Finally, the context in which the experiments were conducted may also have consequences for the degree to which the results of the studies can be generalized. Within-subjects designs were chosen in the experiments. Each Dutch or French participant was compared with him- or herself, as he or she was exposed to all the experimental conditions, such as anecdotal, statistical, causal, and expert evidence in Studies 2 and 3, and to strong and weak evidence in Study 4. This direct comparison is an advantage of the within-subjects design over the between-subjects design, in which participants only receive one condition. A disadvantage of the within-subjects design is a carry over effect: as participants are given different conditions, the response to condition X may affect the reaction to condition Y. More concretely, the participants' judgments of evidence types in the first part of the questionnaire may influence their judgments in the second part. A carry over effect only occurred for the Dutch participants in Study 5, but it was rather small. Still, participants may have become more aware of the goal of the experiment than when a between-subject design had been used, because they were exposed to the different experimental conditions. Another limitation related to the experimental context is the presentation format used in the experiments. In the paper-and-pencil experiments, participants have more time to think than in real life situations when persuasive messages are directed at them. In written (as opposed to oral) questionnaires, participants may more easily give a neutral opinion (i.e., the midpoint of 5-point semantic differential). The limitations on the

generalizability of the results open the door to suggestions for further research. These suggestions will be presented in the next section.

7.3 Suggestions for further research

In this section, I will outline a few directions for further research that naturally follow the discussion of the effects of culture (7.2.1), and the limitations of the experiments (7.2.2). The major problem encountered in determining the effect of culture was that the context variables could not account for the cultural differences in the persuasiveness of evidence types, and of expert evidence in particular. Other context variables should be selected and/or constructed in order to measure the effect of culture. In this process, focus groups could be used in order to raise the understanding of what determines the persuasiveness of evidence types. In a focus group, discussion among group members is stimulated through structured interviewing techniques (Morgan, 1988). Focus groups were, for instance, used in Reinard (2003). In this study, participants were asked to explain why they had marked some evidence instantiations as expectedly high or low on persuasion. This research method may provide valuable information about why people find expert evidence persuasive or not. In a focus group, members can be stimulated to discuss the role of expert evidence in their own culture. Such a discussion may further investigate the Kagitçibasi (1970) explanation by sorting out whether high persuasiveness of expert evidence is a personal or a cultural matter. The context in which participants can freely discuss matters with each other is also more similar to real life than the laboratory setting of the paper-and-pencil experiments. In sum, the use of focus groups may be effective in resolving problems discussed in 7.2.1, and in canceling out limitations of the experimental studies.

A second suggestion for further research would also help solve problems mentioned earlier. In Study 1, expert evidence was the least popular evidence type. One of the explanations for this finding was that text writers, and the French in particular, may think that they do not need further support from other experts, as they are experts themselves. This explanation could be tested by asking students to write an essay about a topic that they are not knowledgeable about. In support of their claims, they would then have to use evidence types that would be provided beforehand in order to ensure a fair comparison of the evidence types. This design looks like Study 2, where students were asked to rank evidence types in terms of their expected persuasiveness. The proposed study, however, is interesting to

conduct, as the task (writing an essay) is more concrete than the task in Study 2. In this way, the persuasiveness of evidence types can be studied in a context that is more similar to real life persuasion.

The route to persuasion (central or peripheral) was suggested as a possible explanation for results of Study 4. Without strong support for the use of either route, a third suggestion for further research is to set up an experiment in which participants can be expected to take ELM's central route to persuasion. A review of Heckler and Childers (1992) demonstrated that people take this route when they are faced with incongruent information. Therefore, Dutch and French participants could be given incongruent information in support of a claim: positive (negative) statistical evidence from a Consumer's Guide, and negative (positive) expert evidence. In different versions, statistical and expert evidence are both normatively strong (weak), or one of them is normatively strong and the other normatively weak.

Limitations mentioned in 7.2.2 can easily be taken into account in further research. Instead of simple texts with claims and evidence, more realistic texts, such as essays, advertisements, or fundraising texts could be employed in an experimental setting. In a similar vein, experiments with between-subjects design could be conducted in order to avoid possible problems related to the carry over effect. Special attention could be paid to the selection of participants. Of course, it is recommended that people other than (Arts) students participate in such experiments. Also, it is important to select other cultures than the Dutch and the French. In this dissertation, the Dutch and the French cultures were selected to represent a large and a small power distance culture respectively. As these cultures also differed in uncertainty avoidance, a comparison of cultures that only differ in power distance is recommended. However, a close look at Hofstede's (1980, 2001) scores of European countries shows that it is quite hard to find European countries that differ in power distance, but not on the three other value dimensions. Italy and Germany are relatively similar with respect to the three other dimensions, but their scores on power distance are not very different (54 and 35 respectively). Another way of selecting cultures is suggested by Van de Vijver and Leung (1997), who argue that the more cultures are included in a study, the more reliable the conclusions are about the effect of culture. In order to further investigate the effect of power distance on the relative expected and actual persuasiveness of expert evidence, then, studies could be conducted in several European cultures that are said to have a large (e.g., Belgium, Greece, Italy, and Spain) or a small power distance (e.g., Denmark, Germany, Great Britain, Sweden, and Switzerland),

regardless of their scores on the other dimensions. An advantage of this procedure is that it would also allow testing across cultures of the Kagitçibasi (1970) suggestion of a personal versus a cultural norm for the persuasiveness of expert evidence.

7.4 Epilogue

The goal of this dissertation was to investigate whether there are cultural differences in the relative expected and actual persuasiveness of anecdotal, statistical, causal, and expert evidence. A number of studies that examined reasoning processes of cultures influenced by ancient Greece (e.g., Europe, United States), and of cultures affected by ancient the Chinese culture (e.g., China, Japan) have demonstrated that culture may influence the ways in which people process and use information to make decisions (Nisbett, 2003; Nisbett, Peng, Choi, & Norenzayan, 2001). In this dissertation, two cultures were selected that are highly similar: the Dutch and the French culture. These cultures have the same Greek roots of reasoning and argumentation. Finding cultural differences, however small, in the relative persuasiveness of evidence types between two such similar cultures provides strong support for the impact of culture. Not only was the relative persuasiveness of evidence types different for the Dutch and French participants, one of the studies also demonstrated that the two cultures differed in the importance of norms for the persuasiveness of strong evidence. Whether normatively strong evidence is indeed more persuasive than normatively weak evidence appears culture-dependent. Together with the results of the other studies in this dissertation, this finding may stimulate further research on the impact of culture on the relative persuasiveness of evidence types and, more generally, on argument quality.

Notes

1. The French participants in Study 2 expected statistical evidence to be the most persuasive type of evidence, but to the other French participants in Study 3 expert evidence was as persuasive as statistical evidence in practice. Anecdotal and causal evidence were the least persuasive types of evidence in both studies.

References

Aaker, J., & Maheswaran, D. (1997). The impact of cultural orientation on persuasion. *Journal of Consumer Research, 24*, 315-328.

Abelson, R. P., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology: Vol. 1.* (pp. 231-309). New York: Random House.

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299-352.

Ajzen, I. (1977). Intuitive theories of events and the effects of base rate information on prediction. *Journal of Personality and Social Psychology, 35*, 303-314.

Aldrich, A. (2003). Considering culture in the analysis of arguments. 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. 3-7). Amsterdam: Sic Sat.

Allen, M., Bruflat, R., Fucilla, R., Kramer, M., McKellips, S., Ryan, D. J., et al. (2000). Testing the persuasiveness of evidence: Combining narrative and statistical forms. *Communication Research Reports, 17*, 331-336.

Allen, M., & Preiss, R. W. (1997). Comparing the persuasiveness of narrative and statistical evidence using meta-analysis. *Communication Research Reports, 14*, 125-131.

Altemeyer, B. (1981). *Right-Wing Authoritarianism.* Winnipeg, Canada: University of Manitoba Press.

Altemeyer, B. (1988). *Enemies of freedom: Understanding Right-Wing Authoritarianism*. San Francisco: Jossey-Bass.

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19*, 93-108.

Anderson, C. A., & Kellam, K. L. (1992). Belief perseverance, biased assimilation, and covariation detection: The effects of hypothetical social theories and new data. *Personality and Social Psychology Bulletin, 18*, 555-565.

Anderson, C. A., Lepper, M. R, & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037-1049.

Areni, C. S. (2002). The proposition-probability model of argument structure and message acceptance. *Journal of Consumer Research, 29 (2)*, 168-187.

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.

Aristotle (1932). *The rhetoric of Aristotle (L. Cooper, trans.)*. New York: Appleton-Crofts.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extreme: Black-white differences in response styles. *Public Opinion Quarterly, 48*, 491-509.

Baesler, E. J. (1997). Persuasive effects of story and statistical evidence. *Argumentation and Advocacy, 33*, 170-175.

Baesler, E. J., & Burgoon, J. K. (1994). The temporal effects of story and statistical evidence on belief change. *Communication Research, 21*, 582-602.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*, 211-233.

Berthelot-Guiet, K., & Ollivier-Yaniv, C. (2001). "Tu t'es vu quand t'écoutes l'état": Réception des campagnes de communication gouvernementale, appropriation et détournement linguistique des messages [Reception of governmental communication campaigns, appropriation and linguistic adjustment of messages]. *Réseaux, 19*, 155-178.

Blom, S. (1995). *Intellectuele vorming in Nederland en Frankrijk: Een vergelijkend onderzoek naar deelname aan intellectuele vorming in het Franse en Nederlandse voortgezet onderwijs [Intellectual formation in the Netherlands and France: A comparative study of the participation in intellectual formation in Dutch and French secondary education].* Groningen, The Netherlands: Wolters-Noordhoff.

Bohner, G., Ruder, M., & Erb, H.-P. (2002). When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology, 41*, 495-519.

Bony, J. de (2003). Individual autonomy and socialization at the Dutch elementary school: The relationship between the individual and the group. In W. Jochems, et al. (Eds.), *Grenzeloos leren: Proceedings van de 30e Onderwijs Research Dagen 2003* (pp. 110-112). Heerlen: Open Universiteit Nederland.

Booth-Butterfield, S., & Welbourne, J. (2002). The Elaboration Likelihood Model: Its impact on persuasion theory and research. In J. P. Dillard, & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 155-173). Thousand Oaks, CA: Sage Publications.

Bradac, J. J., Sandell, K. L., & Wenner, L. A. (1979). The phenomenology of evidence: Information-source utility in decision making. *Communication Quarterly, 27*, 35-46.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573-604.

Breton, P. (2003). *L'argumentation dans la communication [Argumentation in communication] (3rd ed.)*. Paris: La Découverte.

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis, & J. W. Berry (Eds.), *Handbook of cross-cultural psychology: Methodology* (pp. 389-444). Boston: Allyn & Bacon.

Brosius, H.-B., & Bathelt, A. (1994). The utility of exemplars in persuasive communications. *Communication research, 21*, 48-78.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

Cathcart, R. S. (1955). An experimental study of the relative effectiveness of four methods of presenting evidence. *Speech Monographs, 22*, 227-233.

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium: Vol. 5.* (pp. 3-40). Hillsdale, NJ: Lawrence Erlbaum.

Cox, D., & Cox, A. D. (2001). Communicating the consequences of early detection: The role of evidence and framing. *Journal of Marketing, 65*, 91-103.

Darke, P. R., Chaiken, S., Bohner, G., Einwiller, S., Erb, H.-P., & Hazlewood, J. D. (1998). Accuracy motivation, consensus information, and the law of large numbers: Effects on attitude judgment in the absence of argumentation. *Personality and Social Psychology Bulletin, 24*, 1205-1215.

Demangeat, I., & Molz, M. (2003). Frankreich [France]. In A. Thomas, E.-U. Kinast, & S. Schroll-Machl (Eds.), *Handbuch Interkulturelle Kommunikation und Kooperation: Vol. 2. Länder, Kulturen und interkulturelle Berufstätigkeit* (pp. 24-52). Göttingen, Germany: Vandenhoeck & Ruprecht.

Dickson, P. R. (1982). The impact of enriching case and statistical information on consumer judgments. *Journal of Consumer Research, 8*, 398-406.

Doosje, B., Spears, R., & Koomen, W. (1995). When bad isn't all bad: Strategic use of sample information in generalization and stereotyping. *Journal of Personality and Social Psychology, 69*, 642-655.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71*, 5-24.

