

**Children's and adults' interpretation of communicated probabilities: Studies
on directionality and preciseness.**

by

Amélie Nadine Aline GOURDON

A thesis submitted to
the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Psychology
University of Birmingham
March 2013

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Twelve experiments investigated how children and adults interpret verbal probabilities (e.g., *it is likely*). The experiments were designed to determine if and when children and adults use the directionality or the likelihood of verbal probabilities. In Experiments 1a and 1b, I showed that children use only the directionality of verbal probabilities to make decisions. However, they dismiss it when speakers are malevolent. In Experiment 2a, adults showed that they do not consider only the directionality or the likelihood when making decisions. Rather response times suggested that adults are sensitive to the potential conflict between the two features. In Experiment 2b, I showed that, given an unlimited time to decide, adults can show less preference for the positive directionality. However in Experiment 3a, I found that in conversational context, adults prefer the positive directionality even when given more time to decide. In Experiment 3b, adults used the directionality in different ways according to speakers' intentions. In contrast with children in Experiment 1b, they preferred the negative directionality when the speaker was malevolent, rather than dismissed the directionality overall. In Experiments 4a to 4e, counter to expectations, I did not find that speakers using more precise format to communicate probabilities are judged more responsible based on their predictions' accuracy. Instead the results suggest that listeners reward predictions that suggest that speakers wish for the best outcome for listeners. Finally in Experiment 5, I found that the preference for receiving more precise probabilistic information is contingent on speakers' expertise. These results together support a pragmatic account of verbal probabilities. The

directionality of verbal probabilities is a pragmatic cue that influences decision making by shaping listeners' assumptions.

When I was studying towards my M.Sc., Hedya, then 8 years old, told me she would never do a PhD. I am afraid that after watching me getting through the process of writing-up, this is truer than ever. But I hope this experience has taught her that, yes, as a woman, she can be independent and have it all.

To Hedya, first and foremost.

À mes parents, Alain et Nadine Gourdon, et à Justine et Lilia. À mes grands-mères.

To Lazarus Kanhukamwe.

Acknowledgements

This work would not have been possible without the excellent and thorough feedback of my supervisor, Sarah Beck. Her endless patience is also to highlight: she *corrected often* the order of my words, and I hope she will forgive me for using this space to bask in a *French-ed*, un-concise English writing style. I am also particularly grateful that Sarah tried to hide her worries when other duties were getting in the way of writing-up, and instead trusted me to prove that I could face the challenge. Black coffee provision throughout the years was a great bonus too.

Back to the start, in France, I asked Gaëlle Villejoubert for a research placement because I was intrigued that she was pronouncing SPSS the English way. That is, not ES-PEY-ESSE-ESSE. The fact that she used to do so because she had completed a PhD in the UK was not the least to intrigue me. So Gaëlle offered me a placement, and I met verbal probabilities. While she was supervising my M.Sc. project, she then threw at me a simple 'You could look for a co-supervision in England'. Nine months later, I had moved to the UK for a full supervision. As if anyone needed more proof that concise is influential.

The placement with Gaëlle was also the opportunity to get to know better Marie Juanchich, who would become my academic sister, best friend and best collaborator. Endlessly supporting, infinitely inspiring by the quality of her work and her personality. Even when we were in different countries, Marie has provided the best attention. She is also a great resource to share specific French expats' annoyances. I am looking forward to more collaboration, more leisure time ending talking about research and above all to the new team member to come next spring.

I am grateful as well to University of Birmingham and the School of Psychology for having funded this work. The School of Psychology also provided a great work environment. The Cognitive Development and the Social Cognition groups are only examples of the rich research environment I have been offered, but they are to thank more particularly for their feedback throughout the years. The international dimension of the School was not the least feature, providing plenty of foreign people to share social interests with. The list of people to thank

within the School will prove dangerously on the edge of lengthy writing, but need to be mentioned (in pseudo-random order): Nina Powell, Dicle Dövençioğlu, Jess Wang, Laura Kimberley, Katerina Kantartzis, Kate Graham, Andrew Surtees, Nicola Cutting.

I am not sure I can make Hedya Gourdon laugh-until-it-hurts enough times to thank her for playing the parent this last month. I cannot say enough how proud I am of how you adapted after moving abroad in the space of 6 weeks, on pretext that your peculiar mother wanted to pursue a PhD.

Mudiwa, my love, thank you for finding your way to me more than once, just to be the perfect man who won't complain that spending the weekend with me essentially means losing the use of his computer and barely exchanging a word while I am writing. Or in the fun case, being lectured about research and waiting for me to take more architectural pictures than you would. I am looking forward to have again the time to dance, many hours, and maybe now I can be the one who is attentive and low in demands.

Probably none of this would have happened if my parents hadn't failed, most unknowingly and unwillingly, to teach me that you can't always get what you want. They can be reassured that I am not trying to teach it to Hedya.

Finally, I would not be functional without dancers to provide me with amazing shows to satisfy my need for excitement, as well as emotional scapegoating and quality time out with Hedya. Cross-genre, unexpected, weird musicians should be acknowledged for providing mood enhancers and a protective bubble to my expat' self when needed, with a special mention to Gogol Bordello, Yann Tiersen and Mano Negra. Finally, double caramel ice-cream provided sugar rushes.

Dissemination

The contents of this thesis have been presented at conferences of the European Association for Decision Making (2009 & 2011), the Experimental Psychology Society (2010-11), and the European Society for Philosophy and Psychology (2010).

Chapter 3, largely in its current form, has been resubmitted to *Acta Psychologica* for review. Chapter 6 is under revision for resubmission. Chapters 2 and 4 are in preparation for submission.

Table of Contents

CHAPTER 1: INTRODUCTION	1
1.1. Probabilities and probabilistic reasoning.....	2
1.1.1. Probability, likelihood, chance, odds, risk.....	2
1.1.2. Normative and descriptive theories of probabilities	4
1.1.3. Irrationality: the dual model of reasoning and decision making	5
1.1.4. Probabilities as communicated: verbal vs. numerical probabilities	8
1.2. Judging the likelihood of verbal probabilities.....	11
1.2.1. The problem of equivalence between verbal and numerical probabilities.....	11
1.2.2. A numerical solution to the vagueness of verbal probabilities: the membership function	14
1.3. Verbal probabilities are language: pragmatics and argumentative function.....	15
1.3.1. Verbal probabilities and locus of uncertainty	15
1.3.2. Verbal probabilities and directionality.....	16
1.4. Interpretation of verbal probabilities in childhood	20
1.5. Overview and outline of the current experimental work.....	23
CHAPTER 2: CHILDREN'S UNDERSTANDING OF VERBAL PROBABILITIES	26
2.1. Introduction	26

2.2. Experiment 1a	32
2.2.1. Method.....	32
2.2.1.1. Participants.....	32
2.2.1.2. Materials.....	32
2.2.1.3. Procedure	34
2.2.2. Results	34
2.2.2.1. Likelihood judgement.....	34
2.2.2.2. Expected value judgement	36
2.2.2.3. Decision-making	38
2.2.3. Discussion.....	38
2.3. Experiment 1b.....	42
2.3.1. Method.....	42
2.3.1.1. Participants.....	42
2.3.1.2. Materials.....	42
2.3.1.3. Procedure	44
2.3.2. Results	44
2.3.2.1. Likelihood judgement.....	44
2.3.2.2. Expected value judgement	46

2.3.2.3. Decision-making	48
2.3.3. Discussion	50
2.4. General discussion	53
CHAPTER 3: OVERCOMING THE FRAMING EFFECT WHEN MAKING DECISIONS BASED ON	
VERBAL PROBABILITIES: HAVING MORE TIME IS HELPFUL BUT NOT ENOUGH.	58
3.1. Introduction	58
3.2. Experiment 2a	64
3.2.1. Method	64
3.2.1.1. Participants	64
3.2.1.2. Materials	64
3.2.1.3. Procedure	66
3.2.2. Results	68
3.2.2.1. Accuracy	68
3.2.2.2. Directionality	70
3.2.2.3. Response time	72
3.2.3. Discussion	75
3.3. Experiment 2b	77
3.3.1. Method	77

3.3.1.1. Participants	77
3.3.1.2. Materials	77
3.3.1.3. Procedure	77
3.3.2. Results	78
3.3.2.1. Accuracy	78
3.3.2.2. Directionality	81
3.3.2.3. Response time	84
3.3.3. Discussion	88
3.4. General discussion	89
CHAPTER 4: MAKING DECISIONS BASED ON VERBAL PROBABILITIES IN A CONVERSATIONAL CONTEXT: SPEAKERS' INTENTIONS MATTER	95
4.1. Introduction	95
4.2. Experiment 3a	102
4.2.1. Method	102
4.2.1.1. Participants	102
4.2.1.2. Materials	102
4.2.1.3. Procedure	103
4.2.2. Results	104

4.2.2.1. Accuracy.....	104
4.2.2.2. Directionality	107
4.2.2.3. Response time	109
4.2.3. Discussion	114
4.3. Experiment 3b.....	119
4.3.1. Method.....	119
4.3.1.1. Participants.....	119
4.3.1.2. Materials.....	119
4.3.1.3. Procedure	119
4.3.2. Results	122
4.3.2.1. Accuracy.....	122
4.3.2.2. Directionality	126
4.3.2.3. Response time	129
4.3.2.4. Individual analyses.....	132
4.3.3. Discussion	133
4.4. General discussion	136

CHAPTER 5: SHOULD THE MESSENGER BE SHOT? RESPONSIBILITY JUDGEMENTS OF SPEAKERS USING VERBAL AND NUMERICAL PROBABILITIES.	140
--	-----

5.1. Introduction	140
5.2. Experiment 4a	145
5.2.1. Method.....	145
5.2.1.1. Participants.....	145
5.2.1.2. Materials.....	145
5.2.1.3. Procedure	146
5.2.2. Results	147
5.2.3. Discussion	150
5.3. Experiment 4b.....	153
5.3.1. Method.....	153
5.3.1.1. Participants.....	153
5.3.1.2. Materials and procedure	153
5.3.2. Results	154
5.3.3. Discussion	155
5.4. Experiment 4c	156
5.4.1. Method.....	156
5.4.1.1. Participants.....	156
5.4.1.2. Materials and procedure	157

5.4.2. Results	157
5.4.2.1. Responsibility judgements	157
5.4.2.2. Likelihood of recommending.....	159
5.4.3. Discussion.....	160
5.5. Experiment 4d.....	162
5.5.1. Method.....	162
5.5.1.1. Participants.....	162
5.5.1.2. Materials and procedure	163
5.5.2. Results	163
5.5.2.1. Responsibility judgements	163
5.5.2.2. Likelihood of recommending.....	164
5.5.3. Discussion.....	166
5.6. Experiment 4e	168
5.6.1. Method.....	168
5.6.1.1. Participants.....	168
5.6.1.2. Materials and procedure	169
5.6.2. Results	169
5.6.2.1. Responsibility judgements	169

5.6.2.2. Likelihood of recommending.....	172	
5.6.3. Discussion.....	174	
5.7. General discussion.....	176	
CHAPTER 6: RECEIVING UNCERTAIN INFORMATION: PRECISION IS ONLY AS GOOD AS THE		
SPEAKER'S EXPERTISE.....		183
6.1. Introduction.....	183	
6.2. Method.....	186	
6.2.1. Participants.....	186	
6.2.2. Materials.....	186	
6.2.3. Procedure.....	187	
6.3. Results.....	188	
6.4. Discussion.....	188	
CHAPTER 7: GENERAL DISCUSSION.....		192
7.1. Summary.....	192	
7.2. Links to literature.....	198	
7.3. Future research.....	204	
7.4. Conclusion.....	208	
Appendices.....	209	

Appendix I: Verbal probabilities used in the pilot study of Experiment 1a	209
Appendix II: Scenarios used in Experiment 1a.....	210
Example of training scenario.....	210
Experimental scenarios	210
Appendix III: Example page of the scenario booklet used in Experiment 1a	213
Appendix IV: Screenshot of an exercise proposed to students preparing GCSEs (BBC, 2012)	214
Appendix V: Materials used in Experiment 4a.....	215
Scenarios	215
Positive uncertain events	215
Negative uncertain events.....	215
Probabilistic expressions used in Experiment 4a	216
Outcomes used in Experiment 4a	216
Appendix VI: Scenarios used from Experiment 4b to Experiment 4e	217
Scenarios	217
Probabilistic expressions used from Experiment 4b to Experiment 4e	217
Outcomes used from Experiment 4b to Experiment 4e	218
Appendix VII: Material in Experiment 5 (in experimental order)	219

References	221
------------------	-----

List of Figures

Figure 1: Hypothetical membership functions for two probability terms (from Wallsten & al., 1986).....	15
Figure 2: Mean Likelihood Judgement as a Function of Age, Directionality and Likelihood (Experiment 1a)	35
Figure 3: Mean Expected Value Judgement as a Function of Age, Directionality and Likelihood (Experiment 1a).....	36
Figure 4: Proportion of Risky Choices as a Function of Age, Directionality and Likelihood (Experiment 1a)	37
Figure 5: Mean Likelihood Judgement as a Function of Age, Directionality, Likelihood and Speaker's Intention (Experiment 1b).....	45
Figure 6: Mean Expected Value Judgement as a Function of Age, Directionality, Likelihood and Speaker's Intention (Experiment 1b).....	47
Figure 7: Proportion of Risky Choices as a Function of Age, Directionality, Likelihood and Speaker's Intention (Experiment 1b).....	49
Figure 8: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Time Condition (Experiment 2b).....	69
Figure 9: Mean Response Time in the Comparison Conditions (Experiment 2a)	73
Figure 10: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Time Condition (Experiment 2b).....	79

Figure 11: Mean Proportion of Choices in Favour of the Positive Directionality as Function of the Comparison Condition and the Time Condition (Experiment 2b)	82
Figure 12: Mean Response Time (in msec.) as Function of the Comparison Condition and the Time Condition (Experiment 2b)	85
Figure 13: Example of Stimuli Used in Experiment 3a	103
Figure 14: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Time Condition (Experiment 3a)	105
Figure 15: Mean Proportion of Choices in Favour of the Positive Verbal Probability (Standard Deviation) as Function of the Comparison Condition and the Time Condition (Experiment 3a)	108
Figure 16: Mean Response Time (in msec.) as Function of the Comparison Condition and the Time Condition (Experiment 3a)	110
Figure 17: Example of Stimuli Used in Experiment 3b (Malevolent Speaker)	121
Figure 18: Example of Stimuli Used in Experiment 3b (Benevolent Speaker)	122
Figure 19: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Speaker (Experiment 3b).....	123
Figure 20: Mean Proportion of Choices in Favour of the Positive Verbal Probability (Standard Deviation) as Function of the Comparison Condition and the Speaker Condition (Experiment 3b)	127
Figure 21: Mean Response Time (in msec.) as Function of the Comparison Condition and the Speaker Condition (Experiment 3b)	130

Figure 22: Mean Responsibility Judgement as Function of the Communication Mode, the Likelihood the Event was Predicted with, the Valence of the Event, and the Outcome (Experiment 4a) 147

List of Tables

Table 1: Verbal Probabilities Used in Experiments 1a and 1b.....	33
Table 2: Verbal Probabilities Used in Experiments 2a, 2b, 3a and 3b.....	67
Table 3: Post-Hoc Special Contrasts on the Response Time (Experiment 2a)	74
Table 4: Post-Hoc Special Contrasts on the Response Time (Experiment 2b)	86
Table 5: Post-Hoc Special Contrasts on the Response Time (Experiment 3a)	111
Table 6: Post-Hoc Special Contrasts on the Response Time (Experiment 3b)	131
Table 7: Mean Responsibility Judgement (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4b)	154
Table 8: Mean Responsibility Judgement (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4c)	158
Table 9: Mean Likelihood of recommending (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4c)	160
Table 10: Mean Responsibility Judgement (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4d)	164

Table 11: Mean Likelihood of Recommending (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4d)	166
Table 12: Mean Responsibility Judgement (Standard Deviation) as Function of the Question Frame, the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4e).....	170
Table 13: Mean Likelihood of Recommending (Standard Deviation) as Function of the Question Frame, the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4e).....	173
Table 14: Summary of the Main Effects and Interactions Found in Experiments 4a to 4e ..	177

CHAPTER 1: INTRODUCTION

Life is inherently risky. People face uncertainty at any time, for events from the lightest (will that girl's pistachio ice-cream taste as good as expected?) to the most severe (will hurricane Sandy hit a town?). Understanding risk therefore seems to be a skill particularly important to human beings. However, while toddlers can evaluate likelihood accurately based on the information found in their environment (Téglás, Girotto, Gonzalez & Bonatti, 2007), adults are well known for their poor performance when evaluating likelihood based on communicated information, even in numerical terms (e.g., Peters et al., 2006). This can be seen in the recently growing number of popular science books trying to inform about risk (e.g., Gigerenzer, 2002). As one cannot have experience of every possible situation, risk perception cannot be fully understood without understanding how communicated risk information is interpreted.

A particular format of communicating risk information is verbal probabilities (e.g., *It is likely*). Verbal probabilities are phrases communicating a degree of certainty, possibility or obligation of an event by a modal adjective (e.g., *likely, uncertain*) or a verbal form (e.g., *may + verb*). They are embedded in utterances expressing the probability of an event as in *It is likely that x will occur*. The modal term can be combined with a modifier (e.g., *few, strongly*) which will increase or decrease the degree of certainty expressed by the verbal probability (e.g., *It is highly likely that x is true*). This grammatical structure of verbal probabilities, based on a head (the modal term) and a potential modifier gives wide flexibility, so that there is a rich variety of verbal probabilities that can be produced and used. For example, Reagan, Mosteller and Youtz (1989) mentioned encountering 282 different verbal probabilities while reviewing 37 studies.

This thesis aims to examine how children and adults understand risk when it is communicated through the medium of verbal probabilities, and with a particular focus on the conversational features of verbal probabilities. I therefore present a brief discussion of probabilities, the difference between numerical and verbal probabilities, and normative and descriptive probabilistic reasoning. I then review literature on the understanding of verbal probabilities, first in adults, and then in children. Finally I argue in favour of the development of a pragmatics account of verbal probabilities.

1.1. Probabilities and probabilistic reasoning

1.1.1. Probability, likelihood, chance, odds, risk

According to the *Collins English Thesaurus* (2012), probability, likelihood, chance and odds are synonyms. Indeed, *Oxford Dictionaries Online* (2012) defines *probability* as ‘the quality or state of being probable; the extent to which something is likely to happen or to be the case’, and *likelihood* as ‘the state or fact of [something] being likely; probability’. *Odds* are ‘the chances of likelihood of something happening or being the case’ (*Oxford Dictionaries Online*). *Chance* is defined as ‘a possibility of something happening’, but also, as a mass noun, as restricted to positively valenced events (‘the probability of something desirable happening’; *Oxford Dictionaries Online*). Although it is not defined as a synonym of the previous terms, *risk* mirrors the term *chance* in being defined (as singular mass noun) as ‘the possibility that something unpleasant or unwelcome will happen’ (*Oxford Dictionaries Online*).

The term *probability* is however also defined further as a mathematical term, representing ‘the extent to which an event is likely to occur, measured by the ratio of the favourable cases to

the whole number of cases possible' (*Oxford Dictionaries Online*, 2012). In this it fits both what Baron (2008) referred to as the *frequency theory of probability* and the *logical theory of probability*. Under the frequency theory, the probability of an event can be estimated using the known frequencies of the event: for example, one can estimate the probability of having an accident on the M6 on a Sunday evening by comparing the number of accidents which had occurred on the M6 on Sunday evenings, to the number of vehicles which had driven on the M6 on Sundays evenings. Under the logical theory of probability, one can estimate a probability by using a number of mathematical rules. For example, to estimate the probability of obtaining two six when rolling two dices, one has to use the conjunctive rule, where the probability of a conjunction of events is equal to the product of the respective probabilities of each event.

The logical theory of probability can be used only if one can know that the different possibilities are logically equivalent. That is the number of possibilities is known and these possibilities are known to have equivalent probabilities. For example, in (non-biased) card games, the number of different cards is set up from the start and each is known to have only one occurrence. In the frequency theory of probability, one does not need to know the number of possibilities, nor the frequency of each. It is enough to know the frequency of the considered possibility, and the total number of occurrences. If one knows the number of car accidents in the United Kingdom and the total number of transport accidents in the United Kingdom, one does not need to know the number of truck or motorbike accidents to estimate a frequentist probability.

Following Savage (1954), Baron (2008) referred to a third theory of probability, the *personal theory*. In the personal perspective, the estimated probability is not drawn from an objective calculation, but from a personal judgement (Savage). Therefore, 'a probability judgement can be based on any of one's beliefs and knowledge, including knowledge about frequencies or about the set of logical possibilities, but including other knowledge as well' (Baron, p. 109). The personal theory of probability is thus inclusive of *subjectively* estimated probabilities, impacted by the different beliefs of each individual, but also of *objectively* estimated probabilities. For example, a lecturer could tell a student that he has a 40% chance to pass the test, although the pass base rate is usually 70%. This is because the lecturer also takes into account the poor record of attendance of the student. If the student had a high attendance, the lecturer might have told him that his chance of passing was 80%.

This work is focused on the use of probabilities as communicated by others, mainly without using numbers. People cannot reach an estimate by applying mathematical rules or based on frequencies, using communicated probabilities. They can only subjectively estimate the probability. Therefore in this thesis, reference to the term *probability* would always be in the personal probability framework. Thus, for clarity purposes, I will limit the use of the word *probability* (the only one of the synonyms which has a mathematical definition) to situations where an objective estimate can be reached using frequencies or logic. When I will refer to a probability estimate falling under the personal view, I will prefer the term *likelihood judgements*.

1.1.2. Normative and descriptive theories of probabilities

Under the frequency theory and the logical theory of probabilities, probability estimates can be correct or incorrect. There is a norm, defined by the application of mathematical rules (logical theory), or by the understanding of frequencies (frequency theory). Thus, normative theories of probabilities are concerned with the 'ought' of reasoning (Elqayam & Evans, 2011), and one's performance can be checked against the norm. Descriptive theories of probabilities are concerned with the 'is' (Elqayam & Evans), that is how people understand and use probabilities.

In the case of verbal probabilities, a strict norm cannot be reached through the logical or the frequency theory of probabilities. This thesis therefore can only provide a descriptive account of verbal probabilities. However I shall argue later that verbal probabilities can still call for a normative answer, albeit broader. Further, as De Neys (2012) highlighted, one can refer to a normative answer without being prescriptive. A normative answer can be simply used as a benchmark. In the context of reasoning and decision making, this can provide points of reference which make it easier to present a descriptive account (Stuppel & Ball, 2011). In this thesis I will thus sometimes refer to the normative answer (or most often, the correct answer) to help in presenting my descriptive account of verbal probabilities.

1.1.3. Irrationality: the dual model of reasoning and decision making

Piaget and Inhelder (1951) proposed that probabilistic reasoning is possible from 11-12 years old: 'A fundamental set of operational schemas which is also made possible by formal operations is the probabilistic notions which result from the assimilation of chance by those operations' (Piaget & Inhelder, 1966, p. 112; own translation). In the Piagetian theory, early

adolescence is part of the last stage of development, the formal operations stage. Also, in this theory, most individuals develop to this final stage. This entails that most individuals should be able to reason probabilistically.

However the seminal work of Prospect Theory (Kahneman & Tversky, 1979) highlighted that most adults do not reason correctly when using probabilities or when judging likelihood. For example, the conjunctive rule in logical probabilities requires that the conjunction of two events cannot be more likely than any of the two events. However, it can be that the conjunction of two events is more representative of the known or stereotyped environment than any one of the two events. In that case, people judge this conjunction as more likely than one of the two events composing it (conjunction fallacy; Tversky & Kahneman, 1983).

Another bias identified by the Prospect Theory is the tendency to judge likelihood based on the availability of an exemplar in memory (availability heuristic; Tversky & Kahneman, 1973). For example, if one is considering whether it is more likely to die in a transport accident or of a fall in England or Wales, one might think it is the former. The answer, based on the frequencies for 2010, would be that, in fact, a person based in England or Wales is more likely to die of a fall (Office National for Statistics, 2011). However, transports accidents are more likely to be reported in the news than falls (as they are more disruptive of the public life), making their exemplars readily available.

Reasoning biases highlight a gap between the rules of reasoning and reasoning as applied in daily life by lay people. The rule-based solution, that is the norm, is not always applied, suggesting that a normative theory of reasoning (including probabilistic) alone cannot account

for the empirical evidence. Dual-process theories have therefore been proposed (e.g., Kahneman, 2003) to explain adults' failures to attend to the norms of reasoning and decision-making. Under the dual-process frame, people are thought to use either one of two systems of reasoning. System 1 is an intuitive system of reasoning and decision making, relying mainly on the use of heuristics. System 2 is the reasoning system, relying on the use of rules. While System 1 is automatic, and thus fast and requiring little cognitive resources, System 2 operates in a controlled way, hence is slow and cognitively costly (Kahneman, 2003).

As showed in Evans' review (2008), there is considerable empirical evidence to support a dual-process account of reasoning and decision-making. However, Evans argued later (2010) that while the definition of System 1 and System 2 thinking implies different cognitive mechanisms, evidence is scarce, if not contradictory, for different cognitive architectures being involved in the two different types of thinking. In fact as Evans (2010) pointed out, imaging evidence suggests that more than two systems could be involved (see Goel, 2008). He proposed using a terminology of 'Type 1' and 'Type 2' reasoning instead (Evans, 2010). In this terminology, 'System 1 and 2 really correspond to two *families* of systems that have the Type 1 and 2 characteristics' (p. 316). Referring to type rather than system allows therefore accounting for the lack of evidence of precisely two different systems and the possibility that multiple systems are involved. Thus, in this thesis I will follow Evans (2010) and refer to Type 1 and Type 2 when referring to dual-process thinking.

1.1.4. Probabilities as communicated: verbal vs. numerical probabilities

When decision-makers need to judge subjective probabilities, they can observe the environment and/or use their own knowledge. For example, if a couple wants to judge the likelihood of failure of different contraceptive methods before choosing one, they can sample their environment for the frequencies of pregnancies they know to be due to a contraception failure. They can also use what they remember of the information provided to them in high school, in biology or sex education classes, should they have received some. If they do not remember, a quick internet search will provide them with the probability of success of the different methods, most probably in percentages, i.e. in numerical format. However they might find it quicker and simpler to ask a health professional for this information, rather than search for it. If they ask a health professional they are more likely to be provided with this information in a verbal rather than numerical format, through the medium of *verbal probabilities* (Erev & Cohen, 1990).

Erev and Cohen (1990) found that although students preferred to place a bet based on a numerical probability, sports commentators preferred to give their predictions as verbal probabilities. Self-report methods highlighted the so-called Communication Mode Preference paradox: a third of people prefer to receive numerical probabilities but to give verbal probabilities (Wallsten, Budescu, Zwick & Kemp, 1993, for an English-speaking population; Xu, Ye & Li, 2008, for a Chinese-speaking population). In fact, the overall preference for receiving numerical probabilities (disregarding their preference for giving numerical or verbal

probabilities) was found in two thirds of each sample. Whereas two thirds of each sample preferred to give verbal probabilities (whichever their preference for receiving).