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., & Snoeck Henkemans, F. (1997). *Handboek argumentatietheorie [Handbook argumentation theory]*. Groningen, The Netherlands: Martinus Nijhoff.

Ehninger, D., & Brockriede, W. (1963). *Decision by debate*. New York: Dodd, Mead.

Elifson, K., Runyon, R. P., & Haber, A. (1998). *Fundamentals of social statistics (3rd ed.)*. Boston: McGraw-Hill.

Elmes, D. G., Kantowitz, B. H., & Roediger, H. L. (1992). *Research methods in psychology (3rd ed.)*. St Paul, MN: West.

Enschot, R. Van Dijk-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, The Netherlands: Kluwer.

Facione, P. A., & Scherer, D. (1978). *Logic and logical thinking*. New York: McGraw-Hill.

Falcy, S. (1997). *Les processus d'influence publicitaire: Un essai de modélisation intégrant implication vis à vis de la publicité, caractéristiques psychologiques et publicitaires [The processes of influencing through advertising: An attempt at model formation integrating attitude towards the ad, and psychological and advertising characteristics]*. Unpublished dissertation, Université Pierre Mendès France, Grenoble, France.

Feather, N. T. (1998). Reactions to penalties committed by the police and public citizens: Testing a social-cognitive process model of retributive justice. *Journal of Personality and Social Psychology, 75*, 528-544.

Feteris, E. T. (2002). Pragmatic argumentation in a legal context. In F. H. van Eemeren (Ed.), *Advances in pragma-dialectics* (pp. 243-260). Amsterdam: Sic Sat / Newport News, VA: Vale Press.

Fishbein, M., & Ajzen, I. (1981). Acceptance, yielding and impact: Cognitive processes in persuasion. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 339-359). Hillsdale, NJ: Lawrence Erlbaum.

Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology: Vol. 2.* (4th ed., pp. 915-981). Boston: McGraw-Hill.

Fitch, K. L. (2003). Cultural persuadables. *Communication Theory, 13*, 110-123.

Freeley, A. J. (1976). *Argumentation and debate: Critical thinking for reasoned decision making (4th ed.)*. Belmont, CA: Wadsworth.

French, J. R. P., & Raven, R. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Ann Arbor, MI: Institute for Social Research.

Galtung, J. (1981). Structure, culture, and intellectual style: An essay comparing saxonic, teutonic, gallic and nipponic approaches. *Social Sciences Information, 20*, 817-856.

Garssen, 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.

Gilovich, T., & Griffin, D. (2002). Introduction – heuristics and biases: Then and now. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 1-18). Cambridge, England: Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, England: Cambridge University Press.

Ginossar, Z., & Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology, 16,* 228-242.

Gisbergen, M. S. van, Ketelaar, P. E., & Beentjes, H. (2004). Changes in advertising language? A content analysis of magazine advertisements in 1980 and 2000. In P. Neijens, C. Hess, B. van den Putte, & E. Smit (Eds.), *Content and media factors in advertising* (pp. 22-37). Amsterdam: Spinhuis.

Greene, K., & Brinn, L. S. (2003). Messages influencing college women's tanning bed use: Statistical versus narrative evidence format and a self-assessment to increase perceived susceptibility. *Journal of Health Communication, 8*, 443-461.

Gruère, J.-P., & Morel, P. (1991). *Cadres français et communications interculturelles: Comment sont-ils perçus? Comment se perçoivent-ils? Comment perçoivent-ils les autres? [French executives and intercultural communication: How are they perceived? How do they perceive themselves? How do they perceive others?]* Paris: Eyrolles.

Hamill, R., Wilson, T. D., & Nisbett, R. E. (1980). Insensitivity to sample bias: Generalizing from atypical cases. *Journal of Personality and Social Psychology, 39*, 578-589.

Han, S., & Shavitt, S. (1994). Persuasion and culture: Advertising appeals in individualistic and collectivistic societies. *Journal of Experimental Social Psychology, 30*, 326-350.

Harkness, J. A., Vijver, F. J. R. van de, & Mohler, P. Ph. (2003). *Cross-cultural survey methods*. Hoboken, NJ: John Wiley & Sons.

Harte, T. B. (1976). The effects of evidence in persuasive communication. *Central States Speech Journal, 72*, 42-46.

Heckler, S. E., & Childers, T. L. (1992). The role of expectancy and relevancy in memory for verbal and visual information: What is incongruency? *Journal of Consumer Research, 18*, 475-492.

Herk, H. van, Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*, 346-360.

Hesslow, G. (1988). The problem of causal selection. In D. J. Hilton (Ed.), *Contemporary science and natural explanation: Commensense conceptions of causality* (pp. 11-32). Brighton, England: Harvester Press.

Hilton, D. J., & Slugoski, B. R. (2001). Conversational processes in reasoning and explanation. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Vol. 1. Intraindividual processes* (pp. 181-206). Oxford, England: Basil Blackwell.

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. 55-65). Dordrecht, The Netherlands: 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. (2001c). Logos, ethos en pathos in sociaal-wetenschappelijk onderzoek [Logos, ethos, and pathos in social scientific research]. *Lampas, 34*, 424-439.

Hoeken, H. (2005). *Overtuigende taal [Persuasive language]*. Nijmegen, The Netherlands: Radboud Universiteit Nijmegen.

Hoeken, H., Brandt, C. van den, Crijns, R., Domínguez, N., Hendriks, B., Planken, B., et al. (2003). International advertising in Western Europe: Should

differences in uncertainty avoidance be considered when advertising in Belgium, France, the Netherlands and Spain? *Journal of Business Communication, 40*, 195-218.

Hoeken, H., & Hustinx, L. (2003a, May). *The relative persuasiveness of anecdotal evidence compared to statistical and causal evidence: Does the type of claim make a difference?* Paper presented at the 53th Annual Conference of the ICA, San Diego, CA.

Hoeken, H., & Hustinx, L. (2003b). The relative persuasiveness of different types of 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., & Korzilius, H. (2003). Conducting experiments on cultural aspects of document design: Why and how? *Communications, The European Journal of Communication Research, 28*, 285-304.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations, 10*, 301-320.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations (2nd ed.)*. Thousand Oaks, CA: Sage.

Hollihan, T. A., & Baaske, K. (1998). *Arguments and arguing: The products and process of human decision making*. Long Grove, IL: Waveland.

Hornikx, J. (2003). De relatieve frequentie van verschillende evidentietypen in Nederlandse en Franse persuasieve voorlichtingsbrochures [The relative frequency of different evidence types in Dutch and French persuasive public information brochures]. In L. van Waes, P. Cuvelier, G. Jacobs, & I. de Ridder (Eds.), *Studies in Taalbeheersing, volume 1* (pp. 206-217). Assen, The Netherlands: Van Gorcum.

Hornikx, J. (2004). Relative occurrence of evidence types in Dutch and French persuasive communication. In Ch. M. Schmidt, D. Neuendorff, & M. Nielsen (Eds.), *Marktkommunikation in Theorie und Praxis: Inter- und intrakulturelle Dimensionen in der heutigen Wirtschaft* (pp. 291-307). Wiesbaden, Germany: Deutscher Universitäts-Verlag.

Hornikx, J. (2005). De voorkeur voor expertevidentie in Nederland en Frankrijk [Differences in preference for expert evidence in the Netherlands and France]. *Toegepaste Taalwetenschap in Artikelen, 73*, 9-18.

Hornikx, J. (in press-a). 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. (in press-b). Measuring the effect of culture in experimental persuasive effects research. In R. Crijns, & J. Thalheim (Eds.), *Kooperation und Effizienz in der Unternehmenskommunikation: Inner- und außerbetriebliche Kommunikations-aspekte zur Corporate-Identity und Interkulturalität* (pp. 205-214). Wiesbaden, Germany: Deutscher Universitäts-Verlag.

Hornikx, J., & Hoeken, H. (2003, November). *De invloed van cultuur op de relatieve overtuigingskracht van evidentietypen [The influence of culture on the relative persuasiveness of evidence types]*. Paper presented at the Etmaal van de Communicatiewetenschap, Nijmegen, The Netherlands.

Hornikx, J., & Hoeken, H. (2005). Is expertevidentie overtuigender in Frankrijk dan in Nederland? [Is expert evidence more persuasive in France than in the Netherlands?] *Tijdschrift voor Taalbeheersing, 27*, 42-57.

Hornikx, J., & Starren, M. (2004). Publicités en France et aux Pays-Bas: Peut-on standardiser ou faut-il adapter? [Advertising in France and in the Netherlands: Should we standardize or adapt?] *Studies in Communication Sciences, 4*, 219-233.

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.

Iyengar, S., & Kinder, D. R. (1987). *News that matters*. Chicago: University of Chicago Press.

Iribarne, P. d' (1989). *La logique de l'honneur: Gestion des entreprises et traditions nationales [The logic of honour: Business administration and national traditions]*. Paris: Seuil.

Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research, 9*, 169-181.

Jansen, C. (1999). *Zo werkt dat: Het ontwerp van instructieve teksten [That's how it works: The design of instructive texts]*. Nijmegen, The Netherlands: Nijmegen University Press.

Johnson, T., Kulesa, P., Cho, Y.I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*, 264-277.

Jong, M. de, & Schellens, P. J. (2000). Meervoudige doelen en doelgroepen in voorlichtingsbrochures [Multiple goals and target groups in public information brochures]. In R. Neutelings, N. Ummelen, & A. Maes (Eds.), *Over de grenzen van de taalbeheersing* (pp. 263-272). Den Haag, The Netherlands: Sdu.

Kagitçibasi, C. (1970). Social norms and authoritarianism: A Turkish-American comparison. *Journal of Personality and Social Psychology, 16*, 444-451.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Kassin, S. M. (1979). Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology, 37*, 1966-1981.

Kazoleas, D. C. (1993). A comparison of the persuasive effectiveness of qualitative versus quantitative evidence: A test of explanatory hypotheses. *Communication Quarterly, 41*, 40-50.

Kellermann, K. (1980). The concept of evidence: A critical review. *Journal of the American Forensic Association, 16*, 159-172.

Kemmelmeier, M., Burnstein, E., Krumov, K., Genkova, P., Kanagawa, C., Hirshberg, M. S., et al. (2003). Individualism, collectivism, and authoritarianism in seven societies. *Journal of Cross-Cultural Psychology, 34*, 304-322.

Kline, J. A. (1971a). A Q-analysis of encoding behavior in the selection of evidence. *Speech Monographs, 37*, 190-197.

Kline, J. A. (1971b). Dogmatism of the speaker and selection of evidence. *Speech Monographs, 38*, 354-355.

Klopf, D. W., & McCroskey, J. C. (1969). *The elements of debate*. New York: Arco.

Knouse, S. B. (1983). The letter of recommendation: Specificity and favorability of information. *Personnel Psychology, 36*, 331-341.

Koballa, T. R. (1986). Persuading teachers to reexamine the innovative elementary science programs of yesterday: The effect of anecdotal versus data-summary communications. *Journal of Research in Science Teaching, 23*, 437-449.

Koelen, M., & Martijn, C. (1994). Persuasieve voorlichting [Persuasive public information]. In N. G. Röling, D. Kuiper, & R. Janmaat (Eds.), *Basisboek voorlichtingskunde* (pp. 217-239). Amsterdam / Meppel: Boom.

Koopman, P. L., Den Hartog, D. N., Konrad, E., Akerblom, S., Audia, G., Bakacsi, G., et al. (1999). National culture and leadership profiles in Europe: Some results from the GLOBE study. *European Journal of Work and Organizational Psychology, 8*, 503-520.

Krosnick, J. A., & Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology, 59*, 1140-1152.

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum.

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.

Kruglanski, A. W., & Thompson, E. P. (1999a). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry, 10*, 83-109.

Kruglanski, A. W., & Thompson, E. P. (1999b). The illusory second mode or, the cue is the message. *Psychological Inquiry, 10*, 182-193.

Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*, 636-647.

Kunda, Z., & Nisbett, R. E. (1986). Prediction and the partial understanding of the law of large numbers. *Journal of Experimental Social Psychology, 22*, 339-354.

Levasseur, D., & Dean, K. W. (1996). The use of evidence in presidential debates: A study of evidence levels and types from 1960 to 1988. *Argumentation and Advocacy, 32*, 129-142.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109.

Luchok, J. A., & McCroskey, J. C. (1978). The effect of quality of evidence on attitude change and source credibility. *The Southern Speech Communication Journal, 43*, 371-383.

MacIntyre, A. (1988). *Whose justice? Which rationality?* Notre Dame, IN: University of Notre Dame Press.

Maddux, J. E., & Rogers, R. W. (1980). Effects of source expertness, physical attractiveness, and supporting arguments on persuasion: A case of brains over beauty. *Journal of Personality and Social Psychology, 39*, 235-244.

McCroskey, J. C. (1969). A summary of experimental research on the effects of evidence in persuasive communication. *Quarterly Journal of Speech, 55*, 169-176.

McGuire, W. J. (2000). Standing on the shoulders of ancients: Consumer research, persuasion, and figurative language. *Journal of Consumer Research, 27*, 109-114.

McKerrow, R. E. (1990). Overcoming fatalism: Rhetoric/argument in postmodernity. In R. E. McKerrow (Ed.), *Argument and the postmodern challenge: Proceedings of the eight SCA/AFA conference on argumentation* (pp. 119-121). Annandale, VA: Speech Communication Association.

Meloen, J. (1991). De autoritarisme-concepten van Adorno et al. en Altemeyer vergeleken: Theoretische analyse en empirische test [A comparison of the concepts of authoritarianism of Adorno et al. and Altemeyer: A theoretical analysis and an empirical test]. In P. Scheepers, & R. Eisinga (Eds.), *Onderdanig en intolerant: Lacunes en controverses in autoritarisme-studies* (pp. 103-126). Nijmegen, The Netherlands: ITS.

Morgan, D. L. (1988). *Focus groups as qualitative research*. Newbury Park, CA: Sage.

Mulder, M. (1977). *The daily power game*. Leiden, The Netherlands: Martinus Nijhoff.

Munch, J. M., Boller, G. W., & Swasy, J. L. (1993). The effects of argument structure and affective tagging on product attitude formation. *Journal of Consumer Research, 20*, 294-302.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. London: Nicolas Brealey.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32*, 932-943.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristis in everyday inductive reasoning. *Psychological Review, 90*, 339-363.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291-310.

Nisbett, R. E., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science, 26*, 653-684.

Oestermeier, U., & Hesse, F. W. (2000). Verbal and visual causal arguments. *Cognition, 75*, 65-104.

Oestermeier, U., & Hesse, F. W. (2001). Singular and general causal arguments. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 720-725). Mahwah, NJ: Lawrence Erlbaum.

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. (1998). Justification explicitness and persuasive effect: A meta-analytic review of the effects of varying support articulation in persuasive messages. *Argumentation and Advocacy, 35*, 61-75.

O'Keefe, D. J. (2002). *Persuasion: Theory and research (2nd ed.)*. Thousand Oaks, CA: Sage.

O'Shaughnessy, J., & O'Shaughnessy, N. J. (2004). *Persuasion in advertising*. London: Routledge.

Pair, R. le, Crijns, R., & Hoeken, H. (2000). Het belang van cultuurverschillen voor het ontwerp van persuasieve teksten [The importance of cultural differences for the design of persuasive texts]. *Tijdschrift voor Taalbeheersing, 22*, 358-372.

Park, B., & Hastie, R. (1987). Perception of variability in category development: Instance- versus abstraction-based stereotypes. *Journal of Personality and Social Psychology, 53*, 621-635.

Parrott, R., Silk, K., Dorgan, K., Condit, C., & Harris, T. (2005). Risk comprehension and judgments of statistical evidentiary appeals: When a picture is not worth a thousand words. *Human Communication Research, 31*, 423-452.

Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods, 2*, 329-344.

Perloff, R. M. (2003). *The dynamics of persuasion: Communication and attitudes in the 21th century (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Pieters, R. G. M., Verplanken, B., & Modde, J. M. (1987). 'Neiging tot nadenken': Samenhang met beredeneerd gedrag ['Need for cognition': Its relation with reasoned action]. *Nederlands Tijdschrift voor de Psychologie, 42*, 62-70.

Planel, C. (1997). National cultural values and their role in learning: A comparative ethnographic study of state primary schooling in England and France. *Comparative Education, 33*, 349-373.

Poortinga, Y. H., & Vijver, F. J. R., van de (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-cultural Psychology, 18*, 259-282.

Pornpitakpan, C., & Francis, J. N. P. (2001). The effect of cultural differences, source expertise, and argument strength on persuasion: An experiment with Canadians and Thais. *Journal of International Consumer Marketing, 13*, 77-101.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research, 27*, 258-284.

Read, S. (1983). Once is enough: Causal reasoning from a single instance. *Journal of Personality and Social Psychology, 45*, 323-334.

Reinard, J. C. (1988). The empirical study of the persuasive effects of evidence: The status after fifty years of research. *Human Communication Research, 15*, 3-59.

Reinard, J. C. (1998). The persuasive effects of testimony assertion evidence. In M. Allen, & R. W. Preiss (Eds.), *Persuasion: Advances through meta-analysis* (pp. 69-86). Cresskill, NJ: Hampton.

Reinard, J. C. (2003, May). *Persuasive effects of alternative models of introducting evidence*. Paper presented at the 53rd annual conference of the ICA, San Diego, CA.

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.

Rieke, R. D., & Sillars, M. O. (1984). *Argumentation and the decision making process (2nd ed.)*. New York: Harper Collins.

Rietveld, T., Hout, R. van, & Ernestus, M. (2004). Pitfalls in corpus research. *Computers and the Humanities, 38*, 343-362.

Rohan, M. J., & Zanna, M. P. (1996). Value transmissions in families. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The psychology of values: The Ontario symposium: Vol. 8* (pp. 253-276). Hillsdale, NJ: Lawrence Erlbaum.

Rokeach, M. (1960). *The open and closed mind*. New York: Basic Books.

Rook, K. (1986). Encouraging preventive behavior for distant and proximal health threats: Effects of vivid versus abstract information. *Journal of Gerontology, 41*, 526-534.

Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology, 35*, 817-829.

Sanders, J. A., Gass, R. H., & Wiseman, R. L. (1991). The influence of type of warrant and receivers' ethnicity on perceptions of warrant strength. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Proceedings of the second international conference on argumentation* (pp. 709-718). Amsterdam: Sic Sat.

Schellens, P. J. (1985). *Redelijke argumenten: Een onderzoek naar normen voor kritische lezers [Reasonable arguments: A study in criteria for critical reading]*. Dordrecht, The Netherlands: 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 public discussion about biotechnology and food]. In L. van Waes, P. Cuvelier, G. Jacobs, & I. de Ridder (Eds.), *Studies in Taalbeheersing, volume 1* (pp. 416-427). Assen, The Netherlands: 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, The Netherlands: Martinus Nijhoff.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues, 50*, 19-37.

Sherer, M., & Rogers, R. W. (1984). The role of vivid information in fear appeals and attitude change. *Journal of Research in Personality, 18*, 321-334.

Sherman, S. J., Beike, D. R., & Ryalls, K. R. (1999). Dual-processing accounts of inconsistencies in responses to general versus specific cases. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 203-227). New York: Guilford Press.

Siegel, H. (1999). Argument quality and cultural difference. *Argumentation, 13*, 183-201.

Slater, M. D., & Rouner, D. (1996). Value-affirmative and value-protective processing of alcohol education messages that include statistical evidence or anecdotes. *Communication Research, 23*, 210-235.

Slusher, M. P., & Anderson, C. A. (1996). Using causal persuasive arguments to change beliefs and teach new information: The mediating role of explanation availability and evaluation bias in the acceptance of knowledge. *Journal of Educational Psychology, 88*, 110-122.

Smith, P. B. (2004). Nations, cultures, and individuals: New perspectives and old dilemmas. *Journal of Cross-cultural Psychology, 35*, 6-12.

Smith, P. B., & Schwartz, S. H. (1997). Values. In J. W. Berry, M. H. Segall, & C. Kagitçibasi (Eds.), *Handbook of cross-cultural psychology: Vol. 3.* (2nd ed., pp. 77-118). Boston: Allyn and Bacon.

Smith, E. R., & Zárate, M. (1992). Exemplar-based model of social judgment. *Psychological Review, 99*, 3-21.

Spooren, W., Smith, B., & Renkema, J. (2000). De invloed van stijl en type argumentatie op de overtuigingskracht van direct mail [The influence of style and argument type on the persuasiveness of direct mail]. *Tijdschrift voor Taalbeheersing, 22*, 344-355.

Stiff, J. B. (1986). Cognitive processing of persuasive message cues: A meta-analytic review of the effects of supporting information on attitudes. *Communication Monographs, 53*, 75-89.

Strange, J. J., & Leung, C. C. (1999). How anecdotal accounts in news and in fiction can influence judgments of a social problem's urgency, causes, and cures. *Personality and Social Psychology Bulletin, 25*, 436-449.

Swasy, J. L. (1979). Measuring the bases of social power. In W. L. Wilkie (Ed.), *Advances in consumer research: Vol. 6.* (pp. 340-346). Ann Arbor, MI: Association for Consumer Research.

Toulmin, S. E., Rieke, R., & Janik, A. (1984). *An introduction to reasoning (2nd ed.)*. New York: Macmillan.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Vijver, F. J. R. van de, & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Viswanathan, M. (1993). The measurement of individual differences in preference for numerical information. *Journal of Applied Psychology, 78*, 741-752.

Walton, D. N. (1997). *Appeal to expert opinion: Arguments from authority*. University Park, PA: Pennsylvania State University Press.

Warnick, B., & Inch, E. S. (1989). *Critical thinking and communication: The use of reason in argument*. New York: Macmillan.

Warnick, B., & Manusov, V. (2000). The organization of justificatory discourse in interaction: A comparison within and across cultures. *Argumentation, 14*, 381-404.

Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational behavior and human decision processes, 47*, 289-312.

Wells, G. L., & Harvey, J. H. (1977). Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology, 35*, 279-293.

Wiener, J. L., LaForge, R. W., & Goolsby, J. R. (1990). Personal communication in marketing: An examination of self-interest contingency relationships. *Journal of Marketing Research, 27*, 227-231.

Zukier, H., & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology, 47*, 349-360.

Acknowledgments

I would like to thank the following people for their practical help with the five studies:

Study 1

Geneviève Felten (Université de Paris Sorbonne-III), and Marjan Groeneveld

Study 2

Chantal Claudel (Université de Paris Sorbonne-VIII), and Sonia Gouirand (Université de Montpellier-II)

Study 3

Bas Andeweg (Technische Universiteit Delft), Céline Averseng (Université de Montpellier-II), Wim Blokzijl (Technische Universiteit Delft), Jean-François Bonnefon (Université de Toulouse-II), Jean-Pierre Brisacier (Université François-Rabelais de Tours), Jean-Marc Décaudin (Université de Toulouse-I), Dominique Desjeux (Université de Paris Sorbonne-V), Marinel Gerritsen (Radboud Universiteit Nijmegen), Christine Guislin (Université de Lille-III), Anke van Haastrecht (Rijksuniversiteit Groningen), Liesbeth Hermans (Radboud Universiteit Nijmegen), Constanze Hess (Universiteit van Amsterdam), Lettica Hustinx (Radboud Universiteit Nijmegen), Jérémie Ilharragorry (Université François-Rabelais de Tours), Tom van Laer (Radboud Universiteit Nijmegen), Peggy Lafon (Université François-Rabelais de Tours), Laurent Lepaludier (Université d'Angers), Fons Maes (Universiteit van Tilburg), Frank van Meurs (Radboud Universiteit Nijmegen), Lex van Meurs (Universiteit van Amsterdam), Joke Oppenhuisen (Universiteit van Amsterdam), Brigitte Planken (Radboud Universiteit Nijmegen), Jan Renkema (Universiteit van Tilburg), Philip Scheiner (Université de Paris Sorbonne-IV), Joost Schilperoord (Universiteit van Tilburg), Selma Verhoog (Radboud Universiteit Nijmegen), and Evelyne Vos-Fruit (Radboud Universiteit Nijmegen)