Wallsten et al. (1993) reported data suggesting that the Communication Mode Preference paradox relied on the perception that verbal probabilities are easier to use and more natural than numerical probabilities. In contrast, numerical probabilities were perceived as more precise than verbal probabilities. This is consistent with the recent appearance of probability as mathematical concept (from the correspondence between Pascal and de Fermat, c. 1650-60; Hacking, 1975), on the scale of human evolution. According to Gigerenzer (e.g., 1998), this late phylogenetic appearance entails that frequencies are the natural format for the human species to process probabilistic information. That is, it is easier to process frequencies than percentages because the human mind has evolved to do so. Moreover, while *likely*, *certain* and *possible* appeared in the English language circa 1300, and *probable* in the late 14th Century, *probability* was first used in its mathematical sense in 1718 (*Online Etymology Dictionary*, 2012). This could suggest that verbal probabilities feel more natural because they are more established in the daily language thanks to their earlier appearance.

Beyth-Marom (1982) proposed that numerical probabilities are avoided to stop forecasts from being judged on their accuracy. This explanation is not incompatible with the claim that verbal probabilities are more natural. One could prefer verbal probabilities because they feel more natural and consequently have the benefit that one's accuracy is not evaluated. However, this explanation was ruled out by Erev and Cohen (1990). After asking four experts to produce verbal and numerical predictions, Erev and Cohen asked them to predict which of the verbal or

numerical prediction would bring more money to gamblers. Experts received money for answering correctly to this question, either as flat rate or as incentive. The reward format had no effect on the preferences shown by the experts. That is, the experts produced verbal probabilities more often, disregarding the reward condition. This suggests that an incentive to be accurate does not reduce the preference for producing verbal probabilities. Thus verbal probabilities, according to Erev and Cohen, are not chosen because they allow speakers to cover their inaccuracies. However, that speakers are not influenced by rewards based on accuracy is not exactly the same thing as speakers being judged less responsible for using a verbal probability. That is, judgements of accuracy and judgements of responsibility might not overlap. In Chapter 5, I will come back to this possibility to test it indirectly, exploring how receivers judge speakers as responsible when using different formats to predict uncertain information.

Searching language corpora also supports that receiving probabilistic information in a verbal format is likely in everyday life. For example, the *Corpus of Contemporary American English* (Davies, 2008-2012) found *likely*, *possible*, *certain* to all be in the 1,000 most frequent words (625th, 460th and 578th, respectively). In the *British National Corpus* (Leech, Rayson & Wilson, 2001), those three modal adjectives are in the top 500 words, *possible* occurring as frequently as *long*, *likely* as often as *real* and *certain* as often as *difficult*. A month of British news offers 17,300 occurrences of the verbal probability *it is likely*, and 24,900 occurrences of *it is possible* (search run on Google News UK, for one month upwards November, 11th, 2012).

Even in childhood, verbal probabilities are encountered early. For example, in a corpus of French vocabulary found in primary school handbooks (Lété, Sprenger-Charolles & Colé, 2004),

possible occurred as often as *neighbour*. It seems therefore that daily life provides a lot of opportunities to judge likelihood and make decisions based on verbal probabilities rather than numerical, both for adults and for children. Actually, since children are not taught about numerical probabilities until the end of primary school, they may even receive verbal probabilities to a higher extent than adults, even if only the simplest or most common phrases (e.g., *it is likely*).

1.2. Judging the likelihood of verbal probabilities

1.2.1. The problem of equivalence between verbal and numerical probabilities

Lichtenstein and Newman (1967) described the likelihood meaning of a wide range of verbal probabilities. Lichtenstein and Newman asked participants to translate 41 verbal probabilities into a likelihood from .01 to .99. The translations given by their sample were consistent for each verbal probability. However, Lichtenstein and Newman noticed that Cohen, Dearnley and Hansel (1958) reported very different translations in their study of three verbal probabilities. For example, *likely* was translated as lower in Cohen et al. than in Lichtenstein and Newman, but the reverse was true of *improbable*. Following Lichtenstein and Newman's study, most studies consisted of participants translating verbal probabilities into likelihood, via 0 to 1 or 0 to 100 scales. This has been done with general samples (e.g., McGlone & Reed, 1998), expert samples (e.g., Beyth-Marom, 1982) and patient samples (e.g., cancer patients: Sutherland & al., 1991). In other studies, participants were simply asked to rank the verbal probabilities in order of likelihood (Reyna, 1981). What was observed in these translation studies is paradoxical. The likelihood assigned to each verbal probability remained more or less stable between studies, i.e.

there was little variability between different groups of participants. This could lead to the conclusion that verbal probabilities each have a numerical equivalent. But what could also be noticed from these studies is a great variability within each group of participants. So even if on average likelihood translations of verbal probabilities are stable, the interindividual variability may prevent one being sure of being understood when communicating uncertainty verbally.

This observation from translation studies led several authors (e.g., Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986) to characterize verbal probabilities by their vagueness. Nakao and Axelrod (1983) even advised that use of verbal probabilities should be abolished in medical risk communication. Less extremely, several authors (e.g., Hamm, 1991) have suggested specifying a list, as a lexicon, which would guide the use of verbal probabilities. Considering the late appearance of probabilistic reasoning in the human phylogeny, this 'vagueness' may not be surprising however. If adults have not evolved to use probabilistic concepts (e.g., Gigerenzer, 1998), they should not be expected to be able to use those concepts in order to translate verbal probabilities that they might otherwise use with ease in their natural language. It is also well documented that the understanding of the numerical probability scale is incorrect in the general population: adults overestimate small probabilities and underestimate large ones (e.g., Gonzales & Wu, 1999). This can be observed when an increase from 5% to 10% is judged as larger than an increase from 40% to 45% (Quattrone & Tversky, 1988). Conversely an increase from 90% to 95% is judged as smaller than an increase from 40% to 45%. It suggests that the representation of the probability scale is not linear, with equal intervals, but rather has the

shape of an inverted S. This could lead people to have difficulties in using linear probability scales when translating verbal probabilities, creating variability in the results.

One could even argue that asking people to translate language used since the Middle-Ages with concepts which have appeared afterwards is looking at the problem from the wrong side. That is, since verbal probabilities were in use before mathematical probabilities, how they are understood is unlikely to have been shaped by the appearance of the logical theory of probabilities. On this view, having to translate verbal probabilities into mathematical language may not be natural, which prevents us from acquiring fine grained and stable translations. This is not to say that verbal probabilities cannot elicit subjective probability judgements, but rather than they are not a mirror to numerical probabilities, as seems to be assumed in translation studies. This is supported by Windschitl and Wells (1996), who found that verbal probabilities elicit intuitive reasoning, prone to context effects (characteristic of Type 1), while numerical probabilities elicit rule-based reasoning (characteristic of Type 2). In their words, 'people who have been asked to provide a numerical uncertainty estimate think differently about the presented information than those who have been asked to provide verbal uncertainty estimates' (p. 358).

I want to argue that verbal probabilities are only vague if they are considered as communicating numerical information. As just highlighted, their use of verbal probabilities in natural language might in fact be unambiguous, and speakers might perfectly understand each other, even if they do not translate what is said in equivalent numerical probabilities. Providing a pragmatic account, I will suggest that verbal probabilities have foremost an argumentative

function (Anscombe & Ducrot, 1983), and that the interpretation of this function is unambiguous, as shall be showed later in this chapter.

1.2.2. A numerical solution to the vagueness of verbal probabilities: the membership function

To address the problem of defining numerical equivalences for verbal probabilities, Wallsten and colleagues (e.g., Wallsten et al., 1986; Dhimi & Wallsten, 2005) tried to characterize verbal probabilities in a less strict way than translation studies. In Wallsten et al., for example, participants used spinners to indicate what were the lowest and the highest probabilities for which they might use a verbal probability. The range defined by this first answer was divided in a maximum of seven intervals (it could need to be fewer intervals to avoid having intervals smaller than .02). For example if a participant had first defined that *it is likely* was at the lowest .40 and at the highest .75, the range would be divided in seven intervals of .05. *it is likely* could be thus associated with .40, .45... up to .75. For each probability obtained in this way, participant then judged how much it was appropriately described by the initial verbal probability. That is, participants indicated a range of probabilities associated to one particular verbal probability. Within that range, they indicated how much each probability point was appropriately described by the verbal probability. Verbal probabilities can then be characterized by membership functions.

In a membership function, a verbal probability is represented by the range of numerical probabilities it can be translated into, rather than by a single numerical probability. Furthermore, each point of the range is characterized by a level of adequacy with the verbal probability, so that a membership function represents graphically the distribution of

probabilistic meaning of the expression (see Figure 1 for an example). This theoretical development not only took into account the numerical vagueness of verbal probabilities, but also allowed Wallsten et al to include the interindividual variability, allowing more overlap between two people's numerical translations of a verbal probability.

1.3. Verbal probabilities are language: pragmatics and argumentative function

1.3.1. Verbal probabilities and locus of uncertainty

Shortly after the membership function development allowed the numerical vagueness of verbal probabilities to be taken into account, Teigen (1988) proposed that verbal probabilities entail more than a numerical equivalence. Teigen extended Kahneman and Tversky's (1982) proposition to distinguish between *internal* and *external* uncertainty, and Hacking's (1975) distinction between *epistemic* and *aleatory* probabilities, to the case of verbal probabilities. What Hacking referred to as epistemic and Kahneman and Tversky as internal is the uncertainty

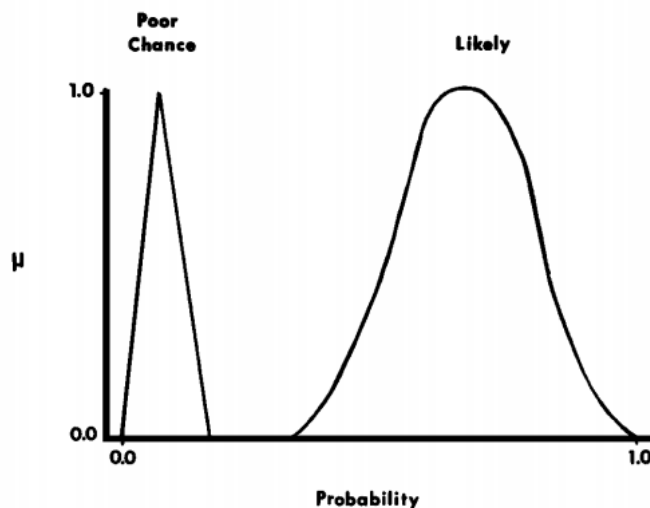


Figure 1: Hypothetical membership functions for two probability terms (from Wallsten & al., 1986)

that reflects the degree of knowledge or belief of an individual. Aleatory, or external, uncertainty relies on the state of the world, that is the distribution of events in the world. Under Teigen's account, *I am certain* and *It is certain* refer respectively to internal and external uncertainty. When using the internal verbal probability *I am certain*, uncertainty is cognitive (epistemic), that is speakers refer to their own state of knowledge. When using the external verbal probability *It is certain*, speakers refer to the uncertainty that resides in the physical world.

This distinction is generally supported by a preference for betting when uncertainty is internal (although when imagining a hypothetical betting scenario, preference is for external uncertainty; e.g., Robinson, Pendle, Rowley, Beck & McColgan, 2009). Research into the different interpretations that can be made from internal and external verbal probabilities has been recently developed (see e.g., Fox & Ülkümen, 2011; Juanchich, Teigen & Gourdon, 2011). Although results are still preliminary, Juanchich et al. found that external verbal probabilities are more likely to be interpreted as indicating that the speaker used statistical information to draw the prediction; they are also interpreted as being more informative. In this thesis however, this aspect of verbal probabilities will not be explored, and in all upcoming experiments, the verbal probabilities will only be external (e.g., *It is...*, *There is...*).

1.3.2. Verbal probabilities and directionality

In 1988, Teigen also proposed that verbal probabilities should be distinguished based on their *directionality*. The concept of directionality was drawn from psycholinguistic work on quantifiers (e.g., Moxey & Sanford, 1986), and can sometimes also be called *polarity* (e.g.,

Sanford & Moxey, 2003; in this thesis I will refer only to directionality for consistency).

Directionality of a quantifier or a verbal probability is dichotomous, being either positive (e.g., *it is likely*) or negative (e.g., *it is unlikely*). Being so, it focuses the listener's or the reader's attention on the occurrence of an uncertain event (if positive) or on its non-occurrence (if negative). The positive or negative directionality of a verbal probability can be established through a continuation task (Teigen & Brun, 1995). Participants are given predictions that include verbal probabilities, for example, *It is very likely that Tom will get a 1st on his exam*. These statements must then be completed with a reason to justify them. A positive directionality (as in the example) elicits the production of reasons in favour of the uncertain event (e.g., *because he has worked hard all term* to complete the example statement). This is thought to be because it focuses on the occurrence of the uncertain event. Conversely, a negative directionality elicits reasons adverse to the uncertain event, since it focuses on its non-occurrence; for example, *It is unlikely that Joe will get a 1st on his exam* will most often be completed by reasons such as *because he missed a lot of classes*.

According to Moxey and Sanford (2000), directionality supports an argumentative function of language (e.g., Anscombe & Ducrot, 1983). According to the argumentative theory of language, every statement aims to inform but is also an argumentation, and its structure frames the inferences that can be drawn from the statement (Anscombe, 1989). The results found by Teigen and Brun (1995) with the continuation task were consistent with such a perspective. When presented with a positive directionality, participants proposed reasons in favour of the uncertain event (pro-reasons); when presented with a negative directionality, they proposed

reasons against the uncertain event (con-reasons). For the majority of verbal probabilities, all the participants provided only pro-reasons, or only con-reasons, illustrating how much the perspective of language can frame inferences. Furthermore, directionality leads to framing effects in the interpretation of verbal probabilities (Teigen & Brun, 1999). A positive (*there is some possibility*) and a negative verbal probabilities (*it is quite uncertain*) were judged by a first group of participants as yielding similar translations in numerical probabilities (30-35 %). A second group of participants then decided to recommend or not a headache treatment, introduced either by *There is some possibility* or by *It is quite uncertain*. Those participants recommended taking the headache treatment more often when presented with the positive verbal probability (*There is some possibility*).

In light of these studies, considering verbal probabilities as only expressing a quantity appears as a restricted view, in that natural language expressions of quantity also convey information about perspective (Sanford & Moxey, 2003). Consistent with this view, they can also convey information about the source of information (see section 1.3.1). By choosing a directionality, speakers indicate a perspective which they wish to be taken over the other. They signal what they think is important, and this information might be critical for listeners to make appropriate decisions.

Moxey and Sanford (2000) even suggested that it may be why people prefer to use verbal probabilities or quantity statements over numerical ones. Verbal probabilities potentially protect speakers from being held responsible by their numerical vagueness (Beyth-Marom, 1982), and they feel more natural (Wallsten et al., 1993). But also, verbal probabilities or

quantity statements allow speakers to try to orient the attention of their addressees where they wish to, and to hint at the nature of the information they used to make their prediction (Juanchich et al., 2010).

Budescu, Karelitz and Wallsten (2003) have, however, argued that directionality is only a secondary feature of verbal probabilities, which relies on the level of chance. That is, verbal probabilities with low likelihood have a negative directionality, and verbal probabilities with a high likelihood have a positive directionality. Yet, Gourdon and Villejoubert (2009) observed that some negative verbal probabilities could be judged as meaning a higher likelihood than would be some positive ones. For example, French-speaking adult participants in this study produced likelihood judgements of 57% for *It is not absolutely sure (Il n'est pas totalement sûr)*, but of 45% for *There is a little chance (Il y a une petite chance)*. Gourdon and Villejoubert argued that directionality should be considered as a pragmatic cue, indicating the communicative intentions of the speaker. They also suggested introducing social factors such as speaker's benevolence to test this possibility, which I do in Chapters 2 (in childhood) and 4 (in adulthood).

Further evidence can be found to support a pragmatic account of verbal probabilities. Juanchich, Teigen and Villejoubert (2010) investigated the choice of verbal probabilities in a conversation after a first verbal probability was uttered to express the likelihood of guilt of a suspect. Participants were given the newly revised likelihood of guilt and had to choose between a positive and a negative verbal probability to express it. They found that participants chose positive verbal probabilities when the likelihood had been revised upwards. When it had been revised downwards, they chose negative verbal probabilities.

More recently, Juanchich, Sirota and Butler (2012) showed that verbal probabilities are the most often interpreted as hedging devices, using vagueness to protect speakers from listeners blaming them, or to protect listeners from a bad news. For example, participants read verbal probabilities before choosing a conversational interpretation among three (the speaker wants to communicate uncertainty; the speaker does not want to deliver the news too harshly; the speaker wants to be cautious in case the prediction would be incorrect). If the uncertain event was positive, participants most often said that the speaker wanted to be cautious in case the prediction turned out to be wrong; if the uncertain event was negative, participants most often said that the speaker was avoiding delivering bad news too harshly.

1.4. Interpretation of verbal probabilities in childhood

Despite the prevalence of uncertainty communication in everyday life (Lété, Sprenger-Charolles & Colé, 2004), very few studies have investigated the interpretation of verbal probabilities in childhood (but see, for exceptions, Moore, Pure & Furrow, 1990; Mullet & Rivet, 1991; Watson & Moritz, 2003; Gourdon & Villejoubert, 2009). Moore et al. found that children as young as 5 years old are able to distinguish between two verbal probabilities to make an appropriate choice. Children had to decide which box (of two) to choose, after receiving two predictions (e.g., *it's probably in the blue box* and *It's maybe in the red box*). Their scores were compared to what the majority of the sample chose to do. This showed a development of understanding of verbal probabilities between 3 and 5 years old, children more reliably choosing the appropriate box from 5 years old.

However, Mullet and Rivet (1991; see also Watson and Moritz, 2003) observed that older children distinguished the different levels of numerical value carried by verbal probabilities less well compared to adolescents. More precisely, Mullet and Rivet found that 9-year-olds could distinguish clearly between only five positions on a likelihood judgement scale. As Moore et al. used only three different verbal probabilities, these different results are not incompatible; comprehension of verbal probabilities could be considered as starting early with simple expressions and developing towards discrimination of more complex ones. Indeed, Moore et al. used single-word verbal probabilities, i.e., non-modified modal terms only (e.g., *possibly*), which are probably simpler than verbal probabilities made of a modal term and a modifier (e.g., *it is quite likely*).

Gourdon and Villejoubert (2009) raised a concern regarding the directional nature of verbal probabilities. None of the three first studies considered directionality as a factor potentially influencing the comprehension of verbal probabilities. In the case of Moore et al. (1990), the authors not only used simple verbal probabilities, but ones with a positive directionality. It cannot be sure therefore if children would have shown the same early understanding of verbal probabilities, should have they been of negative directionality. In the two other studies, both directionalities were used, such that it cannot be sure that the latter development of the understanding of verbal probabilities is not confounded with the different directionalities being used. It is possible that children in these two studies have performed better on the positive verbal probabilities, but this is not possible to know given the designs.

However, consistent with Gourdon and Villejoubert's results, studies of the understanding of quantifiers by children found that children display an early sensitivity to the argumentative function of language. That is, they display an understanding that the language is subjective and can be interpreted differently based on the context. Champaud and Bassano (1987) found that 6-year-olds could already take in consideration directionality of quantifiers. Children were presented with statements such as *I have barely 6 beads (J'ai à peine 6 perles)* or *I am barely reaching the green line (J'arrive à peine au trait vert)*. Children had to identify the speaker who uttered the sentence, in an array of dolls with boxes containing different amount of beads, or in an array of dolls against a wall with different coloured lines drawn on. This was to test their understanding of the informative function of the statements. Children were then asked what the speaker meant, to test their understanding of the argumentative function. This was tested with closed questions, such as *Does he mean that he is tall enough, or not tall enough? (Est-ce qu'il pense qu'il est assez grand ou pas assez grand?)*. In the argumentative task, children performed similarly to adults as early as 6 years old, displaying an understanding of the negative directionality.

Finally, the different results observed by Moore et al. (1990) on the one hand, and Mullet and Rivet (1991) and Watson and Moritz (2003) on the other hand could also be explained in part by the different task used by Moore et al.. While children had to judge likelihood using probability scales in the two other studies, in Moore et al. they had to make a choice regarding a verbal probability. Gourdon and Villejoubert (2009) asked 8-year-olds and adults both to make likelihood judgements on probability scales and to make decisions based on verbal probabilities.

For example, participants were presented with four different verbal probabilities predicting the likelihood of a treasure being inside a treasure chest. The verbal probabilities were chosen to reflect either a high or a low likelihood, and had a positive or a negative directionality. Thus children could judge likelihood and make decisions based on directionality only, on likelihood only, or an interaction of both. Children displayed patterns of answers similar to adults in the decision-making task, with choices reflecting both the likelihood and directionality. However in their likelihood judgements, they only took into account the directionality of verbal probabilities. Gourdon and Villejoubert argued that this difference was due to the decision-making task bearing more concrete consequences (while the likelihood judgement remains only an evaluation), with mistakes entailing punishment.

Moore et al. (1990), and Mullet and Rivet (1991) have used respectively a decision-making task, and a likelihood judgement task. This could explain, following Gourdon and Villejoubert's (2009) claim, why they found different ages at which the discrimination between levels of likelihood appears.

1.5. Overview and outline of the current experimental work

In the review of the literature presented in this chapter, I have outlined that verbal probabilities are not a strict mirror of numerical probabilities. Verbal probabilities can elicit subjective probabilities, but this does not necessarily imply that they are a representation of numerical probabilities. Even if they are a representation of numerical probabilities, because they are language-based, they fulfil argumentative functions (e.g., they focus one's perspective on a particular aspect of a situation; e.g., Teigen & Brun, 1999) and further informative

functions (e.g., they support inferences regarding the source of information of the speaker; Juanchich, Teigen & Gourdon, 2010). Verbal probabilities are also often considered as hedging devices (e.g., Juanchich, Sirota & Butler, 2012), in which cases the likelihood judgements they elicit is different than if they are considered as informative devices. These different studies, taken together, seem to suggest that verbal probabilities are a pragmatic tool, in that they are interpreted in context (e.g., need for politeness) and signal an argumentation by their directionality. This thesis therefore aims to give an account of verbal probabilities that is driven by a pragmatic interpretation.

In Chapter 2, I will aim to develop further the account of verbal probabilities in childhood (Chapter 2), in light of a pragmatic account. Drawing on Gourdon and Villejoubert's work (2009), I will first investigate how English-speaking children understand verbal probabilities. The pragmatic account will be then tested by looking at how the intention of a speaker can interact with directionality of verbal probabilities (Chapter 2 and Chapter 4). That children use only directionality when judging the likelihood of verbal probabilities (Gourdon & Villejoubert) could suggest that directionality bears less cognitive demands. Therefore Chapter 3 will try to disentangle the respective processing demands of directionality and of likelihood in verbal probabilities. Finally, Chapter 5 and Chapter 6 will draw on the suggestion of Beyth-Marom (1982) to try and explain indirectly the Communication Mode Preference paradox (Erev & Cohen, 1990). Chapter 5 will investigate if judgements of the responsibility of forecasters are different when the forecaster has used a verbal or a numerical verbal probability. Chapter 6 will investigate if the preference for receiving numerical probabilities depends on speakers'

expertise; in a Gricean pragmatic account, an expert is expected to know more, and therefore should be expected to use a more precise mode to communicate probabilities.

CHAPTER 2: CHILDREN'S UNDERSTANDING OF VERBAL PROBABILITIES

2.1. Introduction

Imagine that an 8-year-old boy wants to invite a friend to go to the park with him during the week-end. Before inviting him he asks his parents whether it will rain. Should they answer (1) or (2)?

(1) There is a 20% chance that it will rain.

(2) There is a small chance that it will rain.

In many countries an 8-year-old has not yet been taught about numerical probabilities (percentages and frequencies). However, a corpora study of French language used in primary school handbooks (Lété, Sprenger-Charolles & Colé, 2004) revealed that children are confronted with uncertainty words as early as first grade: for example, *possible* occurs as often as *neighbour* in first grade handbooks. What is more, according to the preference paradox (e.g., Erev & Cohen, 1990), adults are likely to prefer to use the verbal probability (*There is a small chance*) rather than the numerical probability to communicate uncertainty.

Besides understanding what his parents mean by *There is a small chance*, the boy will also have to decide if he should invite his friend on the basis of this statement. Children, like adults, have to make decisions daily. To do so they can rely either on experience or on information communicated to them. Since adults are likely to communicate uncertainty to children using verbal probabilities (Erev & Cohen, 1990), the study of children's decision-making activities would benefit from knowledge of how they understand verbal probabilities and make decisions

on this basis. Yet, very few studies have investigated the comprehension and use of verbal probabilities in childhood (exceptions are Moore, Pure & Furrow, 1990; Mullet & Rivet, 1991; Watson & Moritz, 2003; Gourdon & Villejoubert, 2009). Furthermore only two of those studies (Moore & al., 1990; Gourdon & Villejoubert, 2009) investigated decision-making.

Moore et al. (1990) found that children as young as 5 years old are able to distinguish between higher and lower likelihoods of verbal probabilities to make an appropriate decision. For example, when choosing between a blue box that *maybe* contained a candy and a red box that *probably* contained one, 5- to 6-year-olds appropriately picked the 'probable' box. Mullet and Rivet (1991) found development continuing into late childhood with older children distinguishing the levels of likelihood carried by verbal probabilities less well than adolescents. More precisely, Mullet and Rivet (1991) established that 9-year-olds could distinguish clearly between only five positions on a probability judgement scale. In this study, 9-year-olds and 15-year-olds were given 12 expressions of probability and judged them on a probability scale. While the older children could discriminate most judged expressions (that is, placed them on different points of the probability scale), the 9-year-olds only placed the expressions on 5 different points of the probability scale. For example, 15-year olds placed on average *nearly certain* and *small chance* on different points of the scale, the younger group' judgements of the two expressions overlapped.

Watson and Moritz (2003) surveyed a sample of school children from Australian Grade 6 to Grade 11 (presumably 10-11- to 15-16-year-olds, although this was not mentioned by the researchers). They presented them with seven verbal probabilities and a percentage expression,

and asked them to translate them on a probability scale. The verbal probabilities were translated with variability, with the exception of the percentage expression and *impossible*. They also scored how precisely and appropriately students evaluated each expression. They found improvement between Grade 8 and 9, and between Grade 9 and 10. Their level of evaluation scoring method valued translations within a smaller range; for example translating *impossible* as less than 10% was scored as an advanced evaluation, but translating it only as less than 25% was scored as basic evaluation. The scoring method also valued answers consistent with the expected ranking of the verbal probabilities. Although the level of evaluation of students developed with age, only half of the older groups displayed Comprehensive Evaluation (the highest level of evaluation in their scoring system). That is only half of older students were showing fine translations and consistence in the ranking of the verbal probabilities.