Study 4

Bas Andeweg (Technische Universiteit Delft), Thomas Beaufils (Université Marc-Bloch, Strasbourg), Joop Bindels (Université Marc-Bloch, Strasbourg), Wim Blokzijl (Technische Universiteit Delft), Jean-Pierre Brisacier (Université François-Rabelais de Tours), Chantal Claudel (Université de Paris Sorbonne-VIII), Christine Guislin (Université de Lille-III), Corrie de Haan (Technische Universiteit Delft), Marijke Hagedoorn (Universiteit van Amsterdam), Claudia Huisman (Université Marc-Bloch, Strasbourg), Carel Jansen (Radboud Universiteit Nijmegen), Joyce Karreman (Universiteit Twente), Fons Maes (Universiteit van Tilburg), Sabine Nibbeling (Universiteit van Amsterdam), Joke Oppenhuisen (Universiteit van Amsterdam), Jan Renkema (Universiteit van Tilburg), Philip Scheiner (Université de Paris Sorbonne-IV), Joost Schilperoord (Universiteit van Tilburg), Pascale Smorag (Université de Franche-Comté, Besançon), and Wendy Tollenaar (Universiteit Twente)

Study 5

Remieg Aerts (Radboud Universiteit Nijmegen), Thomas Beaufils (Université Marc-Bloch, Strasbourg), Michèle Bitour (Université Marc-Bloch, Strasbourg), Jacqueline de Bony (CNRS), Claudia Huisman (Université Marc-Bloch, Strasbourg), Margot van Mulken (Radboud Universiteit Nijmegen), and Evelyne Vos-Fruit (Radboud Universiteit Nijmegen)

Appendices

Appendix A

Table A1. Normative criteria for the types of evidence

type of evidence	normative criteria
statistical/anecdotal (general claim)	- the cases in the evidence should be representative of the class of cases in the claim - the number of cases in the evidence should be sufficiently large in order to allow valid generalizations about the class of cases in the claim
statistical (specific claim)	- the cases in the evidence should be representative of the class of cases to which the case that is presented in the claim belongs - the number of cases in the evidence should be sufficiently large in order to allow valid generalizations about the class of cases to which the case in the claim belongs - the cases in the evidence should be highly similar to the case presented in the claim
anecdotal (specific claim)	- there should be a large number of relevant similarities between the case in the evidence and the case in the claim - there should be no relevant dissimilarities between the case in the evidence and the case in the claim
causal	- the cause and the effect should covary - the cause should come into play before the effect occurs - there should not be other variables that trouble the causal relationship, nor in a positive way (the relationship only holds if another variable is present), nor in a negative way (the relationship does not hold if another variable is present)
expert	- the expert should be credible - the expert should be reliable - the expert should have a high expertise in the field that the claim he underscores is relevant for

Appendix B

Anecdotal and statistical evidence in cognitive psychological experiments

In Chapter 3, I referred to experiments in cognitive psychology whose results differed from studies presented in the review. In these psychological experiments, the normative criteria of a large sample size and high representativeness did not affect people's judgments, and anecdotal evidence had more impact than statistical evidence. This last difference could be attributed to the type of claim: whereas most of the evidence studies used general claims (see Table 3.1), the majority of the cognitive psychological studies concerned a specific claim. However, there appears to be a more compelling explanation for the difference in results. The objective of this appendix is to show that the persuasive impact of anecdotal evidence, the relatively small impact of statistical evidence, and the lack of effect of differences in sample size and representativeness result from the specific research designs. As reviews on evidence have selected these studies (e.g., Baesler & Burgoon, 1994; Reinard, 1988), and the underlying theoretical background of these studies (Tversky and Kahneman's heuristics) has been adopted as an explanation for the persuasiveness of anecdotal evidence, these studies deserve special attention here. I will first introduce the theoretical background, the design, and the results of cognitive psychological experiments on anecdotal and statistical evidence. Next, the presentation of a series of more recent cognitive psychological studies will demonstrate that statistical evidence has more impact under circumstances that differ from those in the original research designs, and that – as a consequence – the norms of sample size and representativeness do matter.

Theoretical background, design, and results of the cognitive psychological experiments

In experiments on the persuasiveness of evidence types, people are confronted with claims and evidence that they process in order to make a judgment. Cognitive psychologists are interested in this process of information gathering, making sense of it, and using it to reach a certain goal. When normative models on information processing were empirically tested, people's behavior was often characterized by errors and biases in comparison with the normative rules. People sometimes rely on strategies that are normatively weaker but easier to use. These strategies are knowledge structures, such as schemata and beliefs, and heuristics (e.g., Nisbett & Ross, 1980). Heuristics lead to a rapid, easy form of reasoning. In itself, as Nisbett and Ross stress, the use of heuristics is a valid way to arrive at a judgment, but it often results in errors and biases. One of the most important contributions to the study of heuristics in human judgments comes from Tversy and Kahneman's heuristics and biases program (see handbooks of Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). In a number of studies, Tversky and Kahneman have advanced heuristics that explain people's biases and errors in judgments, such as the representativeness heuristic, the availability heuristic, and anchoring and adjustment. The representativeness heuristic is relevant here.

A classic experiment on this heuristic is Kahneman and Tversky (1973). This study was presented in Section 3.1.1. Participants based their judgments on the attributes in the descriptions (anecdotal evidence), and ignored the explicitly stated statistics (statistical evidence). These findings were explained with the representativeness heuristic. People use this heuristic to make inferences about the probability that a person belongs to a group or that an event belongs to a sequence of events. Instead of integrating statistical information

with the individuating information, people ignore the statistical information, which results in a series of errors. The errors that result from the use of this heuristic can be related to three findings that I referred to in Chapter 3:

1. Anecdotal evidence has more impact than statistical evidence
   As was shown in the Kahneman and Tversky (1973) experiment, people neglect statistical information when they have to make a probability judgment, and base their decision on individuating information.

2. The norm of a large sample size is not important
   People do not mind generalizing from one single case. This insensitivity to sample size has been referred to as the belief in the law of small numbers (Tversky & Kahneman, 1971).

3. The norm of high representativeness is not important
   People generalize from single cases regardless of whether the sample is typical of the population or not (e.g., Hamill, Wilson, & Nisbett, 1980; Nisbett & Borgida, 1975; Strange & Leung, 1999).

The original work of Tversky and Kahneman (e.g., Tversky & Kahneman, 1971), replications in psychological studies (e.g., Nisbett & Borgida, 1975), and mass communication studies (e.g., Brosius & Bathelt, 1994) have consistently shown these errors. Below, I will discuss studies that prove that these three findings are the result of the research design in which anecdotal and statistical evidence were presented in competition.

Artefacts of the research design

In the studies on the representativeness heuristic, participants typically did not use statistical information to form a judgment. Some researchers claimed that their behavior is quite normal (Gilovich & Griffin, 2002, p. 12):

> "Consider Kahneman and Tversky's well-known engineer/lawyer problem. When asked whether a given description is likely to belong to an engineer or lawyer, one cannot fail to compute the similarity between the description and each professional stereotype. Both the immediacy and relevance of this assessment of similarity, then, make it highly likely that one's very definition of what the task is about will be hijacked. A question about likelihood is construed as a question about 'fit'. And there is nothing artifactual about the switch"

Bar-Hillel (1980, p. 217) explained that, when base-rate information and individuating information are presented in competition, "people may be ordering information by its perceived degree of relevance to the problem they are judging". High relevance information renders other information irrelevant. Anecdotal evidence is often more specific than statistical evidence in relation to the judgmental task. As participants especially relied on this anecdotal evidence, the base rates were not relevant anymore. As a result, anecdotal evidence had more impact than statistical evidence (1), and the norms of large sample size (2) and high representativeness (3) were not important. Numerous studies have commented on the biased comparison of individuating and base-rate information in Tversky and Kahneman's experiments (e.g., Ginossar & Trope, 1980; Zukier & Pepitone, 1984). These studies claim that base-rate neglect is an artefact of the research design, and show that statistical evidence is persuasive if it is more relevant than in the Tversky and Kahneman

studies. Indirectly, these studies also suggest that the norms of sample size and representativeness are important. Base-rate information is more relevant, for instance, when it is more vidid, and when it is presented after anecdotal evidence. This will be shown below.

Vividness and salience – When base-rates are made more vivid and salient, they are more frequently used by participants to form a judgment (e.g., Ginossar & Trope, 1980; Zukier & Pepitone, 1984). In Ginossar and Trope, for instance, the individual case was made less useful for the task, less consistent or less related to the outcome categories than in the original Kahneman and Tversky (1973) study. Under these conditions, base-rate information had judgmental impact. In three experiments, Zukier and Pepitone showed that the context of the decision in the experiment, and the role of the participants both affect the importance of base-rate information on judgments. Participants who were instructed to solve the judgmental problems as scientists, for instance, assigned greater weight to base rates than participants who were instructed to solve the problems as clinical psychologists.

Order effect – The order effect explanation holds that, as anecdotal evidence was always the second piece of information in the Tversky and Kahneman studies, it was more relevant, and therefore more influential than statistical evidence. This claim was advanced and supported in a series of seven experiments by Krosnick, Li and Lehman (1990). When statistical evidence was given after anecdotal evidence, it had greater influence on the participants' assessment than when it was given before anecdotal evidence. This order or recency effect can be explained with reference to conversational conventions. People expect that, when a text writer decides to give extra information, this information must be highly relevant, and even more relevant than the first piece of evidence (see, for a discussion about conversational conventions and the Tversky and Kahneman studies, Hilton & Slugoski, 2001). The information given in the second place therefore has more impact.

Conclusion

The objective of this appendix was to argue that findings of cognitive psychological experiments about sample size, representativeness, and the persuasive impact of anecdotal and statistical evidence have been mediated by their research design. When statistical information is made more relevant, it has a greater influence on people's judgments than Tversky and Kahneman originally suggested. As a consequence, it is likely that the norms of large sample size and high representativeness are more important than the original studies suggested.

Appendix C

Table C1. Selected Dutch brochures (Study 1)

brochure title (sender)	pages	2nd coder*
interest for other people		
Amsterdam heeft wat met Managua (Stichting Stedenband Amsterdam-Managua)	6	h
De ervaring van je leven (Holland World Youth)	4	s
Geef een kind als Joli een betere toekomst! (Foster Parents Plan)	8	s
Geef om dit kind (Stichting Liliane Fonds)	8	h
Heerlijk smullen! (Solidaridad)	4	h
Het kindje links is de moeder (Terre des Hommes)	6	h
Ik zoek een huis waar ik voor donker binnen moet zijn (Federatie Pleegzorg)	4	h
In Nederland bent u vrij om lid te worden van Amnesty International (Amnesty International Nederland)	8	h
Max Havelaar meer dan lekker (Stichting Max Havelaar)	4	h
Nederlanders naar Srebrenica (Werkgroep Nederland-Srebrenica)	6	h
Teken van geweld (Vrouwen Informatie Punt Nicaragua)	4	h
Wel eens een vluchteling de weg gewezen? (Stichting Voor Vluchtelingen & Nieuwkomers)	2	s
Zuivere koffie (-)	4	h
societal interest		
'Sport'vissen is niet zo sportief ... (Dierenbescherming Afdeling Den Haag)	6	h
receiver interest		
De o is weer in de maand: tijd voor de griepprik! (College voor zorgverzekeringen)	8	h
Geef kanker minder kans eet volop groente en fruit (Nederlandse Kankerbestrijding)	8	h
Hou de spanning erin (-)	4	s
Mist, halveer je snelheid, verdubbel je afstand (Ministerie van Verkeer en Waterstaat)	4	h
Pas bij nul houden we op met tellen (Provinciaal Orgaan Verkeersveiligheid Brabant)	8	h
Rookoverlast? U kunt er wat aan doen (Stichting Volksgezondheid en Roken)	12	s
Tabak (Jellinek Preventie en Consultancy)	12	s
Wat doet drank met u? (Nationaal Instituut voor Gezondheidsbevordering en Ziektepreventie)	12	h