A few differences between these three studies are important to note. In Watson and Moritz (2003), the low verbal probabilities were derived from modals that are frequent in daily language (*unlikely, impossible, in doubt*). But the high verbal probabilities were informal (e.g., *no worries*), and it could even be argued that they were not exactly verbal probabilities (e.g., *looking good*). Mullet and Rivet (1991) used nine verbal probabilities, based on frequent modal words (e.g., *likely, chances*). They also used three verbal expressions of frequency (e.g., *a-one-in-four chance*). Finally Moore et al. (1990) used only three different verbal probabilities (*maybe, probably, possibly*). Moreover, Moore et al. used single-worded verbal probabilities, i.e. non-modified modal terms only (e.g. *possibly*). These differences in materials might suggest that the different results are not incompatible. The comprehension of verbal probabilities could

therefore be considered as starting early with simple expressions (Moore et al.) to develop slowly towards distinction of more complex (Mullet and Rivet) and informal (Watson and Moritz) ones. Furthermore the early abilities observed by Moore et al. (1990), and later competence reported in Mullet and Rivet (1991) and Watson and Moritz (2003), may also result from the types of tasks used. The latter required children to make likelihood judgements on scales, whereas Moore et al. asked them to make a choice regarding a verbal probability. Gourdon and Villejoubert (2009) asked participants to do both: judge probabilities on scale and to make decisions based on verbal probabilities.

Gourdon and Villejoubert (2009) also observed that none of these studies considered directionality as a factor potentially influencing the comprehension of verbal probabilities. In the case of Moore et al., the authors used only positive verbal probabilities. In the two other cases, both directionalities were used, but not investigated systematically. In fact in Watson and Moritz it was possibly a confounded variable as all the low verbal probabilities had a negative directionality while most of the high ones had a positive directionality. This is one aspect of children's understanding that I will investigate in this study, expanding on Gourdon and Villejoubert's work. In their study, they manipulated factorially directionality and likelihood of verbal probabilities. Thus, they presented children and adults with low verbal probabilities of positive directionality, low verbal probabilities of negative directionality, high verbal probabilities of positive directionality, and high verbal probabilities of negative directionality. In their study, French-speaking 8-year-olds behaved very similarly to adults when making decisions. That is they decided based both on directionality and on the likelihood: they chose to

open a chest more often for positive high verbal probabilities, followed by negative high verbal probabilities, positive low verbal probabilities, and negative low verbal probabilities. However, when judging the likelihood of the different verbal probabilities, they seemed to perform more poorly, using only directionality. Higher likelihood judgements were given to positive high verbal probabilities, followed by positive low verbal probabilities, negative high verbal probabilities, and negative low verbal probabilities.

In Gourdon and Villejoubert's (2009) study, the discrepancy between children's decision-making and translation performance could be explained by the paradigm, where the order of the tasks was kept fixed, with the likelihood judgement always coming first. What is more, children and adults were presented with the verbal probabilities twice for each judgement, first in a training task meant to allow calibration of their judgement, and then in the experimental task. This could explain the difference in performance on the two tasks. In order to understand how children interpret and make decisions based on verbal probabilities it is important that task order is controlled.

To sum up, developmental studies had varied in the tasks presented to children. In studies showing younger success, children made decisions, whereas in those that older children found difficult they were asked to translate the verbal probabilities. Gourdon and Villejoubert (2009) tried to contrast these two measures, but task order was not controlled. Furthermore, Gourdon and Villejoubert's findings need to be replicated in English to ensure the findings are generalizable. Gourdon and Villejoubert also studied only one age group. Therefore in my studies I expanded the age range downwards to look for developmental change. However, as

the focus was on the development of the sensitivity of directionality, which was already showed in 8-year-olds in Gourdon and Villejoubert, I did not extend the age range upwards. Finally, in Gourdon and Villejoubert, expressions were selected from the literature but not piloted. As a result, the high and the low expressions were judged differently also by adults (29% to 45% for low expressions, 57% to 71% for high expressions), which introduced a potential confound. In this chapter, piloted data will be used to avoid this.

If Gourdon and Villejoubert's (2009) results were not an artefact of task order or language, then I should observe similar performance by 8-year-olds in my study. That is, 8-year-olds should use directionality only when judging likelihood, but use both directionality and the likelihood to make decisions. To further validate this finding I also added in an expected value judgement. I would expect the judgements made here to follow the same pattern as for the decision-making measure. That is, expected value judgement would be higher for positive high verbal probabilities, followed by negative high verbal probabilities, positive low verbal probabilities, and negative low verbal probabilities. This is because the expected value judgement is supposed to be the product of the likelihood of an event and of its utility. In behavioural economics theory, *utility* is 'the state of being useful, profitable, or beneficial' (*Oxford Online Dictionary*, 2013). By taking into account the utility, the expected value judgement considers the consequences of a decision and it has been proposed that thinking about the consequences is what makes children sensitive to the likelihood (Gourdon & Villejoubert, 2009).

2.2. Experiment 1a

2.2.1. Method

2.2.1.1. Participants

Twenty-three 7- to 8-year-olds (mean age= 7;8; range= 7;3-8;2; 11 girls) and 24 8- to 9-year-olds (mean age= 8;8; range= 8;4-9;2; 15 girls) participated. They were enrolled in a community primary school of a multiethnic and working class area of Birmingham (UK). Participants received a sticker in exchange for their participation.

2.2.1.2. Materials

I chose verbal probabilities based on a pilot study conducted with Psychology undergraduates of the University of Birmingham, UK (N = 38). These participants received course credits in exchange for their participation. They judged 33 expressions (see Appendix I, p. 209) according to their likelihood (answering to the question *What is the probability that the event will happen?* on an 11 point scale) and according to their plausibility (answering to the question *How plausible is it to hear this expression?* on an 11 point scale). I then chose expressions in order to manipulate the likelihood and directionality in a 2x2 within design, so that there was a similar range likelihood within each level of directionality (both positive and negative phrases had either low likelihood = 35-37% or high likelihood = 53-55%). The expressions having highest plausibility ratings were then preferred. The four expressions chosen to be used can be seen in Table 1.

Four different scenarios were then built around common childhood events (see Appendix II, p. 210). In each of these children were given uncertain information via one of the verbal

Table 1

Verbal Probabilities Used in Experiments 1a and 1b

	low likelihood	high likelihood
positive verbal probability	<i>There are a few chances</i>	<i>There is some possibility</i>
negative verbal probability	<i>It is a little unlikely</i>	<i>It is not entirely certain</i>

probabilities, which were counterbalanced across the four scenarios. Therefore, each child received all four scenarios, with all four verbal probabilities, albeit not in the same combinations. Each verbal probability was judged on three measures: to assess the likelihood (*When the clue says [VP], how much do you think that this means that the treasure is in the chest?*), the expected value (*How happy do you think you would be if you [uncertain choice]?*) and decision-making (*If you were him/her, what would you choose to do?*). The order of these questions was counterbalanced between children (i.e., an individual always heard the measures in the same order, to avoid confusion, but the order varied between participants).

A training task (see Appendix II, p.210) was built based on the treasure hunt scenario used by Gourdon and Villejoubert (2009). This scenario was used three times, with two anchoring verbal probabilities (*It is absolutely sure* and *It is impossible*) and one of medium likelihood (*It is not certain*) to learn to use the scales. The order of the questions in the training task matched the order in the experimental task (see Appendix III, p. 213, for an example of how the scenarios and questions were presented to participants). Children were not provided feedback over the training trials.

The training and the experimental tasks were presented in a booklet. Each page of the booklet contained one scenario and the three tasks (likelihood judgement, expected value judgement and decision making; counterbalanced), amounting to seven scenarios. On the top of each page there was a picture illustrating the specific scenario of the page (e.g., a treasure chest for the training scenarios involving a treasure hunt). An example of a scenario page is provided in Appendix III (p. 213).

2.2.1.3. Procedure

Each participant took part individually, in a quiet area of the school. The scenarios booklet was placed between the experimenter and the participant so the participant could follow while the experimenter read through the scenarios. The experimenter first read participants the following statement: 'Here are some stories about children of your age. After I read them to you, you are going to play some little games about these stories.' Then the experimenter read each scenario and each question one by one to the participant, making sure that the participant was meanwhile following on the booklet. After each question, participants answered themselves, in the booklet, with a tick on a scale (likelihood and expected value judgements) or in a box (decision-making). When all scenarios had been read, children received a sticker as a reward.

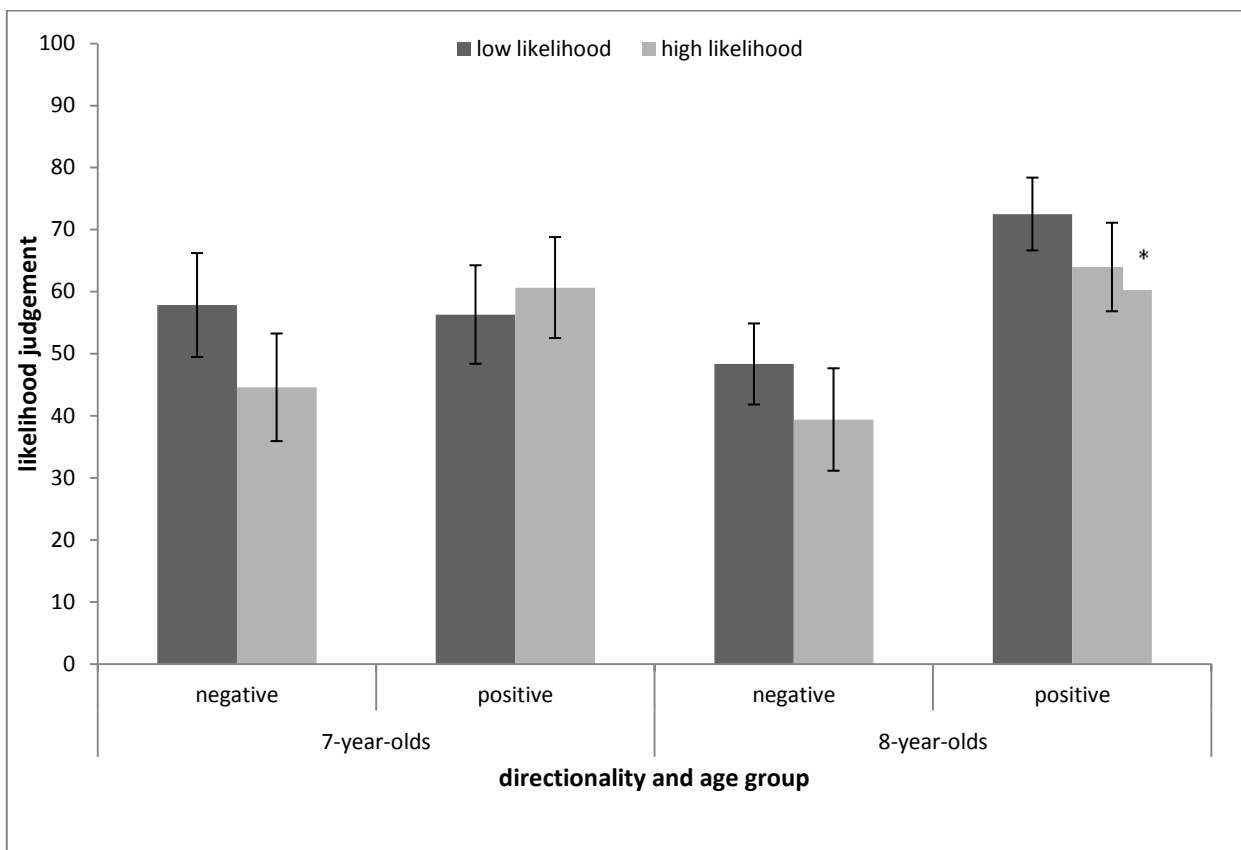
2.2.2. Results

2.2.2.1. Likelihood judgement

Mean likelihood judgements are presented in Figure 2. A 2x2x2 mixed analysis of variance (ANOVA) was conducted with directionality (positive or negative) and likelihood (high or low) as

within-subject factors and age (7 or 8 years old) as a between-subject factor. A main effect of directionality was found: positive expressions were judged as significantly higher ($M = 63.35$, $SD = 29.12$) than negative ones ($M = 47.53$, $SD = 29.43$), $F(1,45) = 6.87$, $p = .012$, $\eta^2_p = .13$.

No main effect of likelihood was found, $F(1,45) = 1.79$, $p = .188$. There was no effect of age either, $F(1,45) < 1$, $p = .843$. No reliable interaction was observed between directionality and age



^a error bars represent standard errors of the mean

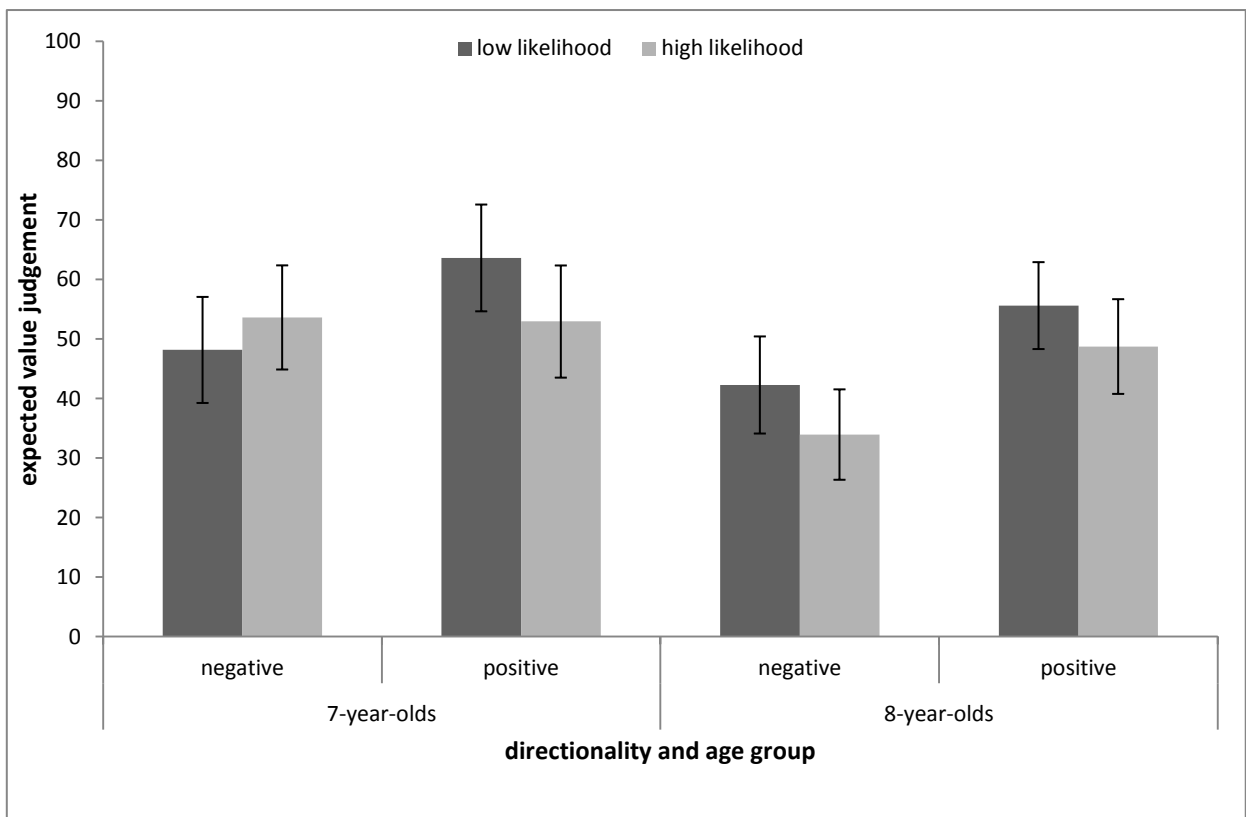
- denotes a significant difference from chance level (50; one-sample t -tests with Bonferroni correction, $\alpha = 0.017$)

Figure 2: Mean Likelihood Judgement as a Function of Age, Directionality and Likelihood (Experiment 1a)

or between directionality and likelihood, $F(1,45) = 2.00, p = .164$ and $F(1,45) = 1.03, p = .315$ respectively. Finally there was no interaction either between directionality, likelihood and age, $F(1,45) < 1, p = .338$.

2.2.2.2. Expected value judgement

Mean expected value judgements are presented in Figure 3. The same 2x2x2 mixed ANOVA as for likelihood judgements was applied. There was no effect of directionality, but a



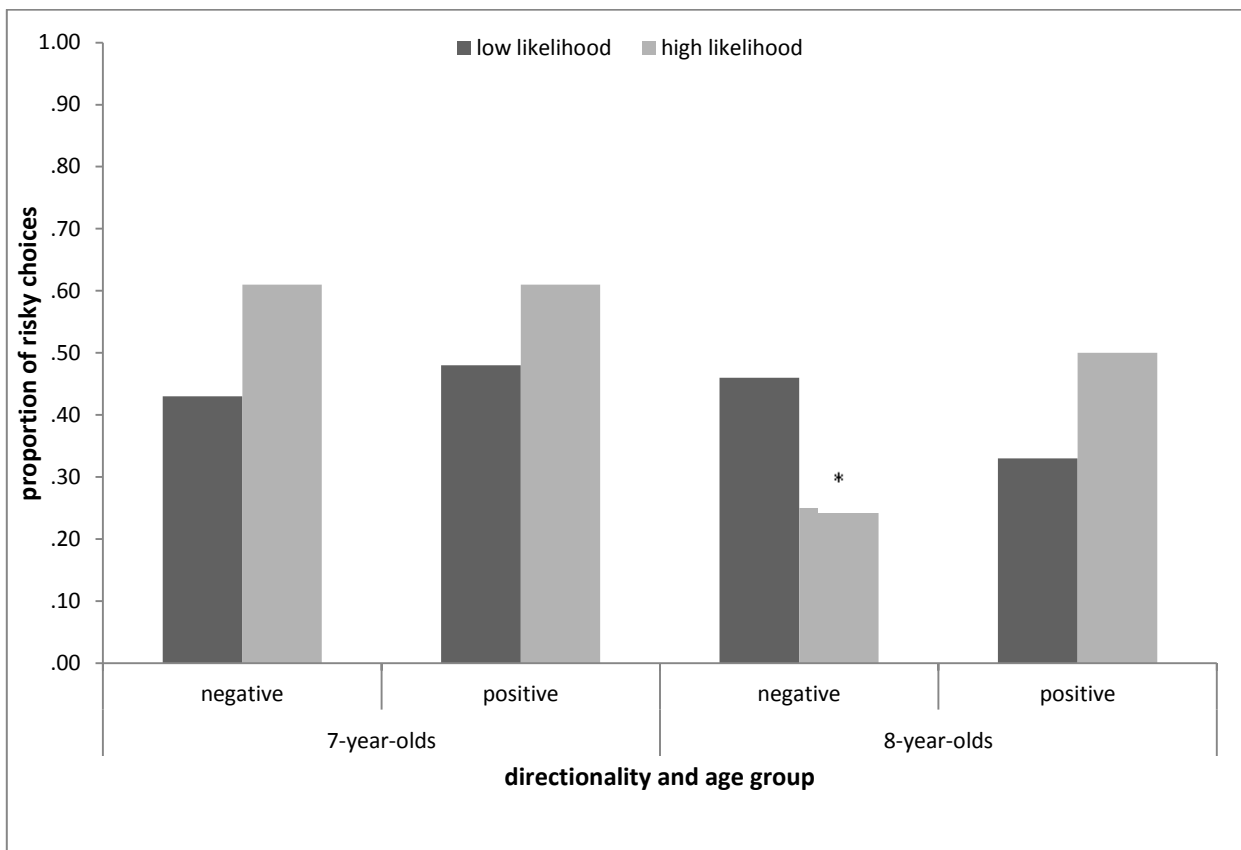
^a error bars represent standard errors of the mean

- denotes a significant difference from chance level (50; one-sample t -tests with Bonferroni correction, $\alpha = 0.017$)

Figure 3: Mean Expected Value Judgement as a Function of Age, Directionality and Likelihood (Experiment 1a)

trend, $F(1,44) = 3.17, p = .082, \eta^2_p = .07$: positive expressions elicited higher expected value judgements ($M = 56.06, SD = 29.12$) than negative ones ($M = 42.11, SD = 29.43$). There was no effect of likelihood, $F(1,44) = 1.03, p = .317$, or of age, $F(1,44) = 2.41, p = .128$.

No reliable interaction was observed between directionality and age, between directionality and likelihood, and between age and likelihood, $F(1,44) < 1, p = .582, F(1,44) < 1, p = .571$ and $F(1,44) < 1, p = .623$ respectively. Finally there was no interaction between



- denotes a significant difference from chance level (0.5; one-sample t -tests with Bonferroni correction, $\alpha = 0.017$)

Figure 4: Proportion of Risky Choices as a Function of Age, Directionality and Likelihood (Experiment 1a)

directionality, likelihood and age, $F(1,44) < 1, p = .498$.

2.2.2.3. Decision-making

Proportions of risky choices (i.e. choosing the uncertain proposition vs. the certain one) are presented in Figure 4. A 2x2x2 mixed General Estimating Equation for probit regression was conducted with directionality (positive or negative) and likelihood (high or low) as within-factors and age as between-subjects factor (7 or 8 years old).

I found no main effect of directionality, likelihood or age, *Generalized Score* $\chi^2(1) < 1, p = .416$, *Generalized Score* $\chi^2(1) < 1, p = .397$ and *Generalized Score* $\chi^2(1) < 1, p = .449$ respectively. There was no interaction between directionality and likelihood, *Generalized Score* $\chi^2(1) < 1, p = .400$, between directionality and age, *Generalized Score* $\chi^2(1) = 1.34, p = .247$, or between likelihood and age, *Generalized Score* $\chi^2(1) < 1, p = .632$. Finally, there was no interaction either between directionality, likelihood and age, *Generalized Score* $\chi^2(1) < 1, p = .422$.

2.2.3. Discussion

Children have been found to be sensitive to the likelihood of verbal probabilities from an early age if asked to make a decision (Moore & al., 1990). When asked to make only a likelihood judgement, sensitivity to likelihood was found at 8 to 9 years old (Gourdon & Villejoubert, 2009), and developing from there (Mullet & Rivet, 1991; Watson & Moritz, 2003). In Experiment 1a, I set out to replicate Gourdon & Villejoubert's findings in an English-speaking sample, but critically, by counterbalancing the task order. I found that my English sample of 7- and 8-year-olds was not influenced by the likelihood when making decisions, which contrasts with the

English-speaking sample in Moore et al. (1990) and the French-speaking sample in Gourdon and Villejoubert (2009). This might be explained by the fact that Moore et al. used only simple verbal probabilities (with only a probability word and no modifier), while I used more complex expressions as can be encountered in everyday language. Perhaps more importantly, I improved on the methodology of Gourdon and Villejoubert by counterbalancing the trials and choosing the expressions on piloted data. This seems to have affected performance and resulted in decreased success on the decision-making measure. In Gourdon and Villejoubert, decision-making was influenced by both the likelihood and directionality, showing patterns of answers similar to adults'. In this experiment, decision-making was found not to be influenced by either directionality or likelihood.

The 7- and 8-year-olds in this experiment also showed no effect of the likelihood when judging probabilities, in contrast with Mullet and Rivet (1991) and Watson and Moritz (2003). However they showed an effect of directionality, judging likelihood as higher for positive verbal probabilities, and as lower for negative verbal probabilities, regardless of their likelihood. This suggests that, as advanced by Gourdon and Villejoubert (2009), the results in Mullet and Rivet and Watson and Moritz may have resulted from a confounded variable between the likelihood and directionality, which was not taken into account in the previous studies. Children can interpret verbal probabilities from 7 years old, but they do so relying on shallow dimensions such as directionality, rather than relying on the likelihood.

Like Gourdon and Villejoubert (2009) I found that children were sensitive to directionality. Indeed, I found no interaction with age and so conclude that this sensitivity emerges from at

least 7 years old. However, this sensitivity was not consistent across the three measures: likelihood, expected value and decision making, unlike in Gourdon and Villejoubert. Where they found an effect of directionality throughout all tasks (likelihood judgement, expected value judgement and decision making), I only found it in the likelihood judgement task, with a trend in the expected value task. This suggests that the order of the tasks, fixed in Gourdon and Villejoubert with likelihood judgement first, resulted in children being primed to rely on directionality and likelihood across the other measures. However, I not only controlled for the order of the tasks but also used a variety of scenarios to achieve a higher ecological validity. Thus each scenario had specific consequences, which might have introduced some variability in the utility of the uncertain choice. For example, the consequences of missing out on birthday cake are perhaps less severe than the consequences of being late at school (or the reverse may be true!). Therefore the utility of the risky option (risking having no cake left and risking being late at school) may have been different in themselves, cancelling out any effect of directionality and/or the likelihood.

The clearest finding from Experiment 1a was that 7- and 8-year-olds seem to be affected by directionality of verbal probabilities. I investigated this further in a follow up study by modifying my task to make it more in line with Gourdon and Villejoubert's original task (2009). Several improvements to the materials were also made: scenarios and questions were presented in the same way as in this first study but the verbal probabilities were highlighted by the use of a bold font and repeated in every question to avoid unnecessary memory load. Only one scenario was used, following Gourdon and Villejoubert, to avoid that different utilities affected results.

However, I maintained my strategy of counterbalancing the measures to make sure any effect on expected value judgement and decision making was not due to an order effect and/or training.

Along with my attempts to reconcile my findings with those of Gourdon and Villejoubert (2009), in Experiment 1b I also investigated the influence of social context on children's handling of verbal probability. Gourdon and Villejoubert argued that directionality should be considered as a pragmatic cue, indicating speakers' intentions, and suggested introducing social factors such as speakers' benevolence to test this possibility. Thus, in Experiment 1b, there were two speakers with different and clearly identified intentions.

Mascaro and Sperber (2009) found that children can use the intention of a speaker as early as 4 years old. When informed that the speaker was lying, children reliably chose the option opposite to the one suggested by the speaker. Therefore in Experiment 1b, children may be expected to use speakers' intentions (benevolence or malevolence) in their judgements and decision making. My assumption is that a positive directionality used by a benevolent speaker should lead to higher judgements and more risky choices, but, used by a malevolent speaker, it should lead to lower judgements and less risky choices. A negative directionality should lead to lower judgements and less risky choices if used by a benevolent speaker, but to higher judgements and more risky choices if used by a malevolent speaker.

2.3. Experiment 1b

2.3.1. Method

2.3.1.1. Participants

Twenty-four 7- to 8-year-olds (mean age= 8;4; range= 7;11-8;10; 1 7- to 8-years old, but of unknown exact age; 12 girls) and 24 8- to 9-year-olds (mean age= 8;10; range= 8;11-9;8; 16 girls) took part in the experiment. They were all enrolled in a faith primary school of a multiethnic area of Birmingham (UK), and of diverse socio-economic backgrounds. Participants received two stickers in exchange for their participation.

2.3.1.2. Materials

The verbal probabilities were the same as in Experiment 1a (see Appendix II, p. 210). I manipulated the likelihood and directionality in a 2x2 within-subjects design, such that phrases in each directionality condition had a similar likelihood (positive and negative low likelihood = 35-37%; positive and negative high likelihood = 53-55%). As in Experiment 1a, the training phase presented participants with two anchoring verbal probabilities (*It is absolutely sure* and *It is impossible*) and one of medium likelihood (*It is not certain*).

A single scenario was used, both in the training and the experimental phase. It involved a treasure hunt, introduced as followed: 'Three friends are playing a game with lots of treasure chests which could contain a treasure or a trap. Each friend has first to find a clue to help him or her decide if the chest should be opened.'

As in Experiment 1a, each verbal probability was judged three times, to assess the likelihood judgement, the expected value judgement and the decision making. However, in order to

reduce the demands on memory, the verbal probability was repeated in each question and written in bold, as in the following examples: ‘When the clue says **it is not certain that the treasure is in the chest**, how much do you think that this means that the treasure is in the chest?’ (likelihood judgement); ‘As **it is not certain that the treasure is in the chest**, how happy do you think Sophie will be if she chooses to open the chest?’ (expected value judgement); ‘As **it is not certain that the treasure is in the chest**, if you were Sophie, what would you choose to do?’ (decision making). The order of the questions was counterbalanced between children (and kept constant within pairs of training and experimental phases). I also counterbalanced the order of the verbal probabilities.

The four verbal probabilities were presented as uttered both by a benevolent speaker (Peter Pan) and by a malevolent speaker (Captain Hook). Half of the participants in each group age received the block with the benevolent speaker first, while the other half received the block with the malevolent speaker first. To make transparent the intentions of each speaker, the following information was given before the benevolent block: ‘We are going to play the game with some other children, but now, Peter Pan is giving them the clues about each chest. Peter Pan is trying his best to help these children.’ Before the malevolent block, participants were told ‘[N]ow, Captain Hook is giving them the clues about each chest. Captain Hook secretly wishes he could keep the treasure for himself.’

As in Experiment 1a, all the scenarios were presented in a single booklet, with one page for each scenario and its subsequent tasks. Each page also presented a picture linked to the

scenario. That is, there was a treasure chest in all cases (the scenarios all being about a treasure hunt).

2.3.1.3. Procedure

The procedure was identical to the procedure of Experiment 1a.

2.3.2. Results

2.3.2.1. Likelihood judgement

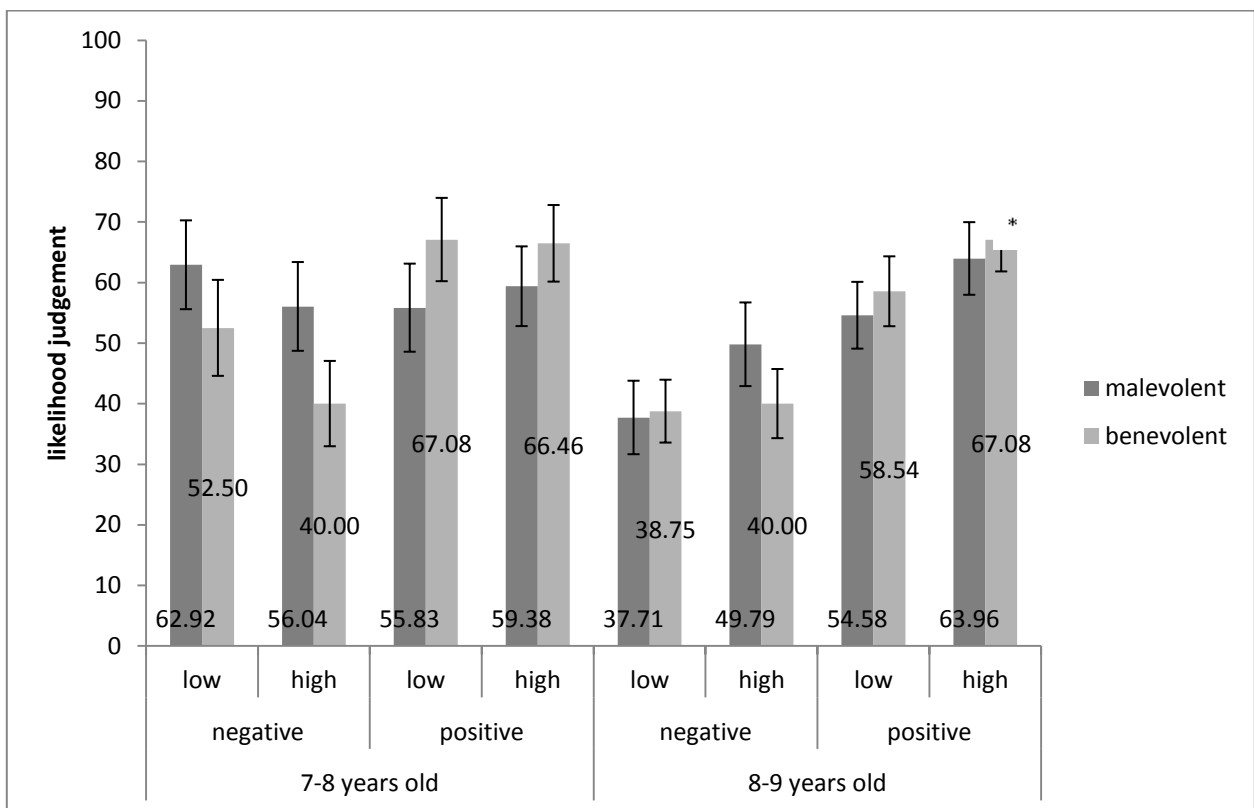
Mean likelihood judgements are presented in Figure 5. A 2x2x2x2 mixed analysis of variance (ANOVA) was conducted with speaker, directionality (positive or negative) and likelihood (high or low) as within-factors, and age as between-factor (7-8 years old or 8-9 years old).

There was a main effect of directionality, $F(1, 46) = 17.04, p < .001, \eta^2_p = .27$. Positive verbal probabilities were judged as meaning significantly higher ($M = 61.62, SD = 18.93$) than negative verbal probabilities ($M = 47.21, SD = 21.48$). There was no main effect of speaker, $F(1, 46) < 1, p = .614$, or of likelihood, $F(1, 46) < 1, p = .582$, or of age, $F(1, 46) = 1.76, p = .191$.

There was an interaction between directionality and speaker, $F(1, 46) = 6.21, p = .016, \eta^2_p = .12$. When the speaker was benevolent (Peter Pan), positive verbal probabilities were judged as significantly higher ($M = 64.79, SD = 23.53$) than negative verbal probabilities ($M = 42.81, SD = 24.92$), $t(47) = 4.53, p < .001, r = .41$. But when the speaker was malevolent (Captain Hook), positive verbal probabilities were judged similarly ($M = 58.44, SD = 21.52$) to negative verbal probabilities ($M = 51.62, SD = 27.46$), $t(47) = 1.52, p = .134$ (Bonferroni correction; $\alpha = .025$).

There was no difference between the benevolent and the malevolent speaker when verbal probabilities were positive, $t(47) = 1.75, p = .086$, or when they were negative, $t(47) = 2.14, p = .038$ (Bonferroni correction; $\alpha = .025$).

There was no interaction between directionality and age, $F(1, 46) = 2.12, p = .152$, between directionality and likelihood, $F(1, 46) = 2.00, p = .164$, between speaker and age, $F(1, 46) < 1$,



^a error bars represent standard errors of the mean

- denotes a significant difference from chance level (50; one-sample t -tests with Bonferroni correction, $\alpha = 0.008$)

Figure 5: Mean Likelihood Judgement as a Function of Age, Directionality, Likelihood and Speaker's Intention (Experiment 1b)

$p = .739$, between speaker and likelihood, $F(1, 46) < 1$, $p = .441$, or between likelihood and age, $F(1, 46) = 3.20$, $p = .080$. There was no interaction between directionality, likelihood and speaker, $F(1, 46) < 1$, $p = .575$, between directionality, speaker and age, $F(1, 46) = 1.42$, $p = .240$, between directionality, likelihood and age, $F(1, 46) < 1$, $p = .356$, and between speaker, likelihood and age, $F(1, 46) < 1$, $p = .946$. There was finally no interaction between speaker, directionality, likelihood and age, $F(1, 46) < 1$, $p = .676$.

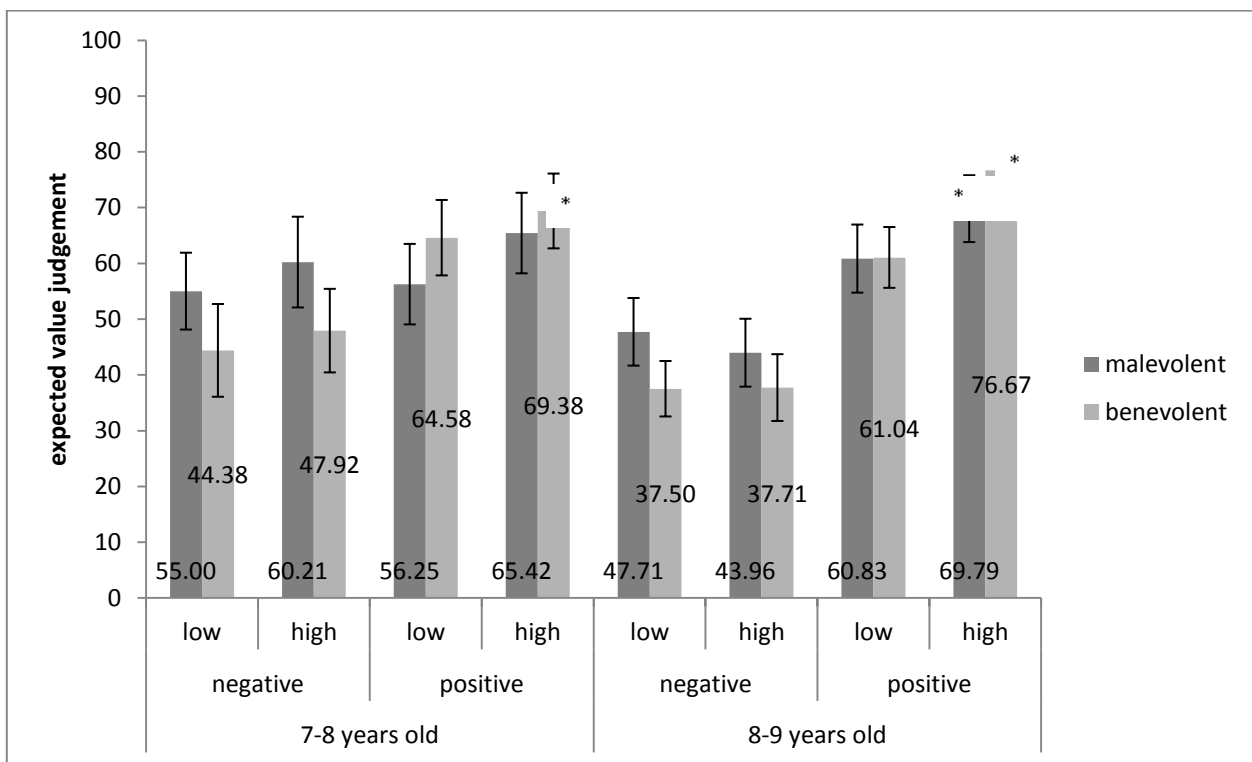
2.3.2.2. Expected value judgement

Mean expected value judgements are presented in Figure 6. The same 2x2x2x2 mixed ANOVA as for likelihood judgements was applied. There was a main effect of directionality, $F(1, 46) = 23.31$, $p < .001$, $\eta^2_p = .34$. Positive verbal probabilities were judged as significantly higher in likelihood ($M = 6.55$, $SD = 1.91$) than negative verbal probabilities ($M = 4.68$, $SD = 2.22$). There was no effect of speaker, $F(1, 46) < 1$, $p = .370$. There was no effect of the likelihood, $F(1, 46) = 3.28$, $p = .078$. There was no main effect of age, $F(1, 46) < 1$, $p = .448$.

There was an interaction between directionality and speaker, $F(1, 46) = 7.09$, $p = .011$, $\eta^2_p = .13$. When the speaker was benevolent (Peter Pan), positive verbal probabilities yielded higher expected values ($M = 6.79$, $SD = 2.27$) than negative verbal probabilities ($M = 4.19$, $SD = 2.43$), $t(47) = 5.86$, $p < .001$, $r = .48$. But when the speaker was malevolent (Captain Hook), positive verbal probabilities did not yield higher expected values ($M = 6.31$, $SD = 2.44$) to negative verbal probabilities ($M = 5.17$, $SD = 2.75$), $t(47) = 2.21$, $p = .032$ (Bonferroni correction; $\alpha = .025$). For negative verbal probabilities, Peter Pan yielded lower expected value judgements

than Captain Hook, $t(47) = 2.65, p = .011, r = .19$. This was not the case for positive verbal probabilities, $t(47) = 1.21, p = .234$ (Bonferroni correction; $\alpha = .025$).

There was no interaction between directionality and age, $F(1, 46) = 2.96, p = .092$, between directionality and likelihood, $F(1, 46) = 3.06, p = .087$, between speaker and age, $F(1, 46) < 1, p = .955$, between speaker and likelihood, $F(1, 46) < 1, p = .847$, or between likelihood and age,



^a error bars represent standard errors of the mean

- denotes a significant difference from chance level (50; one-sample t -tests with Bonferroni correction, $\alpha = 0.008$)

Figure 6: Mean Expected Value Judgement as a Function of Age, Directionality, Likelihood and Speaker's Intention (Experiment 1b)

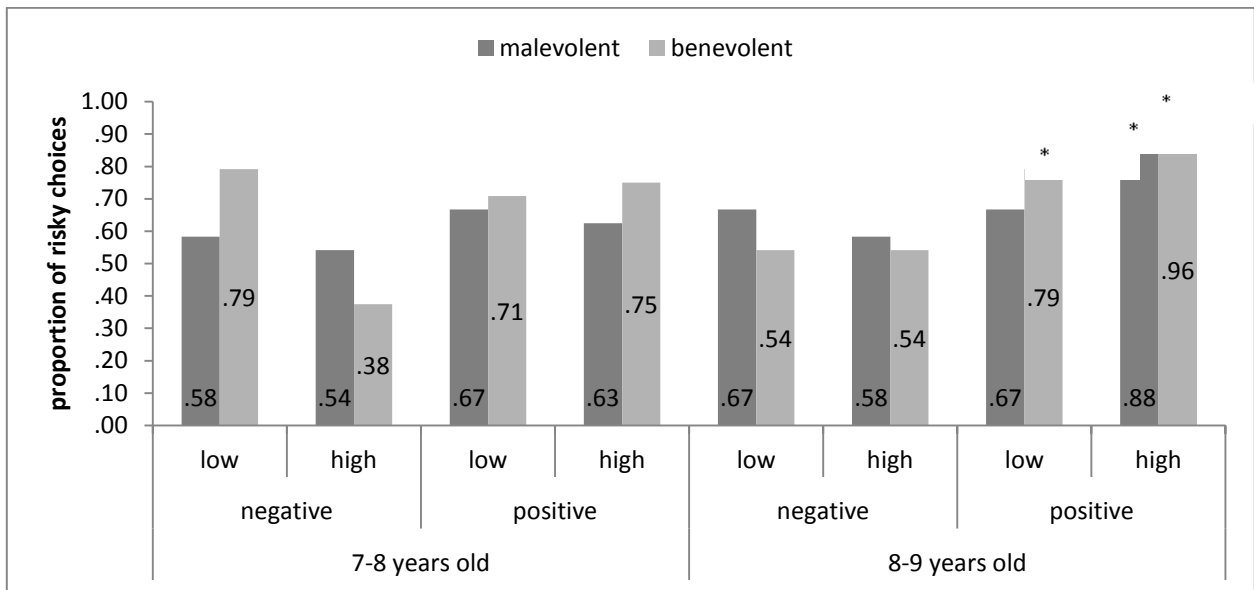
$F(1, 46) < 1, p = .945$. There was no interaction between directionality, likelihood and speaker, $F(1, 46) < 1, p > .999$, between directionality, speaker and age, $F(1, 46) < 1, p = .600$, between directionality, likelihood and age, $F(1, 46) = 1.44, p = .236$, and between speaker, likelihood and age, $F(1, 46) < 1, p = .484$. There was finally no interaction between speaker, directionality, likelihood and age, $F(1, 46) < 1, p = .841$.

2.3.2.3. Decision-making

Proportions of risky choices (i.e. choosing to open the chest) are presented in Figure 7. A 2x2x2x2 mixed General Estimating Equation for probit regression was conducted with directionality (positive or negative), likelihood (high or low) and speaker (benevolent or malevolent) as within-factors and age (7-8 years old or 8-9 years old) as between-factor.

There was a main effect of directionality, *Generalized Score* $\chi^2(1) = 14.16, p < .001$. Children were more likely to make a risky choice (i.e. open the chest) on hearing a positive verbal probability (76%) than a negative verbal probability (53%), *OR* = 0.39. There was also a main effect of age, *Generalized Score* $\chi^2(1) = 4.75, p = .029$. The 8- to 9-year-olds decided to take the risky option more often (70%) than the 7- to 8-year-olds (58%), *OR* = 0.44.

There was no effect of likelihood or of speaker, *Generalized Score* $\chi^2(1) = 1.04, p = .307$ and *Generalized Score* $\chi^2(1) < 1, p = .920$, respectively. There was no interaction between age and speaker, *Generalized Score* $\chi^2(1) = 1.07, p = .300$, between age and directionality *Generalized Score* $\chi^2(1) < 1, p = .424$, or between age and likelihood, *Generalized Score* $\chi^2(1) = 1.53, p = .217$.



- denotes a significant difference from chance level (0.50; one-sample *t*-tests with Bonferroni correction, $\alpha = 0.008$)

Figure 7: Proportion of Risky Choices as a Function of Age, Directionality, Likelihood and Speaker's Intention (Experiment 1b)

There was an interaction between speaker and directionality, *Generalized Score* $\chi^2(1) = 7.32$, $p = .007$. When the speaker was benevolent (Peter Pan), positive verbal probabilities yielded more often the risky option (80%) than did negative verbal probabilities (46%), $\chi^2(1) = 23.62$, $p < .001$, *Cramer's V* = .35. When the speaker was malevolent (Captain Hook), positive verbal probabilities did not yield more often the risky option (71%) than did negative verbal probabilities (59%), $\chi^2(1) = 2.77$, $p = .096$. When the verbal probability was negative, the malevolent speaker (Captain Hook) did not yield more risky choices (59%) than the benevolent speaker (Peter Pan; 46%), $\chi^2(1) = 3.27$, $p = .071$. There was no difference either in the

proportion of risky choices between the malevolent speaker (71%) and the benevolent speaker (80%), when the directionality was positive, $\chi^2(1) = 2.28, p = .131$.

There was no interaction between likelihood and speaker, *Generalized Score* $\chi^2(1) < 1, p = .448$, or between likelihood and directionality, *Generalized Score* $\chi^2(1) = 3.18, p = .075$. Finally, there was no interaction between age, speaker and directionality, *Generalized Score* $\chi^2(1) < 1, p = .991$, between age, speaker and likelihood, *Generalized Score* $\chi^2(1) < 1, p = .899$, between age, directionality and likelihood, *Generalized Score* $\chi^2(1) = 2.29, p = .130$, or between speaker, directionality and likelihood, and *Generalized Score* $\chi^2(1) < 1, p = .864$. There was no interaction between age, speaker, directionality and likelihood, *Generalized Score* $\chi^2(1) < 1, p = .828$.

2.3.3. Discussion

Following Gourdon and Villejoubert (2009), I tested the possibility that directionality of verbal probabilities had the function of pragmatic cue, indicating a speaker's intention. Mascaro and Sperber (2009) found that children were able to take speakers' intention into account. That is, when a malevolent speaker told them to open one box (out of two), children would chose to open the other box, as young as 4 years old. I therefore speculated that children's judgements and decision making would reflect both directionality and speakers' intention: a benevolent speaker would yield the usual higher judgements and more frequent risky choices observed with positive verbal probabilities; a malevolent speaker would yield the reversed pattern (higher judgements and more frequent risky choices observed with negative verbal probabilities).

Experiment 1b also attempted to account for differences between the results of Gourdon and Villejoubert (2009), and the results of Experiment 1a. In Gourdon and Villejoubert 8-year-olds judged likelihood based on directionality, but made decisions based on both directionality and the likelihood of verbal probabilities. I suggested that this could have been an artefact of the fixed order used in Gourdon and Villejoubert (always likelihood judgements first), as well as of the training phase provided on the items used in the experimental task. Experiment 1a therefore counterbalanced the order of the task and provided training with different items than the experimental ones. There I found an effect of directionality on the likelihood judgement, similar to the results of Gourdon and Villejoubert. However, directionality showed only a trend in the expected value judgement and had no effect on decision making. It was then proposed that using different scenarios might have introduced a confounding factor. Therefore Experiment 1b aimed also at replicating Experiment 1a but controlling for the scenarios.

The results of Experiment 1b can be summarized in two parts. First, there was a consistent effect of directionality. As in Experiment 1a when making likelihood judgements, positive verbal probabilities led to higher judgements, and negative ones to lower judgements. Consistent with the trend in Experiment 1a, directionality also had an effect on expected value judgements in Experiment 1b. Finally contrasting with Experiment 1a, the decision making task also showed an effect of directionality (showing a higher frequency of risky decisions with positive verbal probabilities than with negative ones). This effect of directionality is consistent with Gourdon and Villejoubert's results (2009), and supports my suggestion that Experiment 1a failed to replicate this effect because it did not use scenarios with similar utility. Most importantly, I

replicated this effect of directionality even when counterbalancing the different measures (i.e., unlike Gourdon and Villejoubert). The effect of directionality appears to be consistent across different measures of children's understanding of verbal probabilities.

Second, I also consistently found an interaction of directionality with speakers' intentions. When a benevolent speaker was the communicator, directionality yielded the same effect as in Experiment 1a where positive verbal probabilities led to higher judgements and proportions of risky choices than negative verbal probabilities. However, when the speaker was malevolent, directionality had a reduced effect in the three tasks. This is only partially in line with my expectations. In line with Mascaro and Sperber (2009), one might expect that speaker's malevolence should lead to a reversal of the judgements and decision making. If the speaker is trying to frame your decision in the wrong direction, s/he needs to know the actual answer. Therefore this speaker's statement can be used to infer the correct answer. However in this experiment, children acted as if the malevolent speaker was incompetent or unreliable, showing no preference for either directionality, as they would if choosing at random.

I also found a number of trends in Experiment 1b suggesting an emergence of understanding of the likelihood in verbal probabilities. Older children (8 to 9 years old) tended to be more correct in their likelihood judgements, reflecting the likelihood of the expressions. In judging expected values and making decisions, older children tended to integrate the likelihood with directionality; they also showed sensitivity to the likelihood alone when judging the expected value. This is consistent with Gourdon and Villejoubert's results (2009), and supports the possibility that using different scenarios in Experiment 1a may have introduced a confound

variable and may explain why I failed to replicate Gourdon and Villejoubert there. Finally the effect of directionality was moderated by age in the expected value task, with only the older children displaying it.

2.4. General discussion

Scarce empirical research in children's understanding of verbal probabilities has proved difficult to integrate as studies have used either decision tasks and simple expressions (Moore & al., 1990) or likelihood judgement tasks and more complex expressions (Mullet & Rivet, 1991; Watson & Moritz, 2003). More critically most of those previous studies had not taken into account directionality of verbal probabilities. Gourdon and Villejoubert (2009) did, but their paradigm did not test the verbal probabilities in a pilot, did not control for order effects and might have provided too much training. After Gourdon and Villejoubert, in Experiments 1a and 1b I used both likelihood judgement tasks and decision making tasks. As in their paper, I also systematically considered both directionality and the likelihood of verbal probabilities. In line with Gourdon and Villejoubert, I found that likelihood judgements and expected value judgements displayed only an effect of directionality. Interpreting directionality of verbal probabilities appears to be easier for children than interpreting likelihood, suggesting that its interpretation requires less cognitive resources. In contrast with Gourdon and Villejoubert however, I also found that children's decision making showed only an effect of directionality, not of likelihood.

That I found only an effect of directionality on likelihood judgements, expected value judgements and decision-making can be related to the counterbalancing of the task order in

both my experiments. It suggests that the fixed order of the tasks in Gourdon and Villejoubert could explain why they found an effect of likelihood on decision making. But it also suggests that the early sensitivity to directionality was not an order artefact. It is noteworthy that this effect was found in French-speaking children in Gourdon and Villejoubert and in English-speaking children in my experiments.

Secondly, I drew on Mascaro and Sperber's work (2009), suggesting that information delivered by malevolent speakers is considered cautiously and used by young children to draw inferences about the state of the world. In particular, I sought to bring support to my assumption that directionality has a pragmatic function, in that it communicates the intention of a speaker. I only found partial support for this assumption: children's use of directionality changed according to the intention of the speaker, but not in the strictly reversed manner I expected. Instead, children seemed to neglect directionality when the speaker was malevolent, contrasting with Mascaro and Sperber's results. However in Mascaro and Sperber the task was potentially less cognitively demanding as it was displayed through the use of puppets and actual boxes. In my tasks, characters and possibilities were imagined. The language was also more elaborate in my task, due to the use of verbal probabilities. This added a dimension of uncertainty that was not present in Mascaro and Sperber's task. It is therefore possible that the children only failed to use directionality when the speaker was malevolent because this would need more resources.

Further research should be conducted in an adult population with a similar paradigm. In a pragmatic account of verbal probabilities, directionality has been suggested to serve a function

of signalling to listeners about what the speaker knows or wished. In this interpretation, a malevolent speaker would instead use directionality to cue the listener towards the incorrect answer (that he wants you to pick). For this the speaker needs to know what the correct answer is. Therefore, one could consider the chosen directionality in light of their knowledge of the speaker's intention to find out the correct answer. But if directionality is not considered as a usual cue of the correct answer, it cannot be used either as a reversed cue of the correct answer. Then the only choice is to decide randomly. Therefore, if my assumption regarding the pragmatic function of directionality is correct, adults should demonstrate a full reversal of their use of directionality, with a preference for the negative directionality when the speaker is malevolent.

The two experiments I conducted here explored the understanding of verbal probabilities in 7- to 9-year-olds. I found some trends of use of the likelihood, as if children were developing their ability to use it. But it is clear that the age range of my studies is too limited to draw a developmental picture. Further research should aim to investigate more fully the development of the understanding of verbal probabilities, both at an earlier and at an older age. In the former case, it would aim to identify when children start using directionality, and in the latter one, to identify when the likelihood is reliably used in judgements and decision making.

I suggested that directionality is a shallow dimension of verbal probabilities, relying on fewer cognitive resources. Further research should specifically target the cognitive demands of the two dimensions of verbal probabilities, for example by measuring response time. If directionality was found to require fewer cognitive resources (e.g., as shown by shorter

response times to decide only on the basis of directionality), it could explain why children first base their judgements and decisions only on directionality (e.g., Gourdon and Villejoubert, 2009). Furthermore, less cognitive demands from directionality could also explain why adults show a consistent preference for positive verbal probabilities (e.g., Teigen and Brun, 1999). As many decisions are made under time pressure, or at least under the impression of time pressure, adults might chose the option that requires the less cognitive resources in order to decide faster. This would also extend Windschitl and Wells' (1996) work, who showed that verbal probabilities elicit Type 1 reasoning (fast, heuristic-based, cognitively less demanding), while numerical probabilities elicit Type 2 reasoning (slow, rule-based, demanding on cognitive resources). In fact, it could be that verbal probabilities elicit Type 1 reasoning only when cognitive resources are scarce (e.g., under time pressure).