* h = Hans Hoeken, s = Marianne Starren

Table C2. Selected French brochures (Study 1)

brochure title (sender)	pages	2nd coder*
interest for other people		
Amnesty International (Amnesty International)	6	s
Avec vous le défi de la solidarité (Terre des Hommes France)	6	s
Ensemble, brisons l'indifférence (Action des Chrétiens pour l'Abolition de la Torture)	6	g
Grâce à vous, Frères des Hommes les aide à construire leur vie (Frères des Hommes)	10	s
Grandir ensemble (Mouvement pour les Villages d'Enfants)	10	g
La faim c'est quotidien (Action contre la Faim)	10	g
Legs et donations en faveur du CCFD (Comité Catholique contre la Faim et pour le Développement)	6	g
Legs et donations: il y a autre chose à vivre que la misère (Fondation Abbé Pierre pour le Logement des Défavorisés)	12	g
Mouvement pour les villages d'enfants (Mouvement pour les villages d'enfants)	6	s
Œuvrer ensemble (Orphelinat Mutualiste de la Police Nationale)	6	s
On ne peut par dire qu'on ne peut rien faire (Action contre la Faim)	8	s
Parrainez un enfant dans le monde (Centre français de protection de l'enfance)	4	g
Une vraie vie de famille en héritage (SOS Villages d'Enfants)	6	s
societal interest		
Baleines et dauphins en liberté avec SOS Grand Bleu (SOS Grand Bleu)	8	g
receiver interest		
Écoute cancer (Ligue Nationale contre le Cancer)	6	s
Et vous, avec l'alcool, vous en êtes où? (Comité Français d'Éducation pour la Santé)	8	s
Femmes et tabac (Comité Français d'Éducation pour la Santé)	6	s
Le dopage: quelques informations pour votre sécurité (Préfecture de Police)	6	g
Le tabagisme passif (Comité Français d'Éducation pour la Santé)	16	g
Protection rapprochée (Comité Français d'Éducation pour la Santé)	6	g
Quand je serai grande, j'inventerai un vaccin contre la vitesse (Sécurité Routière)	8	s
Si chacun fait un peu, c'est la vie qui gagne (Sécurité Routière)	6	g

* g = Marjan Groeneveld, s = Marianne Starren

Appendix D

Table D1. The origin of the 50 claims, and their presence in Studies 2, 3, 4, and 5

claim	origin*	2	3	4	5
1. A course that is oriented to the labour market leads to better opportunities on that market for art history students.	A	x	x		
2. Riding a bike over bad, bumpy roads promotes stronger bones.					
3. Listening to classical music helps students to absorb a lot of knowledge in a short period of time.	A, C	x	x	x	
4. A walk during lunch leads to higher work productivity in the afternoon.					
5. Twins that grow up apart from each other more often take an identical course in life than twins that grow up in the same family.	B				
6. The introduction of a provisional driver's licence is said to prevent reckless driving by drivers that have just passed their driving test.	B				
7. Watching soaps reduces depression.	C				
8. Drinking at least two liters of water a day reduces cellulite in women.	A, C				
9. If children help with the cooking, they'll have a greater appetite.					
10. A regular change of workplace within a company raises productivity at work.			x	x	x
11. A national campaign in favor of breast-feeding is said to reduce the risk of smoking to children.	C				
12. The regular consumption of pizzas reduces the risk of cancer.					
13. Library membership gives elderly people the feeling that they are less lonely.					
14. Playing computer games has a positive impact on people's sense of direction.			x		
15. The introduction of a new Politics course is said to increase the interest of high school students in the field of politics.	A, C				
16. Being employed as an intermediary by an employment agency helps you to find a job for later.			x		
17. Solving puzzles is a way for elderly people to prevent memory problems.					
18. Multicultural personnel promotes more innovative solutions in companies.					
19. A course of fruit juice helps people to attain their desired weight and maintain that weight.	C		x		
20. The use of Internet reduces the feeling of social isolation that elderly people may have.	C				
21. Playing slow music in supermarkets raises their turnover.	B		x		
22. Boys' performance at school can be improved by putting boys next to girls in class.		x	x	x	
23. Too tight a tie leads to reduced sight.			x	x	x
24. Playing party games helps young criminals to become more socialized.		x	x	x	x
25. The consumption of cold beverages during exams leads to lower achievements in students.	B		x		

* A = Hoeken & Hustinx, 2003a, B = Hoeken & Hustinx, 2003b, C = Hornikx & Hoeken, 2003

claim	origin*	2	3	4	5
26. The introduction of an education protocol against bullying prevents children from falling victim to such behavior.	C		x		
27. The use of relaxation rooms in the office reduces absenteeism.	A				
28. Compulsory work placement abroad leads to better performance in an international context.	C				
29. Places of entertainment that are equipped with cameras give visitors a greater feeling of security.	A				
30. Watching cartoons that were popular in their youth helps publicity agents to regain their creativity.		x	x		
31. Eating paprika leads to better sporting performance.					
32. The use of a lightweight mouse and regular relaxation reduce the risk of RSI for people who often work with computers for hours a day.	C				
33. Salesmen feel more involved in their company, when they use new media, such as a laptop with an Internet connection at home.		x	x		
34. People who live in a city that attracts a lot of visitors have a strongly developed identity.					
35. Art in offices is good for the atmosphere at work.					
36. Descaling drinking water leads to more fractures in elderly people.	B				
37. The consumption of basil in tomato pasta sauce improves sporting performance.			x	x	x
38. Writing poems helps people to become more considerate towards others.					
39. Compulsory driving lessons for people over 70 years old reduces their uncertainty in traffic.	A		x		
40. Reading regional newspapers increases the fear of becoming a crime victim.					
41. Waiters that repeat the orders of customers verbatim, receive a higher tip.			x	x	x
42. Fear of driving cars can be taken away by riding along with a truck driver for a day.		x	x	x	x
43. Babies that are read to by their parents develop fewer language problems at school.	B				
44. Keeping pets leads to better mental health in the elderly.	B				
45. The increase in mobile phone use by car drivers leads to more road accidents.	B				
46. Keeping photos of family and friends on the desk raises productivity at work.		x	x	x	x
47. Fear of flying decreases as a consequence of a balloon flight.			x	x	x
48. Exercising for one hour twice a week is a solution to depression.	A				
49. Swimming once a week helps to prevent RSI.					
50. Confronting young criminals with the victims of their behavior is an effective way to prevent them from reoffending.	A				

* A = Hoeken & Hustinx, 2003a, B = Hoeken & Hustinx, 2003b, C = Hornikx & Hoeken, 2003

Table D2. Fifty Dutch and French claims

1. Een cursus arbeidsmarktoriëntatie leidt tot betere kansen op de arbeidsmarkt voor studenten kunstgeschiedenis. / Les étudiants en histoire de l'art ont de meilleures chances sur le marché du travail après avoir suivi un cours d'orientation sur le marché du travail.
2. Door veel te fietsen over hobbelige, slechte wegen krijg je sterkere botten. / Faire du vélo sur de mauvaises routes cahoteuses renforce les os.
3. Het luisteren naar klassieke muziek helpt scholieren om in korte tijd veel kennis op te nemen. / En écoutant la musique classique lorsqu'ils étudient, les lycéens assimilent beaucoup de connaissances en peu de temps.
4. Een wandeling tijdens de lunch zorgt voor een hoge arbeidsproductiviteit in de middag. / Une promenade pendant la pause de midi stimule la productivité du travail dans l'après-midi.
5. Tweelingen die apart van elkaar opgroeien hebben vaker een identieke levensloop dan tweelingen die in hetzelfde gezin opgroeien. / Les jumeaux qui grandissent séparément ont plus souvent une vie identique que ceux qui évoluent dans la même famille.
6. De invoering van een voorlopig rijbewijs kan roekeloos rijgedrag van automobilisten die net hun rijbewijs hebben tegengaan. / L'introduction d'un permis de conduire provisoire peut prévenir l'imprudence au volant des automobilistes qui viennent d'obtenir leur permis.
7. Het kijken naar soaps leidt tot minder depressiviteit. / Regarder les feuilletons réduit les dépressions nerveuses.
8. Het drinken van minstens twee liter water per dag kan cellulitis bij vrouwen doen verminderen. / La consommation d'au moins deux litres d'eau par jour permet de réduire la cellulite des femmes.
9. Als kinderen zelf helpen bij het koken, krijgen ze meer eetlust. / La participation des enfants à la cuisine stimule leur appétit.
10. Het regelmatig veranderen van werkplek binnen hetzelfde bedrijf leidt tot een verhoging van de arbeidsproductiviteit. / Le changement régulier du lieu de travail au sein de la même entreprise entraîne une augmentation de la productivité du travail.
11. Een landelijke campagne vóór borstvoeding verkleint de kans op roken bij kinderen die geen borstvoeding hebben gehad. / Une campagne nationale en faveur de l'allaitement maternel réduit le risque pour les enfants non allaités par leur mère de développer un penchant au tabagisme.
12. Het regelmatig eten van pizza's verkleint de kans op het krijgen van kanker. / La consommation régulière de pizzas réduit le risque du cancer.
13. Een lidmaatschap bij de bibliotheek geeft bejaarden het gevoel dat ze minder eenzaam zijn. / Une souscription à la bibliothèque aide les personnes âgées à se sentir moins seules.
14. Het spelen van computerspellen heeft een positieve invloed op het oriëntatievermogen. / Les jeux électroniques exercent une influence positive sur le sens de l'orientation.
15. Het instellen van een verplicht vak Politiek op school kan de interesse van middelbare scholieren in politiek verhogen. / L'introduction au programme d'un cours obligatoire de sciences po stimule l'intérêt des lycéens à la vie politique.
16. Werken als intercedent helpt om later zelf een baan te vinden. / En travaillant comme intérimaire on trouvera plus facilement un emploi dans l'avenir.
17. Puzzelen is voor ouderen een manier om vergeetachtigheid tegen te gaan. / Les mots croisés aident les personnes âgées à prévenir les défauts de mémoire.

18. Een multicultureel personeelsbestand op het werk zorgt voor meer innovatieve oplossingen. / En employant des salariés d'origine ethnique variée on favorise la créativité de la prise de decisions.
19. Door het volgen van een sapkuur komen en blijven veel mensen op het gewenste gewicht. / Les régimes à base de jus de fruits conduisent durablement au poids désiré.
20. Het gebruik van het internet kan het gevoel van sociale isolatie van ouderen verminderen. / L'utilisation d'Internet est un des moyens pour réduire l'isolement social dont souffrent les personnes âgées.
21. Het draaien van langzame muziek in supermarkten verhoogt de omzet. / La musique à cadence tempérée permet d'augmenter le chiffre d'affaires des supermarchés.
22. Schoolprestaties van jongens kunnen verbeterd worden door jongens naast meisjes in de schoolbanken te zetten. / Les performances scolaires des garçons s'améliorent quand ils sont placés à côté des filles en classe.
23. Een te strakke stropdas leidt tot een verminderd gezichtsvermogen. / Une cravate trop serrée mène à une vue réduite.
24. Het spelen van gezelschapsspelletjes leidt ertoe dat jonge criminelen socialer worden. / En jouant à des jeux de société, les délinquants juvéniles deviennent plus sociables.
25. Koude dranken drinken tijdens een tentamen leidt tot slechtere prestaties bij studenten. / La consommation de boissons froides pendant les épreuves réduit les performances des étudiants.
26. Het invoeren van een onderwijsprotocol tegen pesten kan voorkomen dat kinderen de dupe worden van pesten. / L'introduction d'un protocole de l'établissement contre les tracasseries est un moyen pour éviter les tracasseries subies par les élèves.
27. Het gebruik van een ontspanningsruimte op het werk brengt het ziekteverzuim omlaag. / L'utilisation d'un espace de détente au travail réduit le taux d'absentéisme des salariés.
28. Een verplichte buitenlandse stage leidt ertoe dat mensen beter functioneren in een internationale bedrijfscontext. / Dans l'environnement économique international, un stage obligatoire à l'étranger mène à un meilleur fonctionnement.
29. Uitgaansgelegenheden die voorzien zijn van camera's geven de bezoekers een veiliger gevoel. / La vidéosurveillance dans les bars, théâtres ou restaurants renforce le sentiment de sécurité des clients.
30. Het kijken naar tekenfilms uit hun eigen jeugd helpt reclamemakers om weer creatief te zijn. / En regardant les dessins animés de leur jeunesse, les publicitaires retrouveront plus facilement leur créativité.
31. Het eten van paprika zorgt voor het neerzetten van betere sportprestaties. / La consommation de poivrons stimule les performances sportives.
32. Het gebruik van een lichtgewicht muis en regelmatig ontspannen vermindert de kans op RSI bij mensen die vaak urenlang achter de computer zitten. / L'utilisation d'une souris ultralégère et le respect de moments de détente réguliers réduisent le risque de LATR pour ceux qui passent souvent de longues heures devant l'écran.
33. Vertegenwoordigers voelen zich meer betrokken bij hun bedrijf door het gebruik van nieuwe media, zoals een laptop met internetverbinding thuis. / Grâce aux nouveaux médias, comme l'ordinateur portable avec une connexion Internet à la maison, les commerciaux se sentent plus impliqués dans leur entreprise.
34. Mensen die in een stad wonen die door veel toeristen wordt bezocht, hebben een sterk ontwikkeld identiteitsgevoel. / Les habitants d'une ville courue par les touristes développent une forte identité.
35. Kunst in bedrijfsruimtes is goed voor de sfeer op het werk. / Les oeuvres d'art exposées dans les entreprises stimulent l'ambiance au site de travail.