Those results have implications for mathematics education. For example in the current British curriculum, teenagers are introduced to a few verbal probabilities and instructed what is each expression's likelihood equivalent (through a life example, e.g., a 50% chance is the chance to get a head if one flips a coin). However the curriculum uses negative verbal probabilities for low likelihood and positive ones for high likelihood (*BBC*, 2012; see Appendix IV, p. 214). Given that children already show a preference for the positive directionality, such a choice of verbal probabilities entails the risk of reinforcing the association between positive/negative directionality and high/low likelihood (respectively). This may further reinforce the preference displayed both by children and by adults for positive directionality, exposing them more to framing.

In this chapter I showed that 7- to 9-year-old children consistently display a use of directionality in verbal probabilities, but did not consistently use the likelihood. I accounted for these results in terms of cognitive demands, assuming that directionality imposes less of such demands. Further I showed that directionality is used by children only when the speaker is deemed benevolent. This partially supported that directionality fulfils a pragmatic function in verbal probabilities, where it communicates speakers' intention.

CHAPTER 3: OVERCOMING THE FRAMING EFFECT WHEN MAKING DECISIONS BASED ON VERBAL PROBABILITIES: HAVING MORE TIME IS HELPFUL BUT NOT ENOUGH.

3.1. Introduction

Everyone who has taken a plane knows what to do if there is a water landing. They also know that this event has a low chance of happening. The safety announcements regarding this special case are very often introduced as follows: 'In the unlikely event of (...)'. This choice of phrasing is surprising, though, if you consider that *unlikely* is generally considered to mean between a 10% and 30% chance of occurrence (e.g., Budescu, Karelitz, & Wallsten, 2003). One may wonder why air companies choose to mislead passengers by overestimating the actual risk. In fact, using negative probability words such as *unlikely* allows speakers to do two things: first they communicate about the chance, in this case very low, that the plane lands on water; second, and maybe more importantly, they drive people's attention to the non-occurrence of this event (Teigen, 1988; Teigen & Brun, 1995). Air companies can therefore hope not to make the possibility of a water landing too salient. This attention-driving property is referred to as directionality. Directionality, i.e. the positive or negative quality of a probability word or a quantifier, leads to framing effects in judgement and decision making. For example *A few people survived* (positive quantifier) is judged as better than *Few people survived* (negative quantifier; Sanford, Fay, Stewart, & Moxey, 2002). In this paper I investigated this framing effect in the case of verbal probabilities (e.g., *There is a chance, It is unlikely*): specifically I set out to identify the

relative processing costs of the different dimensions of verbal probabilities and to explore the conditions that can help to reduce the framing effect resulting from directionality.

Verbal probabilities are typically composed of a modal adjective (e.g., *likely*, *uncertain*) or a probability noun (e.g., *chances*), to which a modifier (e.g., *quite*, *a few*) can be added. Therefore speakers can choose among a large number of combinations (e.g., *There are a few chances*, *It is slightly unlikely*) when they need to communicate uncertainty. To investigate the quantitative meaning carried by verbal probabilities, Lichtenstein and Newman (1967) asked participants to translate 41 expressions into probabilities (from 0 to 1). Mean translations ranged from .06 to .89, covering most of the probabilistic range. It is noteworthy that for some expressions the spread of individual participants' responses also covered the range of probabilities: for example participants estimated *possible* to convey a likelihood between .01 and .99. After Lichtenstein and Newman's paper, studies on verbal probabilities focused primarily on the likelihood people attribute to such expressions (e.g., Reyna, 1981; Dhimi & Wallsten, 2005). A pattern emerged from those translation studies. On the one hand, verbal probabilities are translated in a stable way across studies. *Unlikely*, for example, has a mean translation between 10 and 20%. On the other hand, every study found interindividual variability. That is, verbal probabilities are characterised by such vagueness that when speakers use them to communicate risk, they cannot be sure that listeners understand it in the same way as themselves.

Teigen (1988) suggested that verbal probabilities express more than just a simple likelihood. In particular, he found that different directionalities led to different judgements of 'wrongness': if it turned out that the uncertain event occurred, a statement using a verbal probability with

negative directionality (i.e., a negative verbal probability) was deemed less appropriate than one using a verbal probability with positive directionality (i.e., a positive verbal probability); conversely, if the uncertain event did not occur, a statement using a positive verbal probability was judged less appropriate than one using a negative verbal probability. Directionality focuses a listener's attention on the occurrence of the event in question (when the verbal probability is positive, e.g., *It is likely that I will come tonight*), or on its non-occurrence (when it is negative, e.g., *It is not entirely certain that I will come tonight*). This orientation of attentional focus was observed in a continuation task (Teigen & Brun, 1995). Participants completed sentences about uncertain events described by a verbal probability. When the verbal probability was positive (e.g., *It is likely that Adam will pass his exam*), participants tended to complete the sentence with a reason supporting the occurrence of the uncertain event (i.e. a pro reason; e.g. *because he worked hard for it*). When the verbal probability was negative (e.g., *It is not certain that Adam will pass his exam*), they chose more often a reason in line with the non-occurrence of the event (i.e. a con reason; e.g. *because he skipped a lot of classes*). Directionality has also been shown to frame decision-making. Teigen and Brun (1999) observed that intentions to use a drug were higher if its chances of efficiency were given by a positive verbal probability than if they were given by a negative one *despite another group of participants translating the two verbal probabilities as having similar likelihood*.

According to Budescu et al. (2003), people prefer options given by positive verbal probabilities because they are interpreted as conveying high likelihoods whereas negative ones convey low likelihoods (indifferently for positive and negative events). Thus, they claimed that

the preference for positive verbal probabilities merely reflects a preference for higher probabilities. However, Teigen and Brun (1999) suggested that through directionality, verbal probabilities deliver 'a consistent message, with clear implications for inferential judgments' (p. 185). The implication in the example of the drug choice is that the speaker wants the listener to choose the one described by a positive verbal probability and that is why s/he drives the listener's attention to the drug's efficiency. Therefore the incongruence between judgements of likelihood and the influence of directionality on decision making can be resolved if listeners infer speakers' intention, i.e. I implicate pragmatics: the aspects of meaning that depend on the speaker, the addressee and/or the context and could not be reached based on semantics only. There are two central principles to pragmatics (see e.g., Wilson & Sperber, 2004). First, by the act of communicating a speaker implies to the listener that what is said is relevant (and therefore worth processing). In the case of verbal probabilities, including information about directionality should be seen as a relevant clue to build inferences on. Second, the listener should '[f]ollow a path of least effort in constructing an interpretation of the utterance (and in particular in resolving ambiguities...)' (Sperber & Noveck, 2004, p.6-7). Therefore the framing effect of directionality may result from the related performance costs. It might be that directionality influences our decisions because it requires no more, and perhaps fewer, cognitive resources than interpreting the likelihood of verbal probabilities.

Studies investigating how children judge verbal probabilities support the idea that directionality is less costly to process than the likelihood. Gourdon and Villejoubert (2009) and myself in Chapter 2 observed that when 8-year-olds judged the likelihood of an outcome

described by a verbal probability they used only directionality, ignoring the likelihood: children judged positive verbal probabilities as expressing a higher chance than negative ones, regardless of the likelihood these expressions represented. Processing directionality appears to be easier than interpreting likelihood for children, and a similar pattern may persist in adulthood.

In the two studies I report here, participants made a choice between two verbal probabilities. In previous studies (e.g. Teigen & Brun, 1999), participants were presented with either a positive or a negative verbal probability, in a between-subjects design, and asked if they would recommend that the character in a medical scenario takes a drug. There are plenty of daily situations when a single uncertain event is described to you by a verbal probability, based on which you then decide to take the chance or not. But there are also many situations when two uncertain possibilities are offered, both described by a verbal probability. For example, one could be presented with two different choices of medical treatment, and informed of each treatment's likelihood of success. Giving participants two verbal probabilities might reduce framing effects. The comparison may allow listeners to identify when two verbal probabilities actually have the same likelihood. They may also identify when a negative verbal probability actually expresses a higher likelihood than a positive one, resulting in less frequent framing effects due to directionality.

Furthermore, giving participants two verbal probabilities in my studies meant I could create conditions where decisions needed to be based only on directionality or only on the likelihood, to identify their respective processing costs (through measure of response time). In the task participants saw two treasure chests, each of which was described as perhaps containing gold

coins, using a verbal probability. Verbal probabilities could differ in directionality (positive/negative) and likelihood (high/low). Participants had to decide on a single chest to open. When the likelihood of the two descriptions was similar and directionality was the only thing that differed, I expected participants to choose the positive verbal probability more often, as people prefer options given with positive verbal probabilities (Teigen & Brun, 1999).

Moreover, if directionality needs fewer cognitive resources to be processed than likelihood, I speculated that participants may rely on it inappropriately. Thus, they may choose the positive verbal probability even if directionality contradicts the likelihood (e.g., positive verbal probability conveying a low likelihood). People prefer to use positive verbal probabilities for high likelihoods and negative ones for low likelihoods (Budescu et al., 2003). Thus, they may also choose the positive verbal probability over the negative to a greater extent when both have a high likelihood. Regarding response time, I expected that making decisions between two uncertain events predicted by verbal probabilities might be quicker when directionality is the only thing to rely on (i.e. when the likelihood is equivalent), than when the likelihood is the only thing to base their decisions on (i.e. when directionality is the same).

3.2. Experiment 2a

3.2.1. Method

3.2.1.1. Participants

Nineteen Psychology students from a UK high school (mean age = 17 years and 4 months (17;4); range = 17;0 - 17;10; 15 females) took part in the experiment as part of their psychology curriculum.

3.2.1.2. Materials

Four categories of verbal probabilities combining directionality and likelihood were used: high likelihood conveyed with a positive directionality; low likelihood conveyed with a positive directionality; high likelihood conveyed with a negative directionality; low likelihood conveyed with a negative directionality (see the 12 expressions in Table 2).

Directionality of each verbal probability was established as either positive or negative primarily based on previous research by Teigen and Brun (1995), Budescu et al. (2003), and Honda and Yamagishi (2009). Directionality of nine verbal probabilities that were not used in previous studies was determined by directionality of their modal word (e.g., *unlikely*), combined with the presence or absence of a negative modifier(s), or of a negation. For example, *few chances* was categorized as negative based on the negative directionality of *few* (Sanford et al., 2002), which modified the directionality of *chance* (positive; Teigen & Brun, 2003). Or the directionality of *not absolutely certain* was determined as negative, because of the addition of the negation *not* to an otherwise positive phrase (Teigen & Brun, 1995). The likelihood of the verbal probabilities was categorized either as high or low also based on data collected as a pilot

for Experiment 1a. High verbal probabilities all translated above 50% and low ones all translated below 25%. Of course, as one would expect, when directionality and likelihood were both considered, these combined to affect likelihood judgements (e.g., positive high verbal probabilities were judged as expressing a higher likelihood than negative high verbal probabilities). This is consistent with research by Teigen and Brun (1999), who found verbal probabilities were judged as expressing equivalent likelihood in a pilot study, but used differently in the context of giving advice. It is this effect that I sought to disentangle here by looking at the processing costs.

Pairing each of the four categories with each other I constituted six comparison conditions. In the positive condition (positive expressions with high likelihood vs. positive expressions with low likelihood) and the negative one (negative expressions with high likelihood vs. negative expressions with low likelihood), directionality was controlled and a decision could be made only according to the likelihood. In the high condition (positive expressions with high likelihood vs. negative expressions with high likelihood) and the low one (positive expressions with low likelihood vs. negative expressions with low likelihood), the likelihood was controlled and a decision could be made only according to directionality. In the congruent and the incongruent condition, both dimensions were different, reinforcing each other (congruent condition: positive expressions with high likelihood vs. negative expressions with low likelihood) or contradicting each other (incongruent condition: negative expressions with high likelihood vs. positive expressions with low likelihood).

Nine pairs were created within each comparison condition. Pairs were presented with pictures of two identical treasure chests. The outcome the verbal probabilities referred to was kept constant (i.e. *there are some coins in the chest*).

3.2.1.3. Procedure

Each participant took part individually in a quiet room. The task was implemented on a computer via E-Prime and ran across two blocks of trials (A and B). In Block A, all the possible pairs of verbal probabilities were presented in one order (e.g., *almost certain* on the left and *not guaranteed* on the right). In Block B, the order within the pairs was reversed to avoid any unexpected effect of the side of presentation (e.g., *almost certain* on the right and *not guaranteed* on the left). Half the participants started with Block A followed by Block B, the other half starting with Block B followed by Block A. Furthermore the two blocks were presented a second time, in the same order, so that participants took part in four blocks of trials (A-B-A-B or B-A-B-A).

The following instructions were given to the participants:

‘You are going to be presented pairs of treasure chests. In each pair only one chest contains some gold coins. Plus opening the empty one would make you lose the entire game. To help you deciding which chest you should open, some hints will be given.

Sometimes the two hints will seem very similar to you but they are actually always different within each pair. For each pair you will have to say if you want to open the left chest or the right one. You will have 5 seconds by pair to make this decision.’

Each trial was introduced by a fixation cross and lasted for a maximum of 5 seconds. Although the participants were instructed that losing a single trial would make them lose the entire game, they did not receive feedback after each trial. This would have created a learning effect which was not the focus of this experiment, and would also have entailed the risk of demotivating participants before the end of the experiment. A training block was first administered to familiarize participants with the time limit: eight pairs of verbal probabilities were presented to participants twice each, once in one order and once in the reversed order; those pairs were different from the experimental ones (see Table 2).

Table 2
Verbal Probabilities Used in Experiments 2a, 2b, 3a and 3b

		low likelihood	high likelihood
experimental items	positive verbal probabilities	<i>There is a small possibility</i>	<i>It is very possible</i>
		<i>There are a few chances</i>	<i>It is almost certain</i>
		<i>There is a very poor chance</i>	<i>It is very likely</i>
	negative verbal probabilities	<i>It is almost impossible</i>	<i>It is not absolutely certain</i>
		<i>There are few chances</i>	<i>It is not guaranteed</i>
		<i>It is very unlikely</i>	<i>It is a little unlikely</i>
training items	positive verbal probabilities		<i>It is not doubtful</i>
			<i>It is quite likely</i>
		<i>There is a poor chance</i>	<i>It is probable</i>
	negative verbal probabilities		<i>It is rather certain</i>
			<i>It is a little doubtful</i>
		<i>It is quite unlikely</i>	<i>It is not definite</i>

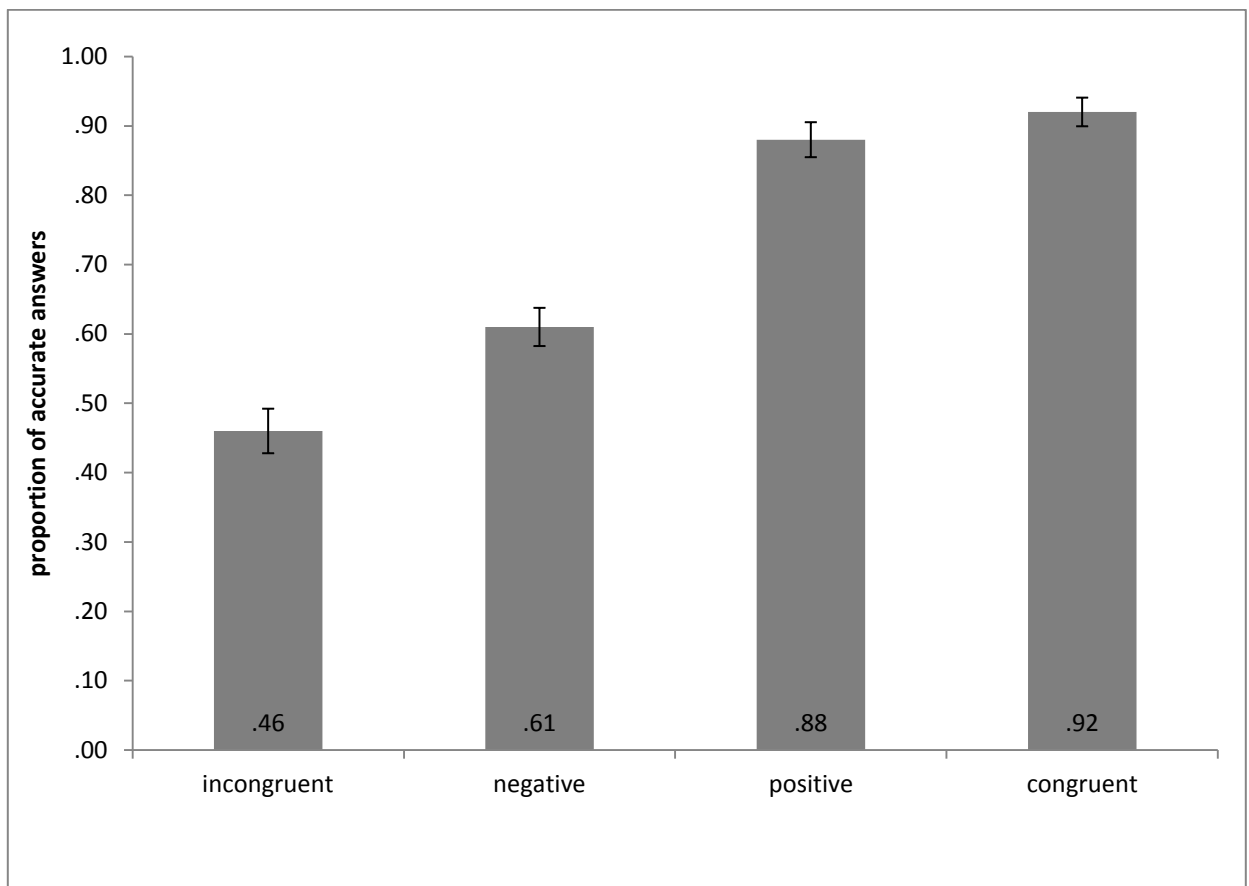
Following the training block, participants completed the four blocks of experimental trials. I recorded which box the participant chose to open on each trial. I also recorded the time elapsed between the presentation of the stimuli and the response.

3.2.2. Results

Each individual verbal probability was included in nine different pairs, and each pair of phrases was compared four times in each experiment (in each of two blocks which were presented twice). Thus participants were presented with each verbal probability 36 times. The phrases I used differed in how familiar they might be to my participants and also in their absolute length. I reasoned that if either of these features were driving any experimental effects they should reduce over time as all the items became more familiar within the context of the experiment. Therefore I conducted my analyses twice, the second time including the experimental order (first or second half of the trials, each half including Block A and Block B, therefore two presentations of each phrase) as additional factor.

3.2.2.1. Accuracy

Accuracy was defined as choosing the chest with the higher likelihood of containing gold coins. In order to test if the accuracy of choice was affected by directionality, I compared the performance between the incongruent and congruent conditions, as well as between the positive and negative conditions. For each trial, participants received a score of 0 (incorrect) or 1 (correct) for their choice. Those scores were averaged so that in each condition, each participant could have an accuracy score between 0 and 1. Mean proportions of accurate answers are shown in Figure 8.



^a error bars represent standard errors of the mean

Figure 8: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Time Condition (Experiment 2b)

ran a repeated-measures analysis of variance (ANOVA) with the comparison condition (incongruent, negative, positive or congruent) as within-participant factor. There was a main effect of comparison condition both for subjects, $F_1(3,54) = 148.70, p < .001, \eta^2_p = .89$ (lower bound adjustment), and for items, $F_2(3,31) = 9.46, p < .001, \eta^2_p = .48$. Accuracy was greatest in the congruent condition, followed by positive, negative and then incongruent. I investigated statistical differences between the conditions in post hoc tests.

Planned repeated contrasts were applied with a Bonferroni correction (for 3 tests, $\alpha = .017$), comparing each pair of means that were closest in accuracy. Accuracy in the incongruent condition ($M = 0.46$, $SD = 0.14$) was significantly lower than in the negative condition ($M = 0.61$, $SD = 0.12$), $F_1(1,18) = 29.82$, $p < .001$, $\eta^2_p = .62$. Accuracy in the negative condition was significantly lower than in the positive condition ($M = 0.88$, $SD = 0.11$), $F_1(1,18) = 138.36$, $p < .001$, $\eta^2_p = .89$. Accuracy in the positive condition was significantly lower than in the congruent condition ($M = 0.92$, $SD = 0.09$), $F_1(1,18) = 7.16$, $p = .015$, $\eta^2_p = .29$.

A second analysis was conducted taking into account the experimental order. The main effect of comparison condition remained for subjects and for items (lowest $F = 9.46$, highest $p < .001$). There was no effect of the experimental order, for subjects or items, and there was no interaction between the comparison condition and the experimental order, for subjects or for items (highest $F = 1.70$, lowest $p = .178$). There was no suggestion that increasing familiarity with items over the course of the experiment affected the results.

3.2.2.2. Directionality

To check if participants preferred the option described by the positive verbal probability when there was no correct answer (low and high conditions), I compared the choices made under those conditions to chance level. For each trial participants received a score of 0 (negative expression) or 1 (positive expression) for their choices. Those scores were averaged so that for each condition where directionality was a relevant measure, each participant had a score between 0 and 1, representing the proportion of trials where they chose the positive expression. Therefore chance level was set as 0.5. One-sample t -tests indicated that participants

chose the outcome described by a positive verbal probability significantly more often than chance: in the low condition ($M = 0.64$, $SD = 0.12$), $t(18) = 4.81$, $p < .001$, $r = .50$; in the high condition ($M = 0.91$, $SD = 0.12$), $t(18) = 14.81$, $p < .001$, $r = .86$. However one-sample t -tests run for items found this preference for the outcome described by a positive verbal probability only in the high condition, $t(10) = 50.16$, $p < .001$, $r < .99$. In the low condition, there was no such difference, $t(5) = 1.10$, $p = .320$.

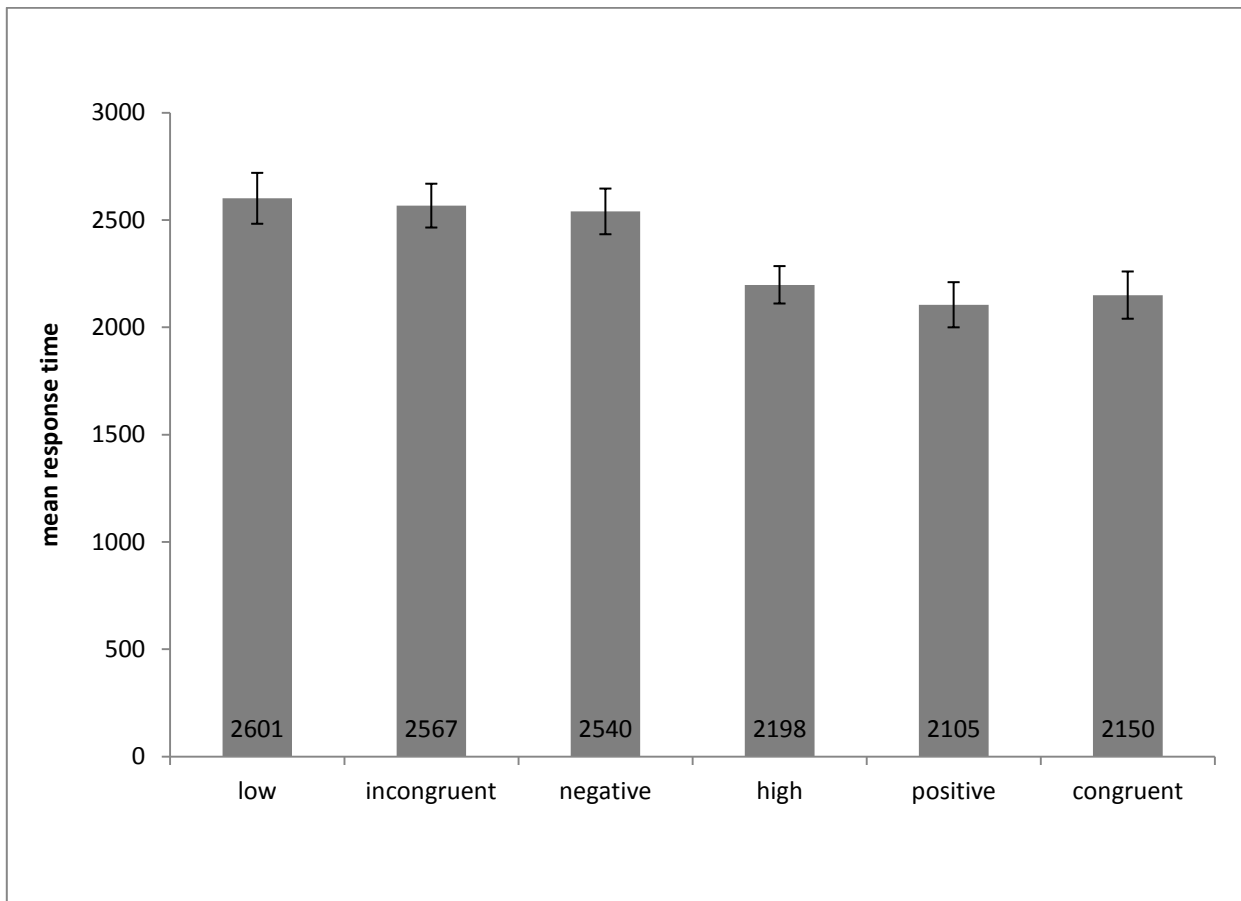
I also ran a paired samples t -test which indicated that when there was no correct answer, the extent of the preference for the positive probability differed according to the comparison condition: participants chose the outcome described by a positive verbal probability significantly more often under the high condition than under the low condition, $t(18) = 8.343$, $p < .001$, $r = .75$. The equivalent analysis by items found, however, no difference between the low and the high condition in the preference for the outcome described by a positive verbal probability, $t(5.041) = 2.12$, $p = .087$ (equal variances not assumed).

I also ran a 2x2 ANOVA analysis to check there was no effect of the order in which the items were presented, with comparison condition and experimental order as factors. I found a main effect of comparison condition both for subjects, $F_1(1,18) = 66.79$, $p < .001$, $\eta^2_p = .79$, and for items, $F_2(1,15) = 8.61$, $p = .010$, $\eta^2_p = .37$. There was no effect of experimental order for subjects or for items, and there was no interaction between comparison condition and experimental order for subjects or for items (all $F < 1$, lowest $p = .782$).

3.2.2.3. Response time

I compared response times under each condition to check if participants were quicker when it was possible to answer using only directionality (low, high, congruent and incongruent conditions) than when there was no difference in directionality and a choice could be based only the likelihood had to be considered (positive and negative conditions). The medians of each participant's response time in each condition were used instead of their means, in order to deal with outliers without losing data points. This method does not affect the α level more than any other method used for cleaning outliers (Ratcliff, 1993). Means of median response times in each comparison condition are presented in Figure 9. I ran a repeated-measures analysis of variance (ANOVA) with the comparison condition (incongruent, low, negative, high, positive or congruent) as within-participant factor. An effect of comparison condition on the response time was confirmed, for both subjects, $F_1(5,90) = 16.11, p < .001, \eta^2_p = .47$, and items, $F_2(5,46) = 8.84, p < .001, \eta^2_p = .49$.