36. Ontkalking van het drinkwater leidt tot meer botbreuken bij bejaarden. / Le détartrage de l'eau augmente le nombre de fractures parmi les personnes âgées.
37. Het eten van basilicum in de pastasaus verhoogt de sportprestaties. / Le basilic ajouté aux sauces pour accompagner les pâtes améliore les performances sportives.
38. Het schrijven van gedichten zorgt ervoor dat mensen attenter in de omgang worden. / Ecrire des poèmes mène à un comportement plus attentif envers l'autre.
39. Verplichte rijlessen voor mensen boven de 70 jaar kunnen hun onzekerheid in het verkeer wegnemen. / En imposant des leçons de conduite aux personnes de plus de 70 ans, on peut supprimer leur incertitude au volant.
40. Het lezen van regionale dagbladen verhoogt de vrees om slachtoffer te worden van een misdrijf. / La lecture de journaux régionaux fait augmenter la peur d'être victime d'une agression.
41. Obers die bestellingen van hun klanten letterlijk herhalen, krijgen een hogere fooi. / Les garçons de café qui répètent littéralement les commandes de leurs clients, touchent des pourboires plus élevés.
42. Angst om auto te rijden kan worden opgelost door een dag mee te rijden met een vrachtwagenchauffeur. / Un des moyens pour supprimer la peur au volant, c'est de passer une journée aux côtés d'un routier qui conduit son camion.
43. Baby's die worden voorgelezen door hun ouders hebben later minder taalproblemen op school. / Si les parents font régulièrement la lecture à leurs bébés, ils auront moins de problèmes linguistiques plus tard à l'école.
44. Het houden van huisdieren leidt tot een betere geestelijke gezondheid van bejaarden. / Prendre un animal domestique améliore la santé mentale des personnes âgées.
45. Het toenemend gebruik van mobiele telefoons achter het stuur leidt tot meer verkeersongelukken. / L'utilisation croissante du téléphone portable au volant fait augmenter le nombre d'accidents de la route.
46. Foto's van familie en vrienden op het bureau verhogen de productiviteit op het werk. / En posant sur le bureau des photos de famille et d'amis, on augmente la productivité du travail.
47. Als gevolg van een vlucht met een luchtballon neemt de angst om te vliegen af. / La phobie de l'avion diminue après un vol en montgolfière.
48. Twee maal per week een uurtje sporten kan een oplossing zijn voor depressiviteit. / Faire du sport deux fois par semaine est un moyen pour résoudre les crises de dépressions nerveuses.
49. Een keer per week zwemmen helpt om RSI te voorkomen. / Faire de la natation une fois par semaine est un moyen pour prévenir les LATR.
50. Een confrontatie met de slachtoffers van hun gedrag is een effectief middel om herhaling bij jonge criminelen te voorkomen. / La confrontation des jeunes avec leurs victimes est un moyen efficace dans la lutte contre la récidive en matière de délinquance juvénile.

Table D3. Probability ratings of 20 claims by Dutch (n = 30) and French (n = 28) participants ordered from relatively improbable (a low score) to relatively probable (a high score) (Dutch and French ratings are not significantly different; the last claim does not precisely meet the first criterion mentioned in Section 5.1.1: the mean probability rating is above 3.50)

claim number*	Dutch	French	mean
42	2.10	2.25	2.18
23	2.23	2.18	2.21
25	2.37	2.21	2.29
3	2.53	2.18	2.36
19	2.43	2.36	2.40
41	2.87	2.50	2.69
37	2.70	2.71	2.71
47	2.83	2.78	2.81
46	2.77	2.96	2.87
26	3.07	2.79	2.93
1	3.10	2.89	3.00
22	3.13	3.00	3.07
24	3.20	3.21	3.21
14	3.30	3.21	3.26
16	3.53	3.04	3.29
21	3.07	3.61	3.34
30	3.50	3.32	3.41
33	3.30	3.57	3.44
10	3.23	3.64	3.44
39	3.50	3.57	3.54
mean	2.92	2.91	2.92

* see Table D1 for the claims

Appendix E

Table E1. Dutch and French templates of the types of evidence (Studies 2, 3, 4, and 5)

evidence	Dutch	French
anecdotal	Sinds [voornaam + achternaam] uit [plaats] de oorzaak heeft meegemaakt, is het gevolg opgetreden.	Depuis que la cause s'est manifestée chez [prénom + nom] de [place], la conséquence s'est produite.
statistical	Uit een Nederlands onderzoek onder N personen is gebleken dat de consequentie bij X % van hen is opgetreden als gevolg van de oorzaak.	Une étude française, se basant sur un échantillon de N personnes, a montré que la conséquence s'est produite à la suite de la cause pour X % d'entre eux.
	De resultaten van een Nederlands onderzoek onder N personen lieten zien dat de consequentie bij X % van hen is opgetreden als gevolg van de oorzaak.	Les résultats d'une étude française, se rapportant à N personnes, ont montré que la conséquence s'est produite à la suite de la cause pour X % d'entre eux.
expert	Volgens prof. dr. [naam] van de universiteit in [plaats], specialist / deskundige op het gebied van X, is [het standpunt[.	D'après [nom], professeur à l'université de [place] et directeur de recherche au CNRS au laboratoire de X, [la thèse].
	Prof. dr. [naam], specialist / deskundige op het gebied van X aan de universiteit van [plaats], is [het standpunt[.	[nom], professeur à l'université de [place] et directeur de recherche au CNRS, est spécialiste en X. Il souligne que [la thèse].

Appendix F

Table F1. Equivalence of specific information used in Dutch and French material (Study 2) (Figures in brackets correspond to the position in the databases given in Section 5.1.3)

information	Netherlands	France
anecdotal names	Bakker	Moreau (10)
	Van de Ven	Fournier (21)
	Heuvelmans	Lambert (24)
	Mulder	Robin (47)
	Glastra	Morin (48)
	Elling	Duval (54)
	Claessens	Picard (88)
	Hesse	Roger (90)
expert names	Timmermans	Bonnet (32)
	Westerveld	Chevalier (41)
	Meijer	Clément (44)
	Wildschut	Masson (57)
	De Groot	Marchand (65)
	De Gauw	Brunet (79)
	Van Zanten	Giraud (89)
	Van Voorschoten	Aubert (98)
male first names	Jeroen (1)	Sébastien (1)
	Martijn (2)	Nicolas (2)
	Peter (3)	Julien (3)
	Patrick (4)	David (4)
	Dennis (5)	Cédric (5)
	Mark (6)	Christophe (6)
	Sander (7)	Frédéric (7)
	Marcel (9)	Stéphane (9)
extra names claim (22)	Wouter (10)	Olivier (10)
	Chantal (6)	Sandrine (6)
cities	Den Haag (3)	Lyon (3)
	Eindhoven (5)	Nice (5)
	Tilburg (6)	Nantes (6)
	Breda (9)	Le Mans (19)
	Apeldoorn (11)	Le Havre (11)
	Enschede (12)	Reims (12)
	's-Hertogenbosch (16)	Toulon (15)
	Haarlemmermeer (18)	Brest (18)
	Dordrecht (20)	Dijon (20)
universities	Utrecht (1)	Strasbourg (2)
	Groningen (3)	Lille (3)
	Rotterdam (4)	Lyon (4)
	VU Amsterdam (5)	Bordeaux (5)
	Leiden (6)	Grenoble (6)
	Maastricht (9)	Toulouse (8)
	Twente (12)	Rennes (12)
	Wageningen (13)	Angers (13)

Appendix G

Table G1. Equivalence of specific information used in Dutch and French material (Study 3) (Figures in brackets correspond to the position in the databases given in Section 5.1.3)

information	Netherlands	France
anecdotal names	Van de Ven	Fournier (21)
	Heuvelmans	Lambert (24)
	Glastra	Morin (48)
	Elling	Duval (54)
expert names	Timmermans	Bonnet (32)
	Westerveld	Chevalier (41)
	De Groot	Marchand (65)
	Van Zanten	Giraud (89)
male first names	Peter (3)	Julien (3)
	Patrick (4)	David (4)
	Mark (6)	Christophe (6)
	Marcel (9)	Stéphane (9)
extra names claim (22)	Chantal (6)	Sandrine (6)
	Wouter (10)	Olivier (10)
cities	Apeldoorn (11)	Le Havre (11)
	Enschede (12)	Reims (12)
	Haarlemmermeer (18)	Brest (18)
	Dordrecht (20)	Dijon (20)
universities	Utrecht (1)	Strasbourg (2)
	Rotterdam (4)	Lyon (4)
	VU Amsterdam (5)	Bordeaux (5)
	Leiden (6)	Grenoble (6)

Table G2. Number of participants in each version in function of country, city, and group (Study 3)

country	city	total	version				
			1	2	3	4	5
Netherlands	Amsterdam	13	2	3	3	3	2
		29	6	6	5	5	7
	Delft	23	3	5	5	6	4
	Groningen	46	11	10	8	7	10
	Nijmegen	39	8	0	5	8	18
		51	13	11	12	10	5
		18	0	14	2	2	0
	Tilburg	16	3	3	4	3	3
		70	15	9	17	17	12
	total	305	61	61	61	61	61
France	Angers	30	6	0	6	11	7
	Paris	75	27	21	9	1	17
		32	0	9	13	8	2
	Roubaix	34	5	0	11	16	2
	Toulouse	26	5	7	2	7	5
	Tours	34	5	7	6	8	8
		64	11	15	12	8	18
	total	295	59	59	59	59	59

Appendix H

Table H1. Equivalence of specific information used in Dutch and French material (Study 4) (Figures in brackets correspond to the position in the databases given in Section 5.1.3)

information	Netherlands	France
anecdotal names	Bakker	Moreau (10)
	Van de Ven	Fournier (21)
	Heuvelmans	Lambert (24)
	Glastra	Morin (48)
	Elling	Duval (54)
expert names	Timmermans	Bonnet (32)
	Westerveld	Chevalier (41)
	De Groot	Marchand (65)
	Van Zanten	Giraud (89)
male first names	Jeroen (1)	Sébastien (1)
	Peter (3)	Julien (3)
	Patrick (4)	David (4)
	Mark (6)	Christophe (6)
	Marcel (9)	Stéphane (9)
cities	Eindhoven (5)	Nice (5)
	Breda (9)	Le Mans (19)
	Apeldoorn (11)	Le Havre (11)
	Den Bosch (16)	Toulon (15)
	Haarlemmermeer (18)	Brest (18)
universities	Groningen (3)	Lille (3)
	Rotterdam (4)	Lyon (4)
	VU Amsterdam (5)	Bordeaux (5)
	Leiden (6)	Grenoble (6)

Table H2. Number of participants in each version in function of country, city, and group (Study 4)

country	city	total	version				
			1	2	3	4	5
Netherlands	Amsterdam	26	3	9	4	5	5
		18	4	4	3	4	3
		12	0	0	0	3	9
		17	3	3	5	3	3
	Delft	14	4	4	2	2	2
		7	3	2	0	1	1
	Enschede	7	2	2	1	2	7
		14	4	5	0	5	0
		7	0	0	7	0	0
	Tilburg	37	7	8	10	5	7
		52	10	11	12	10	9
		12	2	1	1	2	6
	Nijmegen	77	18	11	15	18	15
	total	300	60	60	60	60	60
France	Besançon	49	11	13	7	10	8
	Paris	33	9	6	6	7	5
		23	6	3	6	4	4
	Roubaix	58	10	13	13	10	12
	Strasbourg	11	2	3	1	2	3
		5	1	1	1	1	1
		11	3	2	3	2	1
		12	4	1	3	3	1
		16	0	4	2	4	6
		10	1	2	2	2	3
	Tours	72	13	12	16	15	16
	total	300	60	60	60	60	60

Appendix I

Selected items of the RWA scale (Studies 4 and 5)

RWA items close to the concept of power distance were selected. As Altemeyer (1981, 1988) does not indicate to which of the three concepts each item refers (see Section, 6.4.4), I selected 14 potential items of Altemeyer (1988, pp. 22-23): items 22, 12, 3, 5, 17, 23, 16, 26, 2, 29, 7, 1, 9, and 30. At the same time, nine pairs of Dutch students (Radboud University Nijmegen) were asked to rank Altemeyer's (1988) RWA items from most appropriate to measure power distance to least appropriate to measure power distance. Subsequently, the 10 lists were compared, and finally 10 items were selected. These items were available in Dutch (Meloen, 1991), but were adapted into contemporary Dutch. The adapted items were translated into French by the translation agency France Intermédiaire (Nijmegen), of which a second person back translated the French items into Dutch. The 10 items were (numbers corresponding to Altemeyer, 1988, pp. 22-23):

2. It is wonderful that young people today have greater freedom to protest against things they don't like and 'do their own thing'.