Planned repeated contrasts (order of conditions: incongruent, negative, low, positive, high, congruent) comparing adjacent response times (with a Bonferroni correction for 5 tests, $\alpha = .01$) found a marginally significant difference between the negative and high conditions, $F_1(1,18) = 6.51, p = .02, \eta^2_p = .27$. Therefore I ran post-hoc special contrasts to compare each of the incongruent, low and negative conditions to each of the high, positive and congruent conditions. Those contrasts indicated that under the low condition, the response time was conditions. Those contrasts indicated that under the low condition, the response time was



^a error bars represent standard errors of the mean

Figure 9: Mean Response Time in the Comparison Conditions (Experiment 2a)

significantly longer than the high, the positive and the congruent conditions (see Table 3 for F and p values). Under the incongruent and the negative conditions, the response time was significantly longer than the positive and the congruent conditions, but not than the high condition.

The same analysis was conducted taking into account the experimental order. The main effect of comparison condition remained identical for subjects and for items (lowest $F = 8.84$,

highest $p < .001$). There was also a main effect of experimental order both for subjects, $F_1(1,18) = 42.14, p < .001, \eta^2_p = .70$, and for items, $F_2(1,46) = 114.88, p < .001, \eta^2_p = .71$. As may be expected, participants answered faster in the second half than in the first one. However, there was no interaction between comparison condition and experimental order, for subjects or for items (all $F < 1$, lowest $p = .584$).

Finally, I also ran additional analyses using only response times from successful trials (i.e. trials in which participants chose the chest with the highest likelihood of containing the coins) in the positive, negative, incongruent and congruent conditions. (There was not a correct answer in the high and low conditions.) I again found a similar main effect of the comparison condition, $F_1(3,54) = 12.6, p < .001, \eta^2_p = .41$.

Table 3

Post-Hoc Special Contrasts on the Response Time (Experiment 2a)

compared conditions	F values	p values	η^2_p
low vs. high	24.5	< .001 ^a	.58
low vs. positive	37.69	< .001 ^a	.68
low vs. congruent	59.02	< .001 ^a	.77
incongruent vs. high	9.82	.0057	.35
incongruent vs. positive	19.79	< .001 ^a	.52
incongruent vs. congruent	31.51	< .001 ^a	.64
negative vs. high	6.51	.02	.27
negative vs. positive	23.12	< .001 ^a	.56
negative vs. congruent	19.73	< .001 ^a	.52

^a significant after Bonferroni correction (for 9 tests, $\alpha = .0056$)

3.2.3. Discussion

As expected, when the verbal probabilities differed in directionality, participants were influenced by directionality. When verbal probabilities were incongruent, this led them to choose the chest with the lower likelihood (described by a positive directionality) to contain gold coins half of the time, whereas when the verbal probabilities were congruent, they chose the correct answer (the chest with the highest probability) most of the time. In the low and high conditions, in which there was no correct answer, relying on directionality led participants to choose the chest described by a positive verbal probability more often than chance (even if to a lesser extent in the low condition). On the other hand, response time showed a different pattern from my hypothesis. I expected participants to need more time in both the negative and positive conditions (as participants needed to use the likelihood to make a decision). In fact, the negative, incongruent and low conditions all needed more time than the high, positive and congruent conditions.

The unexpected longer response time in the low condition could result from the task constraints. According to the instructions there was always one chest containing the gold coins and therefore the two verbal probabilities should have added up to approximately 1. But in the low condition their sum was always much less than 1. Thus, the longer response time could reflect participants being sensitive to violations of probability laws. However, the same violation occurred in the high conditions (where the two verbal probabilities would sum to be greater than 1) and nevertheless participants did not take longer to respond. Sensitivity to violations of probability laws is unlikely to explain the longer response time in the low condition.

Instead, I suggest that the longer response time observed in the incongruent and low conditions indicates some sensitivity to inconsistency. In the incongruent condition, the likelihood was contradicted by directionality. In the low condition, the likelihood was inconsistent with the task goal (find some treasure). To detect those two types of inconsistency, participants needed to consider both the likelihood and directionality. Therefore, even when directionality differs and could be used as heuristic to make a decision, people take both dimensions into account. This suggests that the framing effect of directionality cannot be explained by people considering only directionality solely because of its ease of processing.

As in previous studies, participants showed a preference for positive phrases both in the low and the high conditions (although to a lesser extent in the low condition). However, unlike previous studies, my participants had two verbal probabilities to compare. The preference for positive verbal probabilities shown in previous studies (e.g., Teigen & Brun, 1999) therefore cannot be explained by the fact that participants heard only one verbal probability and could not realize its likelihood was similar to another one. Instead the framing effect seems to be a strong property of directionality.

In Experiment 2a, participants had a limited time (5 seconds) to make their decisions. In the second experiment I investigated whether giving people unlimited time to make their decisions affected the framing effect of directionality. I expected that giving more time to participants would allow them to overcome their preference for positive phrases. I also expected that in the incongruent condition, where the accuracy was poor, performance would improve.

3.3. Experiment 2b

3.3.1. Method

3.3.1.1. Participants

Twenty Psychology students (mean age= 19.20; range= 18-26; 19 women) at university of Birmingham (UK) took part in this experiment in exchange for course credit.

3.3.1.2. Materials

The materials were the same as in Experiment 2a.

3.3.1.3. Procedure

Each participant was tested individually. The task was the same as in Experiment 2a except that participants made decisions under two time conditions: in the limited-time condition, each trial lasted 5 seconds; in the unlimited-time condition, participants could take all the time they needed. As in Experiment 2a, every participant first completed a training block (see Table 2). The training block was completed under the time condition the participant would be presented with first (half the participants did the limited-time condition for the first two blocks and the other half did the unlimited-time one for the first two blocks).

In the limited-time condition, participants received the same instructions as in Experiment 2a. In the unlimited-time condition, instead of 'You will have 5 seconds by pair to make this decision.', the instructions read 'You can take all the time you need to make this decision.'

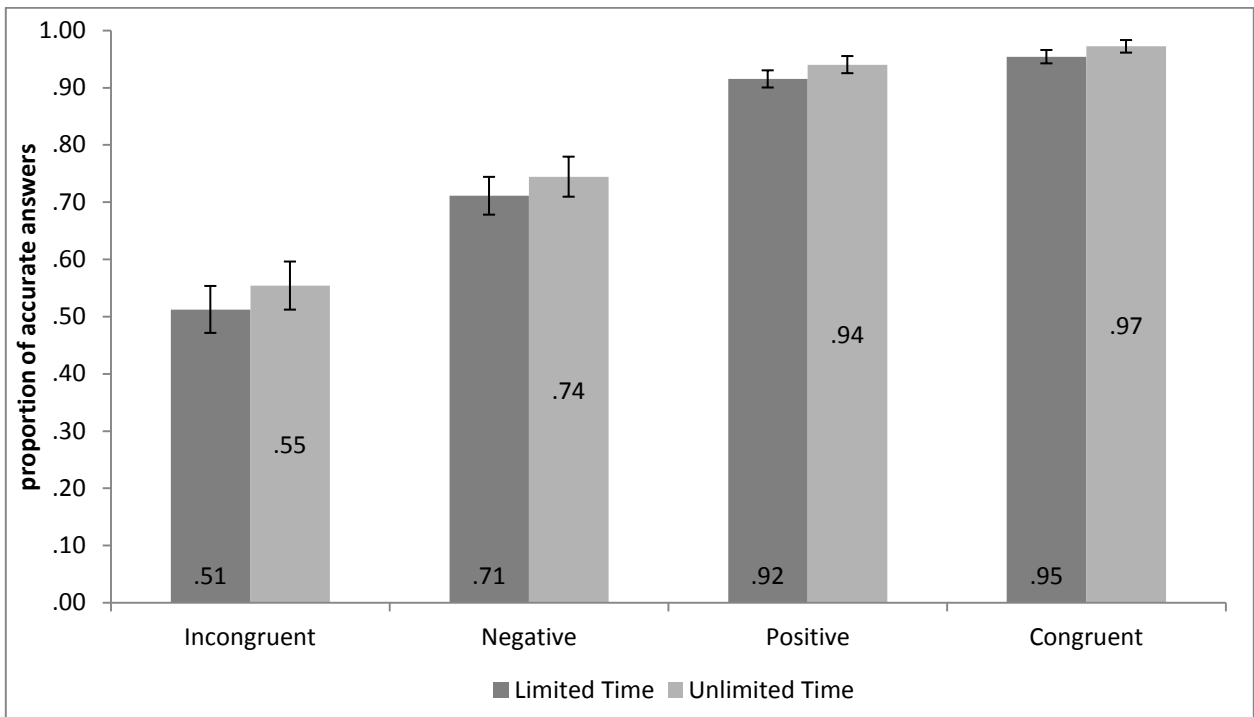
3.3.2. Results

As in Experiment 2a, I conducted secondary analyses, adding in an experimental order factor (first half of trials vs. second half of trials) to confirm whether familiarity or length of phrases in the pairs influenced performance. These are reported after the main analyses.

3.3.2.1. Accuracy

In order to test if participants were more accurate when they had more time to make a decision, I compared the performance on the four conditions where a correct answer existed (positive, negative, incongruent and congruent conditions), under the two time conditions. As in Experiment 2a, the accuracy score could range from 0 to 1. Mean proportions of accurate answers according to the comparison condition (when accuracy is a relevant measure) and the time condition are presented in Figure 10.

A 4 (comparison condition: incongruent, negative, positive or congruent) by 2 (time condition: limited or unlimited) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on accuracy both for subjects, $F_1(1,19) = 95.6, p < .001, \eta^2_p = .83$ (lower bound adjustment), and items, $F_2(3,32) = 11.74, p < .001, \eta^2_p = .52$. Planned



^a error bars represent standard errors of the mean

Figure 10: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Time Condition (Experiment 2b)

repeated contrasts indicated that accuracy in the incongruent condition was significantly lower than in the negative condition, $F_1(1,19) = 59.87, p < .001, \eta^2_p = .76$. Accuracy in the incongruent condition was significantly lower than in the negative condition, $F_1(1,19) = 59.87, p < .001, \eta^2_p = .76$. Accuracy in the negative condition was significantly lower than in the positive condition, $F_1(1,19) = 50.42, p < .001, \eta^2_p = .73$. Accuracy in the positive condition was significantly lower than in the congruent condition, $F_1(1,19) = 41.92, p < .001, \eta^2_p = .69$.

The ANOVA also revealed a main effect of time condition: accuracy was significantly lower under limited time than under unlimited time both for subjects, $F_1(1,19) = 6.72, p = .018, \eta^2_p = .26$, and for items, $F_2(1,32) = 13.72, p = .001, \eta^2_p = .30$. There was no interaction between comparison condition and time condition both for subjects, $F_1(3,57) < 1, p = .790$, and for items, $F_2(3,32) < 1, p = .746$.

A second analysis was conducted taking into account the experimental order. The main effects of comparison condition and of time condition remained the same for subjects and for items (lowest $F = 6.72$, highest $p = .018$). As in the previous analysis, there was no interaction between comparison condition and time condition (highest $F < 1$, lowest $p = .746$). There was no effect of experimental order for subjects, $F_1(1,19) < 1, p = .508$, but there was one for items, $F_2(1,32) = 9.43, p = .004, \eta^2_p = .23$. Across all conditions, participants were more accurate in the second half than in the first half, as would be expected with practice.

There was an interaction between time condition and experimental order for subjects, $F_1(1,19) = 4.46, p = .048, \eta^2_p = .19$, but not for items, $F_2(1,32) < 1, p = .337$. Across all comparison conditions, participants were more accurate under unlimited time than under limited time only in the first half, $t(19) = 3.38, p = .003, r = .26$. By the second half, there was no significant difference in accuracy between the unlimited time condition and the limited condition, $t(19) = 0.767, p = .452$.

However, there was no interaction between experimental order and comparison condition both for subjects, $F_1(3,57) = 1.06, p = .316$ (lower bound adjustment), and for items, $F_2(3,32) < 1, p = .415$. Finally, there was no interaction between time condition, experimental order and

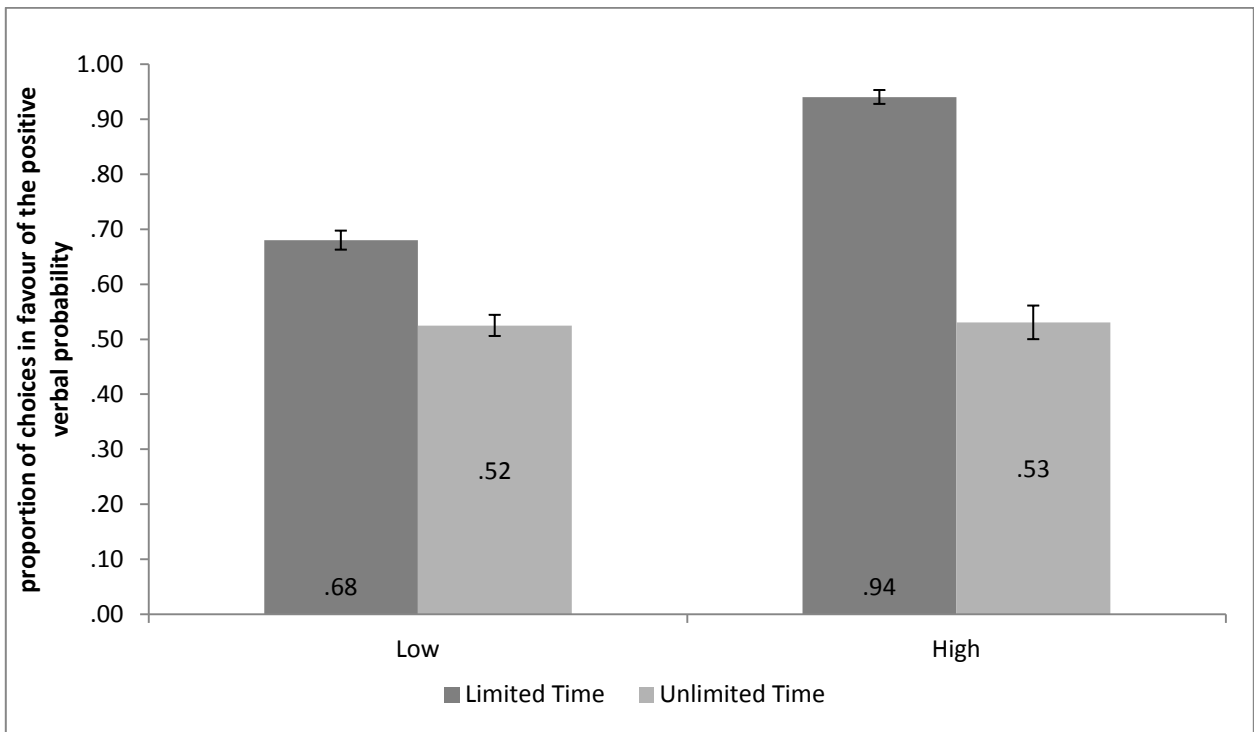
comparison condition, both for subjects, $F_1(3,57) < 1, p = .527$ (lower bound adjustment), and for items, $F_2(3,32) < 1, p = .528$. The effect of the comparison condition was not moderated by the progression in the task.

3.3.2.2. Directionality

I compared the choices made under the two time conditions and under the two conditions with no correct answer (low and high conditions) in order to check if participants' preference for the option presented with the positive verbal probability could be reduced by having more time to make a decision. As in Experiment 1, participants received 0 for each negative expression and 1 for each positive expression chosen, and these were averaged to give a score between 0 and 1. Mean proportions of choices of the positive verbal probabilities in the high and low conditions, by time condition, are presented Figure 11.

A 2 (comparison condition: low or high) by 2 (time condition: limited or unlimited) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on the preference for positive verbal probabilities: participants chose the outcome presented with a positive verbal probability significantly more often under the high condition than under the low condition both for subjects, $F_1(1,19) = 33.93, p < .001, \eta^2_p = .64$, and for items, $F_2(1,16) = 4.54, p = .049, \eta^2_p = .22$.

The ANOVA also revealed a main effect of time condition: participants chose the outcome presented with a positive verbal probability significantly more often under limited time than under unlimited time both for subjects, $F_1(1,19) = 247.81, p < .001, \eta^2_p = .93$, and for items, $F_2(1,16) = 35.81, p < .001, \eta^2_p = .69$.



^a error bars represent standard errors of the mean

Figure 11: Mean Proportion of Choices in Favour of the Positive Directionality as Function of the Comparison Condition and the Time Condition (Experiment 2b)

There was an interaction between comparison condition and time condition both for subjects, $F_1(1,19) = 25.14, p < .001, \eta^2_p = .57$, and for items, $F_2(1,16) = 7.36, p = .015, \eta^2_p = .32$. Under limited time participants chose the outcome described by a positive verbal probability more often in the high condition than in the low condition, $t(19) = 13.01, p < .001, r = .89$. This was not the case under unlimited time, $t(19) < 0.13, p = .90$.

I also ran one-sample t -tests which indicated that under limited time participants chose the outcome described by a positive verbal probability significantly more often than chance, both in the low and high conditions, $t(19) = 10.41, p < .001, r = .76$ and $t(19) = 34.86, p < .001, r = .97$

respectively; this was not the case under unlimited time, $t(19) = 1.30$, $p = .21$ and $t(19) = 1.00$ $p = .33$ respectively. However one-sample t -tests run for items found a preference for the outcome described by a positive verbal probability only in the high condition under limited time, $t(8) = 44.28$, $p < .001$, $r < .99$. In the high condition under unlimited time and in the low condition, under both limited and unlimited time, there was no preference for the outcome described by a positive verbal probability, $t(8) = 1.17$, $p = .274$, $t(8) = 1.72$, $p = .125$ and $t(8) = 1.04$, $p = .329$ respectively.

Finally, I also ran another ANOVA (with comparison condition and time condition as factors) taking also into account the experimental order. I found a similar main effect of comparison condition and of time condition, both for subjects and for items (lowest $F = 4.54$, highest $p = .049$). I also found a similar interaction between comparison condition and time condition, both for subjects and for items (lowest $F = 7.36$, highest $p = .015$).

There was no effect of experimental order, both for subjects, $F_1(1,19) < 1$, $p = .762$, and for items, $F_2(1,16) < 1$, $p = .713$. There was no interaction between experimental order and comparison condition both for subjects, $F_1(1,19) = 1.20$, $p = .288$, and for items, $F_2(1,16) = 1.98$, $p = .179$. There was no interaction either between time condition and experimental order, both for subjects, $F_1(1,19) < 1$, $p = .746$, and for items, $F_2(1,16) < 1$, $p = .737$. Finally, there was no interaction between time condition, experimental order and comparison condition, both for subjects, $F_1(1,19) < 1$, $p = .788$, and for items, $F_2(1,16) < 1$, $p = .801$. The effect of the comparison condition was not moderated by the progression in the task.

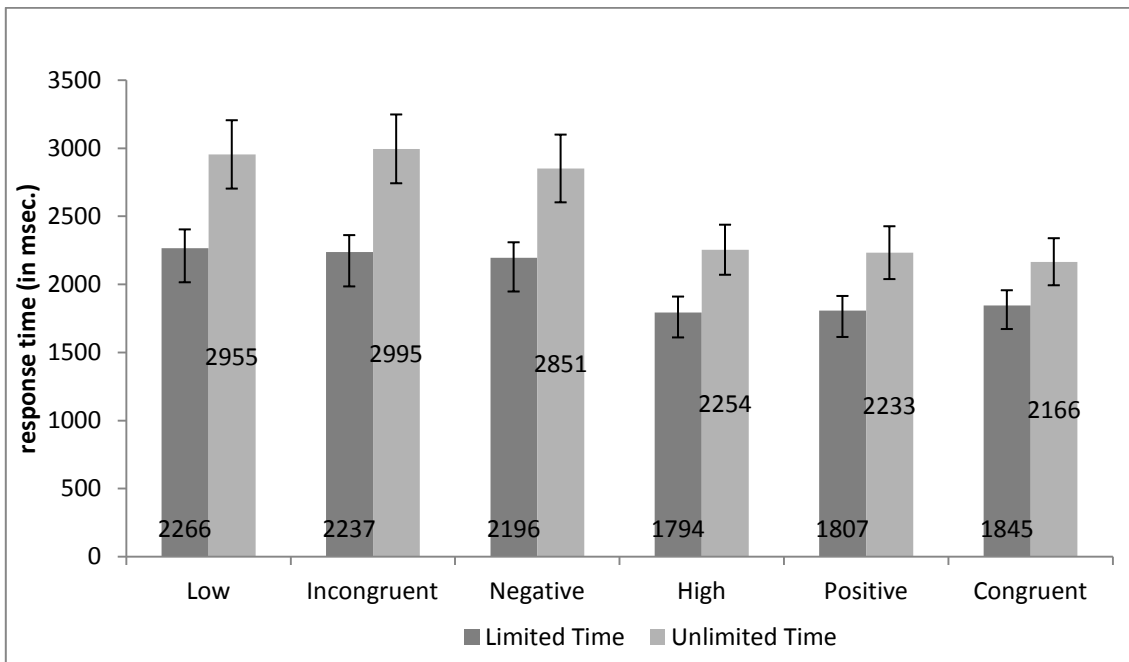
3.3.2.3. Response time

In order to investigate if making more accurate decisions and choices less biased towards the positive directionality was accompanied by taking more time to respond, I compared response times under the two time conditions and under each comparison condition (low, high, positive, negative, congruent and incongruent). I used the medians of each participant in each condition. Means of median response times according to the comparison and time conditions are presented Figure 12.

A 6 (comparison condition: incongruent, low, negative, high, positive or congruent) by 2 (time condition: limited or unlimited) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on the response time both for subjects, $F_1(5,95) = 27.82$, $p < .001$, $\eta^2_p = .59$ (lower bound adjustment), and for items, $F_2(5,48) = 13.06$, $p < .001$, $\eta^2_p = .58$. Planned repeated contrasts (order of conditions: incongruent, negative, low, positive, high, congruent) found only the negative and high conditions to be significantly different, $F_1(1,19) = 24.92$, $p < .001$, $\eta^2_p = .57$. Therefore I ran a post-hoc special contrast which confirmed that response time was longer under the low, incongruent and negative conditions, compared to the high, positive and congruent conditions (see Table 4 for F and p values).

The ANOVA also revealed a main effect of time condition. As would be expected, response time was significantly longer under unlimited time than under limited time both for subjects, $F_1(1,19) = 13.65$, $p = .002$, $\eta^2_p = .42$, and for items, $F_2(1,48) = 76.74$, $p < .001$, $\eta^2_p = .62$.

The ANOVA finally revealed an interaction between comparison condition and time condition, marginally significant for subjects, $F_1(5,95) = 3.29$, $p = .086$, $\eta^2_p = .15$ (lower bound



^a error bars represent standard errors of the mean

Figure 12: Mean Response Time (in msec.) as Function of the Comparison Condition and the Time Condition (Experiment 2b)

adjustment), and significant for items, $F_2(5,48) = 3.04$, $p = .018$, $\eta^2_p = .24$. Paired samples t -tests were applied with a Bonferroni correction (for 6 tests, $\alpha = .0083$). The response time was significantly longer under unlimited time than under limited time in the low condition, $t(19) = 4.41$, $p < .001$, $r = .40$, the incongruent condition, $t(19) = 4.21$, $p < .001$, $r = .442$, and the high condition, $t(19) = 3.29$, $p = .004$, $r = .32$. The difference between unlimited and limited time was not significant in the positive condition, $t(19) = 2.38$, $p = .028$, and the congruent condition, $t(19) = 2.82$, $p = .011$, and the negative condition, $t(19) = 2.91$, $p = .009$.

I ran two separate ANOVAs with the comparison condition (incongruent, low, negative, high, positive or congruent) as within-participant factor, respectively under limited time and unlimited time. An effect of comparison condition on the response time was confirmed both under limited time, $F_1(5,95) = 28.51, p < .001, \eta^2_p = .60$ (lower bound adjustment) and $F_2(5,48) = 21.16, p < .001, \eta^2_p = .69$, and unlimited time, $F_1(1,95) = 16.24, p = .001, \eta^2_p = .46$ (lower bound adjustment) and $F_2(5,48) = 8.45, p < .001, \eta^2_p = .47$. Both under limited and unlimited conditions, planned repeated contrasts applied with a Bonferroni correction (for 5

Table 4

Post-Hoc Special Contrasts on the Response Time (Experiment 2b)

compared conditions	F values	p values	η^2_p
low vs. high	41.26	< .001 ^a	.70
low vs. positive	29.34	< .001 ^a	.62
low vs. congruent	41.97	< .001 ^a	.70
incongruent vs. high	68.66	< .001 ^a	.79
incongruent vs. positive	38.2	< .001 ^a	.68
incongruent vs. congruent	51.8	< .001 ^a	.74
negative vs. high	27.97	< .001 ^a	.61
negative vs. positive	31.48	< .001 ^a	.64
negative vs. congruent	34.78	< .001 ^a	.66

^a significant after Bonferroni correction (for 9 tests, $\alpha = .0056$)

tests, $\alpha = .01$) found only the negative and high conditions to be different, $F_1(1,19) = 73.46$, $p < .001$, $\eta^2_p = .79$ and $F_1(1,19) = 12.73$, $p = .002$, $\eta^2_p = .40$ respectively.

A second analysis was conducted taking into account the experimental order. I found similar main effects of comparison condition and of time condition for subjects and for items (lowest $F = 13.06$, highest $p = .002$). I also found an identical interaction between comparison condition and time condition marginally significant for subjects and significant for items (lowest $F = 3.04$, highest $p = .086$).

There was a main effect of experimental order both for subjects, $F_1(1,19) = 40.26$, $p < .001$, $\eta^2_p = .68$, and for items, $F_2(1,48) = 134.42$, $p < .001$, $\eta^2_p = .74$. There was also an interaction between time condition and experimental order both for subjects, $F_1(1,19) = 14.67$, $p = .001$, $\eta^2_p = .44$, and for items, $F_2(1,48) = 53.32$, $p < .001$, $\eta^2_p = .53$. Participants' response time was shorter under limited time than under unlimited time in the first half, $t(19) = 4.04$, $p = .001$, $r = .47$. In the second half, there was no difference between limited time and unlimited time, $t(19) = 2.36$, $p = .029$, $r = .21$ (Bonferroni correction, $\alpha = .025$). In both conditions, as may be expected with practice, participants' response times were shorter in the second half of the experiment than in the first half, $t(19) = 2.77$, $p = .012$, $r = .18$ for limited time and $t(19) = 5.58$, $p < .001$, $r = .39$ for unlimited time (Bonferroni correction, $\alpha = .025$).

There was no interaction between experimental order and comparison condition both for subjects, $F_1(5,95) = 2.16$, $p = .065$, $\eta^2_p = .10$ (lower bound adjustment), and for items, $F_2(5,48) = 1.75$, $p = .142$. Finally, there was no interaction between time condition, experimental order and comparison condition for subjects, $F_1(5,95) = 2.19$, $p = .155$ (lower bound

adjustment), but there was for items, $F_2(5,48) = 2.88, p = .024, \eta^2_p = .23$. The interaction between the time condition and the comparison condition was significant in the first half, $F_2(5,48) = 3.56, p = .008, \eta^2_p = .27$, but not in the second half, $F_2(5,48) = 1.13, p = .356$.

Additional analyses run using only successful trials yielded similar results, with a main effect of comparison condition and of time condition, and an interaction of time and comparison conditions (lowest $F = 3.34$, highest $p = .003$).

3.3.3. Discussion

Overall, I observed the same patterns of accuracy and response time as in Experiment 1. Even when given unlimited time, participants were more accurate in the positive and the congruent conditions than in the negative and the incongruent conditions. However, an overall improvement of accuracy was observed under unlimited time. Both under limited and unlimited time, the performance in the incongruent condition was consistent with people's preference for positive verbal probabilities (e.g. Teigen & Brun, 1999). Participants failed to choose the correct answer described by a negative verbal probability on half of the trials. This goes against a performance cost account of the framing effect of directionality. If this effect was only due to directionality being easier to process, having more time should result in greater accuracy. That is, in the incongruent condition participants should have been more likely to choose the negative verbal probability over the positive one. However, in the low and high conditions, participants showed a preference for the positive verbal probabilities only under limited time. When given unlimited time, participants were equally likely to choose the positive or negative verbal probability. Even if performance costs are not sufficient to explain the framing effect of

directionality, reducing the time pressure can help to cancel out this effect when the two options have similar likelihood of occurring.

Slower response times in the incongruent and the low conditions, in both time conditions, confirmed that both the likelihood and directionality of a verbal probability are considered when making decisions. However under unlimited time, I observed an increase in the response time restricted to the low, incongruent, negative, and high conditions. Taking more time did not allow participants to increase their accuracy significantly in the incongruent and negative conditions. However, in the low and high conditions, taking more time was accompanied by a reduction of the preference for positive verbal probabilities. This also supports my claim that performance costs alone cannot account for the framing effect of directionality.

Having the opportunity to compare two verbal probabilities, coupled with having more time to make a decision, can help people to overcome the preference for positive verbal probabilities and therefore the framing effect of directionality. Nevertheless in the incongruent condition taking more time was not sufficient for participants to overcome this framing effect. Thus, having more time is helpful when verbal probabilities have equivalent likelihoods. When they carry different likelihoods, especially if these are incongruent with their respective directionality, taking more time is not a guarantee of accuracy.

3.4. General discussion

Empirical data has often shown a framing effect of directionality on decision making: adults prefer options described by positive verbal probabilities even when the options are equally likely (Teigen & Brun, 1999; Gourdon & Villejoubert, 2009). Although the preference for positive

verbal probabilities was not completely overcome in this chapter, it is noteworthy that for the first time, conditions that can reduce the framing effect of directionality were identified. When given the chance to compare two verbal probabilities communicating the same likelihood as well as given unlimited time to make a decision, people no longer systematically preferred the positive verbal probability. I also aimed to determine if the framing effect of directionality could be explained by this feature being easier to process. The relatively long response time in the low condition seems to indicate that participants considered the likelihood even when it was equivalent (and therefore not useful to the decision-making). Such a pattern should not be observed if the framing properties of directionality were relying on ease of processing, in which case the likelihood should have been ignored.

In the incongruent condition directionality was inconsistent with the likelihood. Performance in this condition was relatively inaccurate and slow. It is possible that pragmatic factors explain the remaining difficulty to overcome the framing effect of directionality under this particular condition. Pragmatic constraints require that speakers be as relevant as possible and that listeners make the least costly interpretation (e.g., Wilson & Sperber, 2004). Therefore if a listener detected the incongruence between the likelihood and directionality s/he may consider it as indicating a puzzling communicative intention from the speaker. If directionality had misled participants only because of its lower performance costs, I should have seen no increase in response times and an overwhelming preference for the positive verbal probabilities (in this condition the incorrect answer). On the other hand, if it is the case that the preference for positive verbal probabilities relies not only on performance costs but also on pragmatic

factors, I might expect that in settings that emphasise these factors (e.g. a naturalistic conversation) the preference for positive verbal probabilities might be even stronger and remain even when participants have unlimited time to respond. Future experiments should explore this question by replicating the second experiment with a clearly indicated conversational context.

We know that despite their vagueness, verbal probabilities are a preferred tool to communicate about uncertain events (e.g., Erev & Cohen, 1990). They are likely to be used in many settings, for example medical communication, although it was once recommended that they be banned from this type of interaction (Nakao & Axelrod, 1983). Teigen and Brun (2003) note that it would be hard to put this constraint on people's natural language and, elsewhere, it has been established that people have difficulties reasoning with the numerical probabilities that would need to replace them (e.g., Gigerenzer, Hertwig, van den Broek, Fasolo, & Katsikopoulos, 2005). However, I showed here that when comparing two options described by verbal probabilities with unlimited time, people were able to overcome their preference for positive verbal probabilities. Therefore, one way to establish better risk communication when using verbal probabilities is to give every option to the listener and to allow time to make a decision.

There are situations in everyday life when one has to make decisions under time pressure. Imagine you are in a consultation with your doctor and she offers you the choice between two treatments with descriptions of their likely success. Often one is expected to make the decision promptly, but it might be that people make better informed decisions if they are given more

time to decide. Ensuring patients have the best chance to make decisions is important as it is thought that physicians underestimate patients' need for fuller disclosure of risks (Bismark, Gogos, Clark, Gruen, Gawande & Studdert, 2012). If doctors are encouraged to make efforts to improve this disclosure, then how to communicate the risk is an important concern. An interesting possibility is that simply allowing decision makers to feel that the time pressure is relieved may even be sufficient in itself. Notably even when their decisions were not subject to the framing effect of directionality (under unlimited time in the low and high conditions), my participants did not take a very long time: their mean response times under the low and the high conditions were still under 5 seconds. However, such strategies will lead to improvement only in cases when the options have similar likelihoods of success/failure. Indeed in the incongruent condition, where there was a more likely option, participants still performed at chance level despite being offered all the time they needed and using it. If the options to decide between are known to have different likelihoods, one strategy would be to present people with verbal probabilities where directionality was congruent with this likelihood.

My studies demonstrated that measuring response times and manipulating the time pressure are useful tools to research decision making based on verbal probabilities. One limitation is that my paradigm does not distinguish between time taken to read the verbal probabilities and the time taken to decide between them. However, it is unclear how these processes could be separated in such a task as participants will no doubt start to interpret verbal probabilities while they are reading them. This observation does not undermine my results as the different verbal probabilities were repeated in conditions that had slow and fast

response times. Furthermore, the cost of processing negation cannot explain the different patterns observed in response times as conditions with both slow and fast response times contained negative phrases (the only exception was the positive condition). Thus, it is reasonable to conclude that it is the decision making time which differs between conditions.

I mentioned that in translation tasks (including unpublished data I used to select the materials), some of the positive and negative verbal probabilities which I categorized as having a similar likelihood are often judged rather differently most likely due to directionality (e.g., *not absolutely certain* and *very likely*). However, it is noteworthy that given unlimited time in the low and high conditions, where only the likelihood could drive the decision participants showed no preference for the expressions usually translated as having a higher likelihood (the positive ones). This supports my claim that the materials I used was adequately chosen and suggests that the effect of directionality in translation tasks can be overcome in some circumstances.

One could also argue that my samples were limited: they were relatively small, and predominantly of female participants and of psychology students. There is some evidence that women are more prone to framing effects (e.g., Fagley & Miller, 1990), yet these gender differences in framing are modulated by the task domain (e.g., Huang & Wang, 2010): for example, women were more sensitive to negative attribute framings when it comes to life or death, while men showed a sensitivity to negative attribute framings when the task was related to finances. I expect that my results should also apply to men, if and when the task taps in the right domain. Indeed, it would be useful to explore whether the results differ between the genders when the task describes different domains. Furthermore across the two experiments, I

considered it a strength was that the samples were drawn from different populations. In Experiment 2, the sample was made of undergraduate psychology students, and in Experiment 1, participants were 17-year-old high school students who had only recently begun their introductory study of psychology. More critically, the high school participants had rarely if ever participated in psychology experiments. Despite this, I found the same pattern in both experiments, suggesting that my results are unlikely to be the product of task demands and/or of the lack of naïvety of participants.

I demonstrated that the framing effect of directionality in verbal probabilities can be reduced by relieving the time pressure. I also observed patterns of results indicating that this framing effect does not rely on directionality being easier to process. Instead, I suggest that pragmatic factors also contribute to the framing effect of directionality. Verbal probabilities are not merely a way to communicate risk, they may also communicate information about a speaker's intention.

CHAPTER 4: MAKING DECISIONS BASED ON VERBAL PROBABILITIES IN A CONVERSATIONAL CONTEXT: SPEAKERS' INTENTIONS MATTER

4.1. Introduction

When choosing a verbal probability to express a level of uncertainty, a speaker is confronted with a wide range of phrases, from the rare (e.g., *a ghost of a chance*), to the most informal (e.g., *a snowball in hell's chance*), to the well-established (e.g., *unlikely*). A speaker thus needs to choose a level of language and the appropriate phrase to communicate a level of uncertainty. Furthermore, a speaker needs to choose the directionality of the phrase, that is the positive or negative quality of a verbal probability. Choosing one directionality or the other is as important as appropriately translating the level of uncertainty the speaker wants to communicate. Directionality drives listeners' attention to the occurrence or the non-occurrence of the event (Teigen & Brun, 1995), which results in framing effects in judgement and decision making (Teigen & Brun, 1999). Between equally likely options, people chose the one described with a positive verbal probability more often than the one described with a negative phrase (e.g., Gourdon & Villejoubert, 2009). In Chapter 3 I suggested that this may be because directionality signals the speaker's intention to the listener. In this chapter I investigated this possibility by extending Experiment 2b: first, by setting out a conversational context; second, by setting out the speakers' benevolent or malevolent intention.

Weber and Hilton (1990) showed that the numerical translation of verbal probabilities into a risk estimate is sensitive to contextual factors such as the severity of the predicted event. For

example, *possible* was interpreted as representing a 50% chance of developing a wart, whereas it was interpreted as meaning a 62% chance of having a sprained ankle. More importantly, Bonnefon and Villejoubert (2006) showed that verbal probabilities were not only prone to the severity bias, but also to the listener's pragmatic interpretations. When the uncertain event was severe (deafness), the numerical translation of *It is possible* was higher (67%) than when the uncertain event was mild (insomnia; 59%). Crucially, when they interpreted that the speaker used a verbal probability to be tactful, participants gave a higher numerical translation of *It is possible* (e.g., 71% for insomnia) than when they interpreted that the speaker used a verbal probability because he was not sure (e.g., 57% for insomnia). Participants also interpreted more often that the speaker was being tactful in the case of deafness (60%) than in the case of insomnia (17%). Thus, verbal probabilities seem to be prone to pragmatic influences, where a speaker's intentions are taken into account in the numerical interpretation of the statements.

Juanchich, Teigen & Villejoubert (2010) extended the pragmatic account of verbal probabilities to include their directionality. They found that speakers used positive and negative verbal probabilities differentially in order to contradict someone else's prediction. Participants were asked to select what two profilers, Tom and David, would say in a conversation where they disagree on the likelihood of a suspect's guilt. When David thought that the likelihood was higher than Tom had just said, participants chose a positive verbal probability more often to express David's prediction. When David thought that the likelihood was lower than the one predicted by Tom, participants were more likely to choose a negative verbal probability to express David's prediction. Challenging Budescu, Karelitz and Wallsten's position (2003) that

directionality merely reflects the numerical information of a verbal probability, this supported the claim that directionality also conveys 'a consistent message, with clear implications for inferential judgments' (Teigen & Brun, 1999, p. 185). In the case of revised judgements as in Juanchich et al., directionality seems to be chosen so it can frame listeners' perspective in the direction of the wished revision. As far as I know however, no one has yet investigated such pragmatic effects within a conversational context, on the interpretation rather than the production of directionality.

In Chapter 3 I showed that adults chose the least likely option over the most likely one in about half of the cases where the least likely option was presented by a positive verbal probability (while the most likely one was presented by a negative verbal probability). In contrast with conditions where the likelihood of the two options was similar, this was the case even when the time pressure was relieved. A crucial feature in the former case was that the likelihood and directionality were incongruent. I suggested that listeners interpret that speakers may signal their intention by choosing incongruent verbal probabilities. If it did not allow speakers to signal intention, choosing an incongruent verbal probability would have no pragmatic relevance, violating the maxim of manner (i.e., that a speaker should avoid obscuring the communication; Grice, 1975). Therefore listeners infer that speakers choose it because they want the listener to choose the option presented by a positive verbal probability, even if it is the least likely.

According to this pragmatic interpretation of directionality, the preference for positive verbal probabilities would be expected to be manifested even more strongly in a more

conversational context. In this context, the speaker is expected to produce utterances that are useful to process (e.g., Wilson & Sperber, 2004). Therefore, like the case of incongruent verbal probabilities, the listener would interpret that the speaker's choice of a positive directionality is not arbitrary but signals something. Therefore, people should have more difficulty overcoming this preference, even under favourable conditions (i.e., with unlimited time). Furthermore, if a positive directionality is indicative of what the speaker suggests the listener chooses, one would not want to choose the option presented by a positive verbal probability if the speaker is malevolent. Imagine a poker game: player A is teasingly asking player B if s/he has a good hand. Player B has a good hand but does not want to deter player A from betting more money and making the pot bigger. Player B therefore may reply that 'It is not absolutely certain'. But, not being a naive player, player A will infer that B is trying to mislead her and will understand that 'It is likely'. Thus she will fold. Therefore another prediction of the pragmatic interpretation of directionality is that the preference for positive verbal probability should depend on the intentions attributed to speakers.

Such a prediction is supported by recent work in the theoretical frame of epistemic vigilance (e.g., Sperber et al., 2010). Epistemic vigilance can be seen as the set of cognitive tools that ensure that an individual does not waste cognitive resources on understanding a misleading communication. These mechanisms range from reasoning abilities to evaluation of trustworthiness, and target the content as well as the speaker. For example, Mascaro and Sperber (2009) found that, as early as 4 years old, children do the opposite of what was indicated if the speaker is thought to be malevolent. Just like the poker player A, they evaluated

the trustworthiness of the speaker. When they deemed him untrustworthy, they changed their interpretation of the information they received from the speaker.

In Chapter 2, I also found that children take into account the intention of the speaker when judging and making decisions based on verbal probabilities. However I failed to find the reversal expected in a pragmatic account of directionality. While children changed their use of directionality when the speaker was malevolent, they neglected directionality, rather than using it as a reverse cue. One possibility I suggested then was that it might be too costly for children to process the interaction of a speaker's intention and of directionality. Dismissal of directionality to choose randomly would then be the best next option. If this is the case, it could be expected that adults however combine the intention and directionality in order to make a decision. If directionality is an indicator to listeners of what the speakers want, listeners should infer that a malevolent speaker would use this cue to mislead them. To do so the speaker however needs to know the correct answer. So listeners can in fact infer the correct answer from what the speaker says and from his intention. Therefore adults could use directionality as a cue, but in reverse. This would support a pragmatic account of directionality. But if listeners only dismissed the speaker's answer to choose randomly, like children did in Experiment 1b, it would suggest that they consider that directionality is not a cue of what the speaker wants.

This chapter aimed to test the pragmatic account of directionality in two experiments based on the paradigm I used on Chapter 3. Two treasure chests were presented to participants, each chest being described as possibly containing gold coins, using a verbal probability. Verbal probabilities could differ in directionality (positive/negative) and probabilistic meaning

(high/low). Participants had to decide on a single chest to open, under limited time and without time limit. In Experiment 3a, silhouettes and speech bubbles were added to the paradigm to provide participants with a schematic conversational context. In Experiment 3b, silhouettes and speech bubbles were also used; I did not manipulate the time limit (giving only limited time), but crucially manipulated speakers' intentions as benevolent or malevolent, by introducing two different characters (Peter Pan and Captain Hook, as in Experiment 1b).

In Experiment 3a, I expected to find patterns of decisions and reaction times similar to those in Chapter 3, except that, even when relieved from time pressure, participants would prefer the positive verbal probabilities in every comparison condition. Given the conversational context, listeners would be expected to interpret that the positive directionality has been chosen purposely; so listeners would most of the time rely on directionality to make their decision, choosing the positive one, since they would think it indicates what the speaker wants. This would lead them to choose the incorrect option more often when asked to choose between two incongruent verbal probabilities.

In Experiment 3b, I again expected to find a similar pattern of decisions and reaction times as in Chapter 3 (i.e., similar to those in Experiment 3a), but only in the case of the benevolent speaker. When the speaker was malevolent, I expected that decisions would be reversed, showing a preference for the option presented by a negative verbal probability. In conditions where directionality was the same, if listeners interpreted that the speaker was malevolent and trying to mislead them, they could interpret that the speaker is describing the option that is actually less likely as more likely. Therefore, they would show a preference for the option

presented as the least likely. I also expected that response times would be longer when the speaker was malevolent, as a consequence of listeners reversing their initial interpretation.

Thus, in Experiment 3a, under any time condition, when the probabilistic meaning of the two descriptions was similar and directionality was the only thing changing, it was expected that participants would choose the positive verbal probability more often, as people prefer options given with positive verbal probabilities (Teigen & Brun, 1999). Moreover, it was expected that participants would choose the positive verbal probability even if directionality contradicted the probabilistic meaning (e.g., positive verbal probability conveying a low probabilistic meaning), then making more incorrect decisions (defined as choosing the least likely option). In the conversational context, listeners would interpret that the positive directionality has been chosen purposely, therefore choosing more often the positive verbal probability as it would signal to them what the speaker wants.

People prefer to use positive verbal probabilities for high probabilistic meanings and low ones for low probabilistic meanings (Budescu et al., 2003). Thus, I expected participants to choose the positive verbal probability over the negative to a greater extent when both had a high probabilistic meaning, and that this would be true under both time conditions. Regarding response times, following the results of Chapter 3, I expected participants to take longer to decide between: two negative verbal probabilities (because of the cost of negation); between two verbal probabilities conveying a low likelihood (because of the inconsistency with the task goal, finding some treasure); and between two incongruent verbal probabilities (e.g., a positive verbal probabilities conveying a low likelihood; because of the inconsistency between the

message of directionality and the message of the likelihood). Shorter response times were expected when a decision was made between two positive verbal probabilities, two congruent verbal probabilities and two verbal probabilities carrying a high likelihood.

4.2. Experiment 3a

4.2.1. Method

4.2.1.1. Participants

Nineteen psychology students (mean age = 19.45 years; range = 18 - 23; 18 girls) at university of Birmingham (UK) took part in the experiment against course credits.

4.2.1.2. Materials

The materials used were very similar to those used in the Experiments 2a and 2b. The only difference was that each verbal probability was presented in a speech bubble coming out of one of two identical (but mirroring) silhouettes (see Figure 13 for an example). As in Experiment 2a and 2b, the outcome the verbal probabilities referred to was kept constant (i.e. *there are some coins in the chest*).

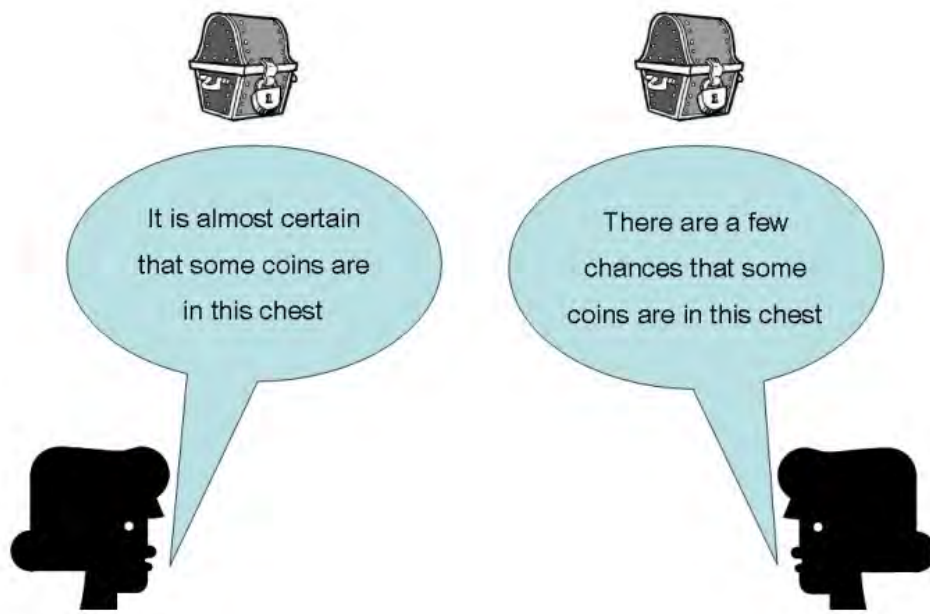


Figure 13: Example of Stimuli Used in Experiment 3a

4.2.1.3. Procedure

The procedure was very similar to Experiment 2b. Differences included changes in the instructions to reinforce the conversational nature of the task, and in the stimuli presentation, for the same reason. Instructions were as follows:

‘You are going to be presented pairs of treasure chests. In every pair one chest is more likely to contain some gold coins. To help you deciding which chest you should open, some people will give you some hints. Sometimes the two hints will seem very similar to you but they are actually always different within each pair. For each pair you will then have to say if you want to open the left chest (z key) or the right one (m key).’

Each trial was introduced by a fixation cross. In order to reinforce the impression of a conversation, I first presented the silhouettes without speech bubbles, for one second, immediately followed by the silhouettes uttering the two verbal probabilities in the speech bubbles.

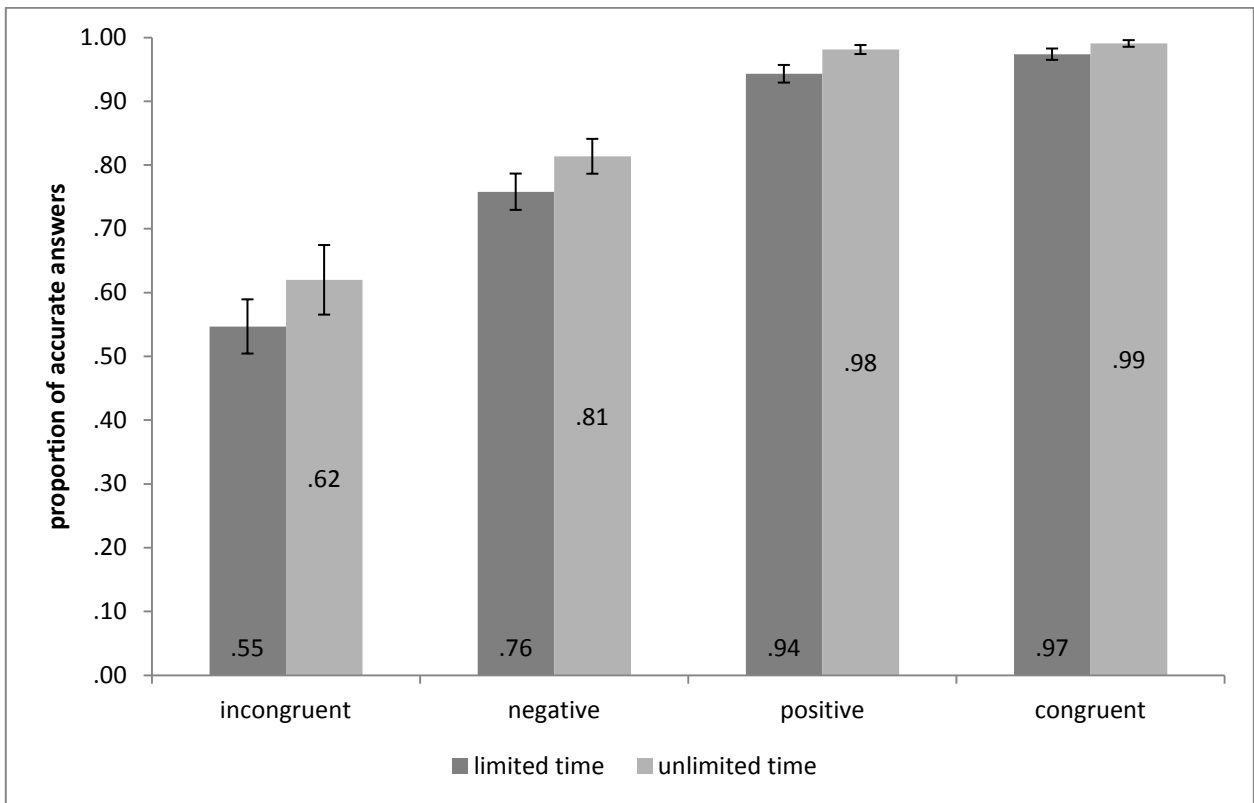
4.2.2. Results

As in Chapter 3, participants were presented with each verbal probability 36 times, having the opportunity to familiarize themselves with the materials. Hence each analysis was conducted twice, the second time including the experimental order as additional factor to exclude familiarity and/or length of the pairs as confounding variables.

4.2.2.1. Accuracy

In order to test if participants were more accurate when they had more time to make a decision, I compared the performance on the four conditions where a correct answer existed (positive, negative, incongruent and congruent conditions), under the two time conditions. As in Experiments 2a and 2b, accuracy ranged from 0 to 1. Mean proportions of accurate answers according to the comparison condition (when accuracy is a relevant measure) and the time condition are presented in Figure 14.

A 4 (comparison condition: incongruent, negative, positive or congruent) by 2 (time: condition: limited or unlimited) revealed a main effect of comparison condition on accuracy both for subjects, $F_1(3,54) = 61.28, p < .001, \eta^2_p = .77$, and items, $F_2(3,32) = 12.86, p < .001, \eta^2_p = .55$. Planned repeated contrasts (with Bonferroni correction: for 3 tests, $\alpha = .0166$)



^a error bars represent standard errors of the mean

Figure 14: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Time Condition (Experiment 3a)

indicated that accuracy in the incongruent condition was significantly lower than in the negative condition, $F_1(1,18) = 27.51, p < .001, \eta^2_p = .60$. Accuracy in the negative condition was significantly lower than in the positive condition, $F_1(1,18) = 57.55, p < .001, \eta^2_p = .76$. Accuracy in the positive and the congruent conditions were similar, $F_1(1,18) = 4.70, p = .044$. The ANOVA also revealed a main effect of time condition: accuracy was significantly lower under limited time than under unlimited time both for subjects, $F_1(1,18) = 19.25, p < .001, \eta^2_p = .52$, and for items, $F_2(3,32) = 41.23, p < .001, \eta^2_p = .56$.

There was no interaction between comparison condition and time condition for subjects, $F_1(3,54) = 2.00, p = .125$, but there was one for items, $F_2(3,32) = 3.56, p = .025, \eta^2_p = .25$. Accuracy was lower under limited time than under unlimited time under the incongruent and the positive conditions, $t(8) = 5.99, p < .001, r = .16$ and $t(8) = 3.56, p = .007, r = .52$, respectively. Under the congruent and the negative conditions, there was no difference between the limited time and the unlimited time, $t(8) = 2.82, p = .023$, and $t(8) = 2.33, p = .049$, respectively (Bonferroni correction; $\alpha = .0125$). The effect of the comparison condition was significant both under limited time, $F_1(3,54) = 72.92, p < .001, \eta^2_p = .80$, and unlimited time, $F_1(3,54) = 41.20, p < .001, \eta^2_p = .70$.

A second analysis was conducted taking into account the experimental order. The main effect of comparison condition and of time condition remained for subjects and for items (lowest $F = 12.86$, highest $p < .001$). The interaction between comparison condition and time condition was the same, that is, not significant for subjects, significant for items (highest $F = 3.56$, lowest $p = .0250$). There was no effect of the experimental order, for subjects or items. There was no interaction between time condition and experimental order, for subject or for items. Finally there was no interaction between time condition, experimental order and comparison condition, for subjects or for items, (highest $F = 1.81$, lowest $p = .195$).