3. It is always better to trust the judgment of the proper authorities in government and religion than to listen to the noisy rabble-rousers in our society who are trying to create doubt in people's minds.

5. It would be best for everyone if the proper authorities censored magazines and movies to keep thrashy material away from the youth.

7. The sooner we get rid of the traditional family structure, where the father is the head of the family and the children are taught to obey authority automatically, the better. The old-fashioned way has a lot wrong with it.

12. Obedience and respect are the most important virtues children should learn.

17. In these troubled times laws have to be enforced without mercy, especially when dealing with agitators and revolutionaries who are stirring things up.

22. If a child starts becoming unconventional and disrespectful of authority, it is his parents' duty to get him back to the normal way.

23. In the final analysis the established authorities, like parents and our national leaders, generally turn out to be right about things, and all the protesters don't know what they're talking about.

26. The real keys to the 'good life' are obedience, discipline, and sticking to the straight and narrow.

29. Students in high school and university must be encouraged to challenge their parents' ways, confront established authorities, and in general criticize the customs and traditions of our society.

Appendix J

Table J1. Equivalence of specific information used in Dutch and French material (Study 5) (Figures in brackets correspond to the position in the databases given in Section 5.1.3)

information	Netherlands	France
expert names	Timmermans	Bonnet (32)
	Westerveld	Chevalier (41)
	Meijer	Clément (44)
	De Groot	Marchand (65)
	De Gauw	Brunet (79)
	Van Zanten	Giraud (89)
	Van Voorschoten	Aubert (98)
universities	Utrecht (1)	Paris (1)
	Groningen (3)	Lille (3)
	Rotterdam (4)	Lyon (4)
	VU Amsterdam (5)	Bordeaux (5)
	Leiden (6)	Grenoble (6)
	Maastricht (9)	Toulouse (8)
	Twente (12)	Rennes (12)
	Wageningen (13)	Angers (13)

Summary in Dutch

Overtuigende teksten hebben vaak als doel mensen over te halen iets te doen of iets te laten. Als er in die teksten argumenten worden aangedragen en mensen deze argumenten kritisch beoordelen, dan is de kwaliteit van de aangedragen argumenten van belang: sterke argumenten zijn overtuigender dan zwakke argumenten. Helaas is er nog weinig bekend in welk opzicht sterke argumenten zich onderscheiden van zwakke. In dit proefschrift wordt argumentkwaliteit bestudeerd door de overtuigingskracht van evidentie te onderzoeken. Evidentie wordt als ondersteuning voor standpunten gebruikt in argumentatie. Er zijn verschillende evidentietypen: statistische, anekdotische, causale en expertevidentie. Door de relatieve overtuigingskracht van deze typen te bestuderen, wordt het inzicht in argumentkwaliteit vergroot. Hoewel verschillende onderzoekers hebben onderstreept dat de overtuigingskracht van evidentietypen zou kunnen afhangen van cultuur, is hiernaar nog nooit onderzoek verricht. Recent psychologisch onderzoek heeft laten zien dat mensen die variëren in culturele achtergrond kunnen verschillen in de manieren waarop ze redeneren. Omdat evidentie een rol kan spelen in deze redeneerprocessen, is het mogelijk dat ook de relatieve overtuigingskracht van de vier evidentietypen door cultuur wordt beïnvloed. De algemene onderzoeksvraag luidt daarom:

> Onderzoeksvraag 1 Zijn er culturele verschillen in de relatieve overtuigingskracht van statistische, anekdotische, causale en expertevidentie?

Meer specifiek zou de overtuigingskracht van expertevidentie gerelateerd kunnen zijn aan machtsafstand, de mate waarin er in een cultuur geaccepteerd wordt dat macht (bijvoorbeeld geld, invloed en kennis) ongelijk verdeeld is in de samenleving. In culturen met een grote machtsafstand zoals de Franse wordt een ongelijke machtsverdeling verwacht en geaccepteerd. In dergelijke culturen zou expertevidentie overtuigender kunnen zijn dan in culturen met een kleine machtsafstand zoals de Nederlandse. Immers, mensen accepteren dan dat er

anderen zijn die meer weten over bepaalde onderwerpen dan zijzelf. De specifieke onderzoeksvraag luidt daarom:

> Onderzoeksvraag 2 Is de overtuigingskracht van expertevidentie groter in Frankrijk dan in Nederland?

Om deze vragen te beantwoorden zijn vijf studies uitgevoerd. In de eerste twee studies is de *verwachte* overtuigingskracht onderzocht: hoe overtuigend verwachten mensen dat de vier evidentietypen zijn voor anderen in hun cultuur? In de derde en vierde studie is de *daadwerkelijke* overtuigingskracht onderzocht. Deze studies doen uitspraken over hoe overtuigend de vier evidentietypen in werkelijkheid zijn. De vijfde en laatste studie is opgezet om een verklaring voor een resultaat in de vierde studie te toetsen.

Normatieve criteria en overtuigingskracht

Als er in overtuigende teksten gebruik wordt gemaakt van argumenten, dan gebeurt dit meestal door in te gaan op de voordelen van het aanbevolen gedrag (een snelle internetverbinding door een ADSL-aansluiting), of op de nadelen van het afgeraden gedrag (misselijkheid door dagelijkse sherryconsumptie). Deze vorm van argumentatie wordt pragmatische argumentatie of argumentatie op basis van voor- en nadelen genoemd. De redenering in pragmatische argumentatie is als volgt: als de gevolgen van gedrag (on)wenselijk zijn, dan is het gedrag het ook. Een voorbeeld: ‘misselijkheid is onwenselijk, dus het dagelijkse drinken van sherry is onwenselijk’. Pragmatische argumentatie wordt vaak gebruikt, maar leidt het er ook toe dat de lezer wordt overtuigd? Alleen als mensen gemotiveerd en capabel zijn om de argumenten in een tekst kritisch te beoordelen, is de kwaliteit van de argumenten belangrijk. Sterke argumenten zijn dan overtuigender dan zwakke argumenten. In de eerste twee hoofdstukken laat ik zien dat een sterk pragmatisch argument in overtuigende teksten een argument is dat het (a) heel waarschijnlijk maakt dat een gevolg gaat optreden en dat het (b) heel aannemelijk maakt dat dit gevolg wenselijk of juist onwenselijk is. Een lezer die argumenten kritisch beoordeelt, zal niet zomaar aannemen dat het oorzakelijke verband zal optreden. Evidentie kan dan worden ingezet om het waarschijnlijker te maken dat bijvoorbeeld het dagelijks drinken van sherry tot misselijkheid leidt.

Er zijn grofweg vier evidentietypen: (1) statistische evidentie met een getalsmatige samenvatting van een aantal gevallen, (2) anekdotische evidentie waarbij een specifiek geval wordt gepresenteerd, (3) causale evidentie met een

verklaring en (4) expertevidentie, ten slotte, waarbij een deskundige wordt aangehaald. Hieronder staan deze evidentietypen voor het standpunt over misselijkheid als gevolg van sherry:

(1) Uit een steekproef van 267 mensen bleek 78% van hen last van misselijkheid te hebben na het dagelijks drinken van sherry

(2) Sinds Johan Kaarsemaker uit Assen dagelijks sherry drinkt, heeft hij regelmatig last van misselijkheid

(3) Alcohol komt in het bloed terecht en tast zo de maag en het evenwichtsorgaan in de hersenen aan

(4) Volgens alcoholdeskundige prof. dr. F. Klassen van de Universiteit Utrecht leidt het dagelijks drinken van sherry tot misselijkheid

In de argumentatietheorie zijn normen geformuleerd voor sterke en zwakke argumentatie. Voor mensen die argumenten kritisch beoordelen zouden die normen van belang moeten zijn: normatief sterke argumenten zijn dan overtuigender dan normatief zwakke argumenten. In hoofdstuk 2 worden deze normatieve criteria toegepast op de vier evidentietypen. Expertevidentie, bijvoorbeeld, is normatief sterk als de expert die wordt aangehaald geloofwaardig is, betrouwbaar is en een hoge mate van expertise heeft op het gebied waarover het standpunt gaat dat hij onderstreept. Als mensen een tekst kritisch verwerken, dan zou normatief sterke evidentie overtuigender moeten zijn dan normatief zwakke evidentie. In hoofdstuk 3 wordt een overzicht gegeven van empirische studies waarin onderzocht is of de normen voor overtuigende evidentie er in de praktijk ook echt toe doen. Hoewel het aantal onderzoeken niet groot is, suggereert dit overzicht dat dit inderdaad het geval is. In het laatste gedeelte van hoofdstuk 3 volgt een tweede overzicht, namelijk van de relatieve daadwerkelijke overtuigingskracht van verschillende evidentietypen. De voorzichtige conclusie van dit overzicht is dat statistische, causale en expertevidentie overtuigender zijn dan anekdotische evidentie.

Studies 1 en 2: verwachte overtuigingskracht

In de enkele onderzoeken naar de verwachte overtuigingskracht van evidentietypen is de verwachte overtuigingskracht op twee manieren gemeten: *indirect* aan de hand van het gebruik van evidentie door taalgebruikers en *direct* aan de hand van experimenten. Bij de indirecte onderzoeken is de aanname dat het gebruik van evidentietypen door schrijvers of proefpersonen een indicator is voor de mate waarin ze verwachten dat de evidentietypen overtuigend zijn. De twee manieren van meten zijn beide in dit proefschrift toegepast: de corpusanalytische benadering in Studie 1 en de experimentele benadering in Studie 2.

Studie 1 is opgezet met het doel te onderzoeken of er (a) culturele verschillen zijn in het relatieve gebruik van de vier evidentietypen in overtuigende communicatie in Nederland en Frankrijk en of (b) expertevidentie daarbij vaker wordt gebruikt in Frankrijk. Onafhankelijke beoordelaars analyseerden het vóórkomen van evidentie als ondersteuning voor standpunten in pragmatische argumentatie in 22 Nederlandse en 22 Franse overtuigende voorlichtingsbrochures. De resultaten lieten in de eerste plaats zien dat er inderdaad culturele verschillen waren in het gebruik van de vier evidentietypen in het Nederlandse en het Franse corpus. In de tweede plaats bleek expertevidentie relatief vaker te worden gebruikt door de Franse tekstschrijvers dan door hun Nederlandse collega's. Expertevidentie werd nauwelijks gebruikt in Nederlandse overtuigende folders en in beperkte mate in de Franse folders. Omdat tekstschrijvers professionals zijn van wie verwacht mag worden dat ze ideeën hebben over wat wel en wat niet overtuigend werkt, is het aannemelijk om te veronderstellen dat de culturele verschillen in het gebruik van evidentietypen een indicatie zijn voor de verwachte overtuigingskracht van deze evidentietypen. Een beperking van dit corpusonderzoek is dat het twijfelachtig is of de tekstschrijvers bij elke beslissing om evidentie te gebruiken de beschikking hadden over alle vier de evidentietypen. Om deze tekortkoming weg te nemen is er in Studie 2 voor gezorgd dat de proefpersonen vrij konden kiezen uit de vier evidentietypen.

Studie 2 is een experiment uit de serie van vier experimenten waarvoor hetzelfde basismateriaal is ontwikkeld: een aantal standpunten met bijbehorende evidentietypen. In een vooronderzoek zijn twintig standpunten geselecteerd die even waarschijnlijk waren gevonden door Franse en Nederlandse studenten. Een voorbeeld van een standpunt is 'Het eten van basilicum in de pastasaus verhoogt de sportprestaties'. Voor elk standpunt zijn vier evidentietypen ontwikkeld. De evidentie voldeed aan de normatieve criteria uit hoofdstuk 2. Voor anekdotische evidentie was dit onmogelijk, omdat dit evidentietype normatief zwak is. Immers,

om te generaliseren naar het standpunt zou het gebaseerd moeten zijn op een groot aantal gevallen, maar het gaat juist uit van één geval. Voor causale evidentie bleken de normen ook moeilijk om te operationaliseren. Door het vertalen en terugvertalen van de Nederlandse en Franse standpunten en evidentie was het materiaal in beide talen in hoge mate vergelijkbaar.

De vragenlijsten waren in elk experiment schriftelijk en werden in hoor- of werkcolleges door studenten ingevuld. In die vragenlijsten is ook een schaal opgenomen om eventuele culturele verschillen te kunnen verklaren. Mocht expertevidentie overtuigender blijken in Frankrijk dan in Nederland, dan zou dat aan machtsafstand kunnen liggen, maar ook aan heel andere factoren. Daarom kan er gecontroleerd worden of de Franse proefpersonen inderdaad hoger scoren op waarden die zijn gerelateerd aan machtsafstand dan de Nederlandse proefpersonen. Omdat het meten van waarden op proefpersoonniveau problemen met zich meebrengt, zo wordt betoogd in hoofdstuk 5, is de Voorkeur voor Expertinformatie (VVE) schaal ontwikkeld. Van deze schaal wordt verwacht dat hij samenhangt met de overtuigingskracht van expertevidentie. Daarnaast wordt van de Franse proefpersonen verwacht dat ze een grotere voorkeur voor expertinformatie hebben dan de Nederlandse proefpersonen.

In Studie 2 kregen Franse en Nederlandse studenten acht standpunten met telkens de vier evidentietypen. Voor elke standpunt moest een rangschikking gemaakt worden van het evidentietype waarvan ze verwachtten dat het het *meest* overtuigend is tot het type waarvan ze verwachtten dat het het *minst* overtuigend is. Proefpersonen vulden ook de VVE-schaal in. De resultaten van dit experiment lieten zien dat er culturele verschillen waren in relatieve verwachte overtuigingskracht van de vier evidentietypen. Nederlandse proefpersonen verwachtten dat statistische evidentie overtuigender was dan de Franse proefpersonen en het omgekeerde gold voor anekdotische evidentie. De gemiddelde rangschikking voor expertevidentie was hetzelfde in beide culturen, maar expertevidentie stond wel vaker bovenaan de rangschikking bij de Franse proefpersonen dan bij de Nederlandse proefpersonen. Dit culturele verschil kon niet worden toegeschreven aan de VVE-schaal, omdat de scores op deze schaal in beide culturen niet verschilden. Studies 1 en 2 tonen aan dat culturen kunnen verschillen in de verwachtingen die ze hebben van de overtuigingskracht van evidentietypen. Of de evidentietypen ook daadwerkelijk verschillen in overtuigingskracht in Nederland en Frankrijk, is onderzocht in de drie andere onderzoeken.

Studies 3, 4 en 5: daadwerkelijke overtuigingskracht

In hoofdstuk 6 worden drie onderzoeken gepresenteerd die zijn uitgevoerd om na te gaan of er (a) culturele verschillen zijn in de relatieve daadwerkelijke overtuigingskracht van de vier evidentietypen en of (b) expertevidentie overtuigender is in Frankrijk dan in Nederland. Overtuigingskracht is in deze studies als volgt gemeten. De proefpersonen kregen een serie standpunten voorgelegd met telkens een ander evidentietype of geen evidentie. De verdeling van de evidentietypen over de standpunten was verschillend in de versies van de vragenlijst. Zodoende is een standpunt in elke versie door een ander evidentietype ondersteund. De overtuigingskracht van evidentie is het verschil tussen de waarschijnlijkheid van een standpunt mét evidentie en dat van een standpunt zónder evidentie.

In Studie 3 is de overtuigingskracht van de vier evidentietypen onderzocht door Nederlandse en Franse proefpersonen de waarschijnlijkheid van twintig standpunten met evidentie te laten beoordelen. De evidentie was, indien mogelijk, normatief sterk. De studenten vulden ook de VVE-schaal in. De resultaten van Studie 3 lieten zien dat er culturele verschillen waren in de relatieve overtuigingskracht van de vier evidentietypen. Expertevidentie was in *absolute* zin niet overtuigender voor de Franse proefpersonen dan voor de Nederlandse, maar wel in *relatieve* zin. Terwijl expertevidentie voor de Nederlandse proefpersonen samen met causale evidentie een middenpositie innam, was expertevidentie voor de Franse proefpersonen overtuigender dan causale en anekdotische evidentie. De VVE-schaal kon de relatief hogere overtuigingskracht van expertevidentie niet verklaren, aangezien er geen verschil was in de score op deze schaal voor de Franse en Nederlandse proefpersonen. De score op de VVE-schaal bleek daarnaast positief samen te hangen met de overtuigingskracht van expertevidentie voor de Nederlandse proefpersonen, maar niet voor de Franse.

In Studie 3 was normatief sterke expertevidentie in absolute zin niet overtuigender in Frankrijk dan in Nederland. Op basis van wat bekend is over machtsafstand kan verwacht worden dat er ook culturele verschillen zijn als expertevidentie normatief zwak is. In culturen met een grote machtsafstand zouden geloofwaardige en betrouwbare experts met een expertisegebied dat irrelevant is voor het standpunt nog steeds overtuigend kunnen zijn. De norm voor een relevant expertisegebied zou er niet toe doen, omdat de expert, bijvoorbeeld een hoogleraar, een hoge status heeft. In culturen met een kleine machtsafstand zou normatief zwakke expertevidentie wel minder overtuigend zijn

dan normatief sterke expertevidentie. Studie 4 is opgezet om deze verwachting te onderzoeken.

Dit experiment bevatte dezelfde standpunten als in Studie 3, alleen nu met normatief sterke of zwakke statistische en expertevidentie. De zwakke varianten hadden een irrelevant expertisegebied (expertevidentie) of een kleinere steekproef en lagere percentages (statistische evidentie). Naast de VVE-schaal werd ook de Right-Wing Authoritarianism (RWA) opgenomen. Van deze schaal – die onderdanigheid, agressie en conventionaliteit meet – is bekend dat hij samenhangt met de culturele dimensie machtsafstand. Er bleken echter geen verschillen te zijn in de scores op de VVE- en RWA-schaal voor de Franse en Nederlandse proefpersonen. Om te controleren of de ingeschatte expertise van normatief sterke experts hoger was dan van normatief zwakke experts moesten proefpersonen voor elke expert de ingeschatte expertise met betrekking tot het standpunt aangeven. Er was ook een controlevraag voor de manipulatie van normatief sterke en zwakke statistische evidentie. Beide manipulaties bleken succesvol, maar in veel hogere mate voor de Nederlandse dan voor de Franse proefpersonen.

Uit de resultaten bleek dat er een cultureel verschil was in de overtuigingskracht van normatief sterke en normatief zwakke evidentie. Nederlandse proefpersonen vonden sterke evidentie overtuigender dan zwakke, terwijl evidentiekwaliteit er voor de Franse proefpersonen niet toe deed. Voor hen was normatief zwakke expertevidentie dus even overtuigend als normatief sterke expertevidentie. Dit culturele verschil kan worden toegeschreven aan het kleine verschil dat de Franse proefpersonen zagen in de expertise van normatief sterke en normatief zwakke experts. Op basis van machtsafstand kan gesuggereerd worden dat het expertisegebied onbelangrijk is voor Fransen, als de expert een hoge status heeft, zoals de professoren in het experimentele materiaal. Zo beredeneerd zou het belang van het expertisegebied moeten toenemen wanneer de experts minder hoog in aanzien staan. Studie 5 onderzocht of dit het geval is.

In Studie 5 gaven Nederlandse en Franse proefpersonen voor vier soorten experts de ingeschatte expertise met betrekking tot het standpunt aan: hoogleraren met een relevant of een irrelevant expertisegebied (zoals in Studie 4) en onderzoekers met een relevant of een irrelevant expertisegebied. De resultaten in Studie 4 werden in dit experiment gerepliceerd voor de hoogleraren en de onderzoekers. Ook bij experts met een lagere status (onderzoekers) had het expertisegebied geen effect op de ingeschatte expertise voor de Franse proefpersonen. Het resultaat in Studie 4 kan dus niet worden verklaard door de hoge status van de experts.

Conclusie en discussie

De studies waarover in dit proefschrift wordt gerapporteerd, laten zien dat er culturele verschillen zijn in de relatieve verwachte en daadwerkelijke overtuigingskracht van statistische, anekdotische, causale en expertevidentie. Daarnaast blijkt expertevidentie overtuigender te zijn in Frankrijk dan in Nederland, maar zijn de verschillen genuanceerd. De verwachte overtuigingskracht van expertevidentie is slechts in geringe mate hoger in Frankrijk dan in Nederland (Studies 1 en 2). Normatief sterke (Studie 3) en normatief zwakke (Studie 4) expertevidentie is in absolute zin niet overtuigender in Frankrijk dan in Nederland, maar wel in relatieve zin.

Bij het optreden van deze culturele verschillen met betrekking tot expertevidentie is het belangrijk om vast te stellen in hoeverre de verschillen kunnen worden toegeschreven aan machtsafstand. Een probleem dat werd gesignaleerd in alle experimenten, is dat de schalen op individueel proefpersoonniveau (VVE, RWA) de culturele verschillen niet konden verklaren. De gevonden resultaten kunnen dan ook niet empirisch worden toegeschreven aan verschillen in machtsafstand. In hoofdstuk 7 passeren enkele andere verklaringen de revue, waaronder onzekerheidsvermijding, het onderwijssysteem en de mate van kritisch verwerken. Veel van deze verklaringen zijn gerelateerd aan machtsafstand. Concluderend lijkt machtsafstand aan de ene kant een aannemelijke verklaring, maar aan de andere kant kan de invloed ervan in deze studies niet worden aangetoond.

In hoofdstuk 7 wordt een aantal beperkingen van de studies besproken die betrekking hebben op het materiaal (standpunten met evidentie zonder de context van een overtuigende tekst) en de proefpersonen (alleen studenten). Deze beperkingen en de problemen die zijn opgetreden bij het meten van het effect van cultuur leiden tot een serie suggesties voor verder onderzoek, zoals het gebruik maken van focusgroepen, meer uitgebreide teksten met evidentie, andere proefpersonen en andere culturen. Dit soort onderzoek kan meer inzicht verschaffen in het optreden en in het verklaren van de invloed van cultuur op de relatieve overtuigingskracht van evidentietypen.

Curriculum vitae

Jos Hornikx was born in Eindhoven (the Netherlands) on February 17, 1979. After graduating from Sondervick College in Veldhoven in 1997, he studied Business Communication Studies at the Radboud University Nijmegen, and – for one semester – Applied Foreign Languages at the University François-Rabelais in Tours (France). In 2001, he took his Master's degree in Business Communication Studies, with a focus on International Business Communication (cum laude). Subsequently, he was appointed on a 'White raven' grant at the same department in October 2001 as a junior researcher on a PhD project on the effect of culture on the persuasiveness of evidence types. During the project, Hornikx taught persuasive effects research at the Radboud University Nijmegen, and marketing at the University of Montpellier (France). In October 2005, he was awarded a Niels Stensen research grant to work with Prof. Daniel J. O'Keefe at Northwestern University (USA) on a project titled 'A meta-analysis of the persuasive effect of adapting persuasive messages to cultural and personality characteristics of the audience'. He will conduct this research at Nortwestern University from February to November 2006. Hornikx' research interests are the effect of culture on the persuasion process, human reasoning, and the appreciation of foreign language use in advertising.