There was however an interaction between comparison condition and experimental order, not for subjects, $F_1(3,54) = 1.81, p = .195$ (lower bound adjustment), but one for items, $F_2(3,32) = 2.95, p = .047, \eta^2_p = .22$. There was no difference between the first and the second block in the positive, the congruent, the incongruent and the negative conditions, $t(8) = 2.26,$

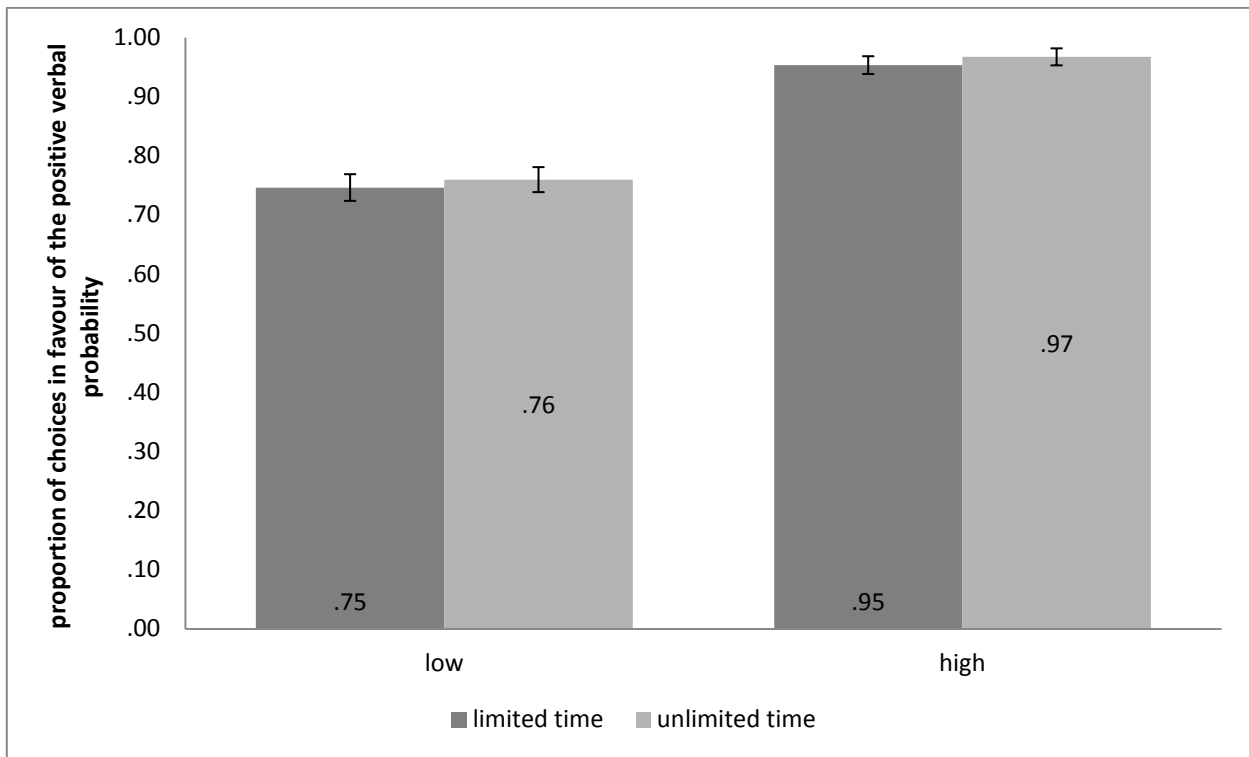
$p = .054$ and $t(8) = 1.79$, $p = .111$, $t(8) = 1.61$, $p = .147$ and $t(8) = 0.56$, $p = .589$, respectively (Bonferroni correction; $\alpha = .0125$). There was a main effect of condition in the first half, $F_2(3,32) = 12.97$, $p < .001$, $\eta^2_p = .55$, as well as in the second half, $F_2(3,32) = 12.41$, $p < .001$, $\eta^2_p = .54$. There was no suggestion that increasing familiarity with items over the course of the experiment affected the results.

4.2.2.2. Directionality

I compared the choices made under the two time conditions and under the two conditions with no correct answer (low and high conditions) in order to check if participants' preference for the option presented with the positive verbal probability could be reduced by having more time to make a decision. As in Experiment 3a and 3b, participants received 0 for each negative expression and 1 for each positive expression chosen, and these were averaged to give a score between 0 and 1. Mean proportions of choices of the positive verbal probabilities in the high and low conditions are presented in Figure 15.

A 2 (comparison condition: low or high) by 2 (time condition: limited or unlimited) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on the preference for positive verbal probabilities. Participants chose the outcome presented with a positive verbal probability significantly more often under the high condition than under the low condition for subjects, $F_1(1,18) = 123.94$, $p < .001$, $\eta^2_p = .87$, and marginally for items, $F_2(1,16) = 4.13$, $p = .059$, $\eta^2_p = .21$.

The ANOVA revealed no effect of time condition: participants chose the outcome presented with a positive verbal probability as often under limited time as under unlimited time both for



^a error bars represent standard errors of the mean

Figure 15: Mean Proportion of Choices in Favour of the Positive Verbal Probability (Standard Deviation) as Function of the Comparison Condition and the Time Condition (Experiment 3a)

subjects, $F_1(1,18) = 1.91, p = .184$, or for items, $F_2(1,16) = 1.08, p = .315$. There was no interaction either between the comparison condition and the time condition both for subjects, $F_1(1,18) < 1, p = .977$, or for items, $F_2(1,16) < 1, p = .949$.

One-sample t -tests indicated that under limited time participants chose the outcome described by a positive verbal probability significantly more often than chance, both in the low and high conditions, $t(18) = 10.91, p < .001, r = .78$ and $t(18) = 30.16, p < .001, r = .96$

respectively; this was also the case under unlimited time, $t(18) = 12.28$, $p < .001$, $r = .82$ and $t(18) = 32.49$, $p < .001$, $r = .97$ respectively.

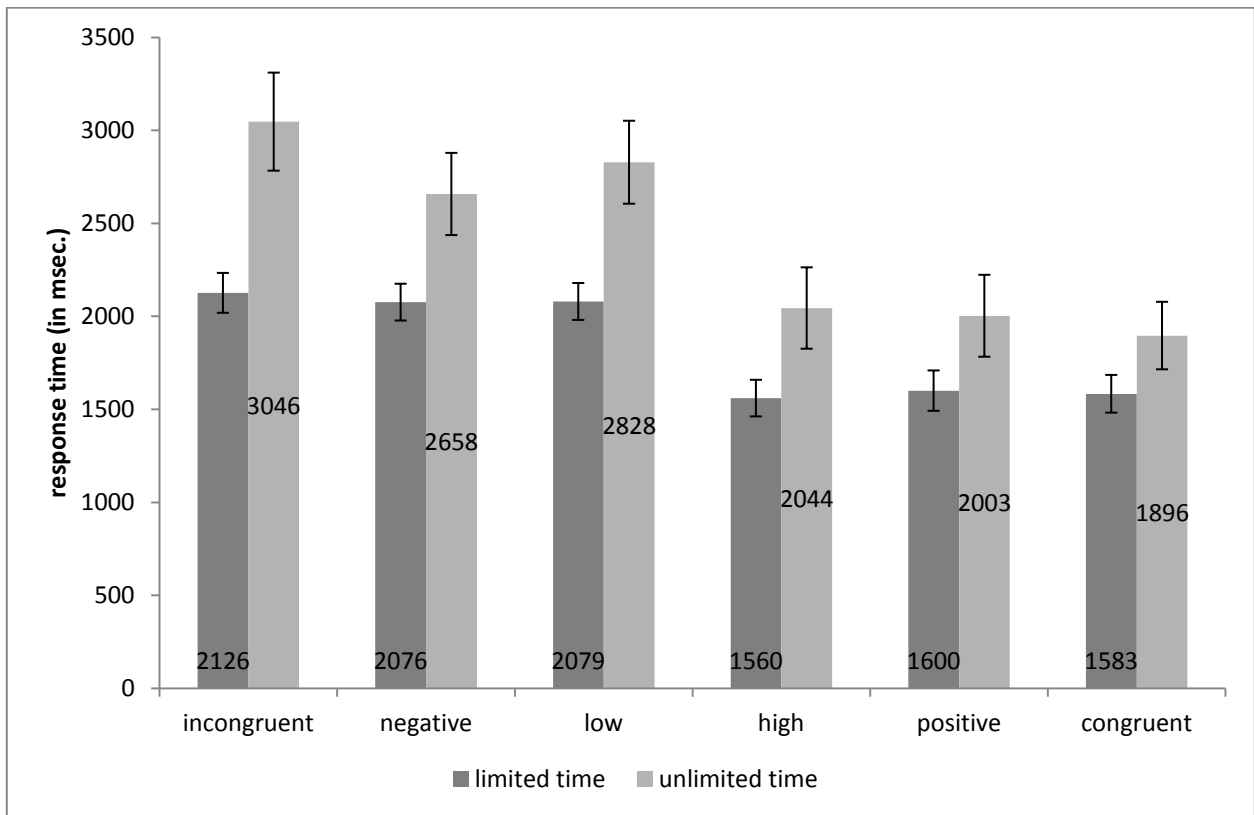
A second analysis was conducted taking into account the experimental order. The main effect of comparison condition remained significant for subjects and marginal for items (lowest $F = 4.13$, highest $p = .059$). There was still no main effect of time condition and no interaction of comparison condition and time condition (highest $F < 1$, lowest $p < .995$) There was a marginal effect of experimental order for subjects, $F_1(1,18) = 3.73$, $p = .069$, $\eta^2_p = .17$, but not for items, $F_2(1,16) = 2.59$, $p = .127$. In the second half, participants chose more often the option described by the positive verbal probability than in the first half.

There was no interaction between comparison condition and experimental order, for subjects or for items, between comparison condition and experimental order, for subjects or for items, and between time condition, experimental order and comparison condition, for subjects or for items (highest $F = 2.30$, lowest $p = .147$).

4.2.2.3. Response time

In order to investigate if making more accurate decisions and choices that were less biased towards the positive directionality was accompanied by taking more time to respond, I compared response times under the two time conditions and under each comparison condition (low, high, positive, negative, congruent and incongruent). As in Experiments 3a and 3b, the medians of each participant/item in each condition were used to deal with outliers. Means of median response times according to the comparison conditions are presented Figure 16.

A 6 (comparison condition: incongruent, low, negative, high, positive or congruent) by 2 (time condition: limited or unlimited) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on the response time both for subjects, $F_1(5,90) = 53.49$, $p < .001$, $\eta^2_p = .75$, and for items, $F_2(5,48) = 39.91$, $p < .001$, $\eta^2_p = .81$. Planned repeated contrasts (order of conditions: incongruent, negative, low, positive, high, congruent) found only the low and high conditions to be significantly different, $F_1(1,18) = 64.71$, $p < .001$, $\eta^2_p = .78$.



^a error bars represent standard errors of the mean

Figure 16: Mean Response Time (in msec.) as Function of the Comparison Condition and the Time Condition (Experiment 3a)

Therefore I ran a post-hoc special contrast analysis which confirmed that response time was longer under the low, incongruent and negative conditions, compared to the high, positive and congruent conditions (see Table 5 for F and p values).

The ANOVA also revealed a main effect of time condition. As would be expected with a time pressure paradigm, response time was significantly longer under unlimited time than under limited time both for subjects, $F_1(1,18) = 8.16, p = .010, \eta^2_p = .31$, and for items, $F_2(1,48) = 214.72, p < .001, \eta^2_p = .82$.

Table 5

Post-Hoc Special Contrasts on the Response Time (Experiment 3a)

compared conditions	F values	p values	η^2_p
low vs. high	64.71	< .001 ^a	.78
low vs. positive	101.30	< .001 ^a	.85
low vs. congruent	149.71	< .001 ^a	.89
incongruent vs. high	64.85	< .001 ^a	.78
incongruent vs. positive	97.36	< .001 ^a	.84
incongruent vs. congruent	75.18	< .001 ^a	.81
negative vs. high	47.22	< .001 ^a	.72
negative vs. positive	88.84	< .001 ^a	.83
negative vs. congruent	100.91	< .001 ^a	.85

^a significant after Bonferroni correction (for 9 tests, $\alpha = .0056$)

The ANOVA revealed an interaction between comparison condition and time condition, for subjects, $F_1(5,90) = 7.51, p = .013, \eta^2_p = .29$ (lower bound adjustment), and for items, $F_2(5,48) = 10.89, p < .001, \eta^2_p = .53$. Paired samples *t*-tests were applied with a Bonferroni correction (for 6 tests, $\alpha = .0083$). The response time was significantly longer under unlimited time than under limited time in the low condition, $t(18) = 3.41, p = .003, r = .45$, and the incongruent condition, $t(18) = 3.37, p = .003, r = .47$. The difference between unlimited and limited time was not significant in the positive condition, $t(18) = 2.17, p = .044$, the congruent condition, $t(18) = 1.76, p = .096$, the high condition, $t(18) = 2.32, p = .033$, and the negative condition, $t(18) = 2.75, p = .013$.

I ran two separate repeated-measures ANOVAs with the comparison condition (incongruent, low, negative, high, positive or congruent) as within-participant factor, respectively under limited time and unlimited time. An effect of comparison condition on the response time was confirmed both under limited time, $F_1(5,90) = 49.38, p < .001, \eta^2_p = .73$ and $F_2(5,48) = 32.59, p < .001, \eta^2_p = .77$, and unlimited time, $F_1(5,90) = 32.94, p < .001, \eta^2_p = .65$ (lower bound adjustment) and $F_2(5,48) = 33.31, p < .001, \eta^2_p = .78$. Under limited time, planned repeated contrasts applied with a Bonferroni correction (for 5 tests, $\alpha = .01$) found only the low and high conditions were different, $F_1(1,18) = 111.67, p < .001, \eta^2_p = .86$, with shorter response times in the high condition than in the low condition. Under unlimited time, the low and high conditions were different too, $F_1(1,18) = 30.51, p < .001, \eta^2_p = .63$. Moreover, the negative and the low conditions were also different, $F_1(1,18) = 8.36, p = .010, \eta^2_p = .32$, with shorter response times in the negative condition than in the low one.

The same analysis was conducted taking into account the experimental order. The main effect of comparison condition and of time condition remained for subjects and for items, as well as the interaction between comparison condition and time condition (lowest $F = 8.24$, highest $p = .010$). There was a main effect of experimental order both for subjects, $F_1(1,18) = 7.12, p = .016, \eta^2_p = .28$, and for items, $F_2(1,48) = 99.27, p < .001, \eta^2_p = .67$. Across all conditions, response times were shorter in the second half of the task than in the first one, as would be expected from the effect of practice.

There was an interaction between comparison condition and experimental order both for subjects, $F_1(5,90) = 2.59, p = .031, \eta^2_p = .13$, and for items, $F_2(5,48) = 3.15, p = .015, \eta^2_p = .25$. Response times were significantly shorter in the second half compared to the first one in the congruent and the negative conditions, $t(18) = 3.25, p = .004, r = .28$ and $t(18) = 2.36, p = .029, r = .29$ respectively. In the incongruent, low, high and positive conditions, the first and the second half yielded similar response times, $t(18) = 2.20, p = .041, t(18) = 2.29, p = .034, t(18) = 2.13, p = .047$ and $t(18) = 0.39, p = .703$, respectively. The main effect of comparison condition observed in the first half, $F_1(5,90) = 46.27, p < .001, \eta^2_p = .72$, remained in the second half, $F_1(5,90) = 24.39, p < .001, \eta^2_p = .58$.

There was no interaction between experimental order and time condition for subjects, $F_1(1,18) = 1.23, p = .282$, but there was for items, $F_2(1,48) = 27.45, p < .001, \eta^2_p = .36$. Participants' response time was shorter under limited time than under unlimited time in the first half, $t(53) = 9.45, p < .001, r = .46$, as well as in the second half, $t(53) = 8.10, p < .001, r = .35$, with a smaller effect in the second half. Participants' response time was shorter from the

first half to the second half, both under limited time, $t(53) = 6.57, p < .001, r = .25$, and under unlimited time, $t(53) = 7.88, p < .001, r = .36$, with a bigger effect under unlimited time.

Finally, there was no interaction between time condition, experimental order and comparison condition both for subjects, $F_1(5,90) = 1.64, p = .216$ (lower bound adjustment), or for items, $F_2(5,48) = 1.96, p = .102$.

Additional analyses run using only successful trials yielded similar results, with a main effect of comparison condition and of time condition, and an interaction of time and comparison conditions (lowest $F = 6.89$, highest $p = .017$).

4.2.3. Discussion

Experiment 3a repeated Experiment 2b in a more conversational context. Following a pragmatic account of directionality, it was expected that participants would show a preference for positive verbal probabilities in all conditions where directionality differed and under each time condition. In line with this expectation, in the incongruent condition participants chose the incorrect chest (the one with the lowest likelihood, but described by a positive directionality) half of the time, whereas when the verbal probabilities were congruent, the correct chest (with the highest likelihood to contain gold coins) was picked most of the time. In the low and high conditions, in which there was no correct answer, participants preferred the chest described by a positive verbal probability more often than chance. However the preference for the positive verbal probability was significantly lower in the low condition than in the high condition, reflecting that negative verbal probabilities are preferred for lower likelihoods, and positive verbal probabilities are preferred for higher likelihood (Budescu et al., 2003). Crucially, this was

the case even when participants were given more time to decide, in opposition to what was observed in Experiment 2b.

While I suggested that having more time to compare each verbal probabilities can help to overcome the preference for positive verbal probabilities (Experiment 2b), I believe that this is only the case when the task is set up without any conversational context. One could argue that the schematic conversational context provided here was minimalistic. However, the pattern of results I observed in Experiment 3a was different from Experiment 2b, suggesting that the conversational context was sufficient to elicit answers that would be expected in a pragmatic framework. When the two verbal probabilities are provided within a conversation, pragmatic rules seem to lead listeners to interpret that the speaker must have chosen purposely to use a positive directionality. In a Gricean perspective of pragmatics, it would otherwise violate the maxim of quality (i.e., that a speaker should tell what s/he knows is true; Grice, 1975). In a relevance interpretation (e.g., Wilson & Sperber & Wilson, 2004), it would violate that principle that utterances should be relevant to be worth being processed. The preference for positive verbal probabilities both under limited time and under unlimited time therefore supports that directionality is a pragmatic feature of verbal probabilities.

Response times showed the same patterns as in Experiments 2a and 2b, with the negative, incongruent and low conditions eliciting longer times than the high, positive and congruent conditions. In the incongruent condition, the conflict between directionality and the likelihood seem to necessitate more resources. Take the case of a positive verbal probability conveying a low likelihood (e.g., *there are a few chances*): as a pragmatic cue, directionality tells the decision

maker that the speaker wants him or her to open the chest, but the likelihood is indicating that s/he should not. The conflict between those two pieces of information can explain the longer response times. In the low condition, as I advanced in Chapter 3, the longer response times can also be interpreted in terms of conflict. Both likelihoods being low, they might be in conflict with the task goal, since it is to choose the chest with the highest likelihood to contain gold coins. The task goal might prime participants to reject low likelihoods. But in the low condition this would lead participants to end up with no options to choose. Finally in the negative condition, the cost of negation is a potential interpretation for the longer response times. For example, Gough (1966) showed that negative sentences are verified more slowly than affirmative ones. This explanation is not incompatible with an explanation in terms of conflict as well. Both verbal probabilities have a negative directionality, which under a pragmatic interpretation of directionality, could be puzzling for a participant. It would suggest that the speaker recommends neither of the options, since s/he is using a negative directionality, while the task instructions indicate that in each pair there is always a correct chest. Similarly to the low condition, the use of both negative directionality could generate a conflict with the task goal itself.

As I suggested in the introduction of the present chapter, further support for the pragmatic account of directionality would be that directionality is used differently according to a speaker's intentions. If directionality is a pragmatic cue to what a speaker wants you to choose, one could read this cue as what a malevolent speaker does not want you choose. Therefore one could still use directionality as a reversed cue. I have shown in Experiment 1b that children used

directionality differently when the speaker is malevolent than when the speaker is benevolent. However, the expected reversed pattern of judgements was not found. Rather children seemed to make probability judgements at chance level. This could suggest that a malevolent speaker is placed on the same level as an ignorant or incompetent speaker, and that his/her input is simply dismissed. In a study with children, Nurmsoo and Robinson (2009) found that the information from speakers who made mistakes previously, for reasons out of their control, is later taken into account. However, the information from speakers who made mistakes previously, without a good reason, is discarded seemingly because those speakers are judged incompetent. It remains that in Experiment 1b, it could not be excluded that processing together directionality and the intention of the speaker was too costly to be fully done by children.

Experiment 3b replaced the time manipulation with a manipulation of the speaker's benevolence or malevolence, in an adult's population. Participants were informed that either Peter Pan or Captain Hook would give them information, and that Captain Hook had unhelpful intentions. Therefore Peter Pan was set up as the benevolent speaker and Captain Hook as the malevolent one. It was expected that in the benevolent case the pattern of accuracy, preference for the positive directionality and response times would be similar to those found in Experiment 2a, the limited time condition of Experiment 2b and both conditions of Experiment 3a. That is, accuracy would be higher in the positive and congruent conditions, participants would prefer positive verbal probabilities in the high and low condition (but to a less extent in the latter), and response times would be longer in the incongruent, negative and low conditions.

However in the malevolent case, it was expected that accuracy would be low in the congruent conditions, as adults would use directionality as a reversed cue. Therefore they should choose the least likely option (conveyed with a negative directionality in the case of a congruent verbal probability). This would support both the pragmatic case in general and the suggestion that children did not use directionality as a reversed cue because it was too costly. Conversely adults' performance in the incongruent condition would benefit from this use of directionality as a reversed cue. In this condition, the positive verbal probability is the one with a low likelihood, while the high likelihood is expressed by a negative verbal probability. Therefore if a listener chose the negative verbal probability because he thinks that the speaker is malevolent, he would in fact chose the verbal probability with a high likelihood. Higher accuracy than in the case of a benevolent speaker was thus expected in the incongruent condition.

In the positive and the negative conditions, directionality could not be used as a reversed cue, and participants would choose the least likely option. In a pragmatic account, participants who think that the malevolent speaker is trying to mislead them should try to use what the speaker says to make their decisions. But as they could not use directionality, the next best option would be to consider that the speaker describes as having a low likelihood what has a high likelihood, and vice-versa. Therefore participants could use the likelihood also as a cue to reverse their interpretation. Therefore lower accuracy was also expected in those conditions. In the high and low conditions, a reversal of the preference for positive verbal probabilities was expected, participants choosing the negative one more often. However it could be that children

in Experiment 2b behaved like adults instead of lacking cognitive resources. In that case, the malevolent speaker would be simply ignored, considered as incompetent. In that case, the performance and choices in the malevolent condition would not be expected to be different from chance. Finally, response times were expected to be generally longer, as the processing of the speaker's intention would add one more step (reversing the cue) in the processing of the information. However, if the malevolent speaker was deemed ignorant and ignored, response times should not be different than in the benevolent speaker condition.

4.3. Experiment 3b

4.3.1. Method

4.3.1.1. Participants

Nineteen Psychology students (mean age= 19.16; range= 18-21; 18 women) at university of Birmingham (UK) took part in this experiment in exchange for course credit.

4.3.1.2. Materials

The materials were the same as in Experiment 2a, 2b and 3a. Similarly to Experiment 3a, the verbal probabilities were presented in speech bubbles, told by the silhouettes on the screen. The only difference with Experiment 3a is that the silhouettes telling the verbal probabilities were either Peter Pan or Captain Hook (see Figures 17 and 18 for examples).

4.3.1.3. Procedure

The procedure was similar to Experiment 3a, with differences as follows. Participants made decisions under two speaker conditions: in half of the trials, Peter Pan spoke both speech bubbles; in the other half, Captain Hook was speaking. In order to reinforce the impression of a

conversation, the silhouettes of one trial's speaker were presented first without speech bubbles, for one second, immediately followed by the silhouettes uttering the two verbal probabilities in the speech bubbles. The first image, without speech bubbles, also mentioned who was speaking ('Hook/Peter says...'), to ensure that participants identified the correct speaker for each trial.

To ensure that participants attributed malevolent intentions to Captain Hook, the following instructions were given to the participants on a first screen:

'You are going to be presented pairs of treasure chests. In every pair only one chest contains some gold coins. Plus choosing the empty one would make you lose the entire game. To help you deciding which chest you should open, either Peter Pan or Captain Hook will give you some information about each chest. Captain Hook secretly wishes he could keep the treasure for himself.'

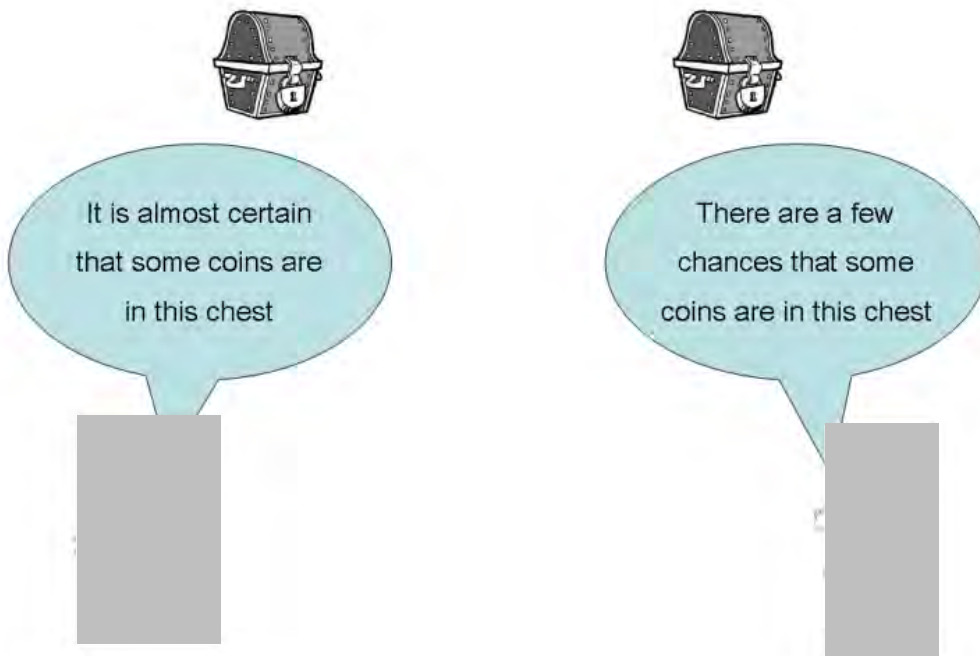


Figure 17: Example of Stimuli Used in Experiment 3b (Malevolent Speaker)

A second screen of instructions read:

‘For each trial it will be first indicated who is going to give you information. Then the pieces of information will be presented under their respective chests. Sometimes the two pieces of information will seem very similar to you but they are actually always different within each pair. For each pair you will have to say if you want to open the left chest (x key) or the right one (m key). You will have only 5 seconds to make your decisions. Before playing the game for real, here are some training pairs. Press the spacebar to continue’.

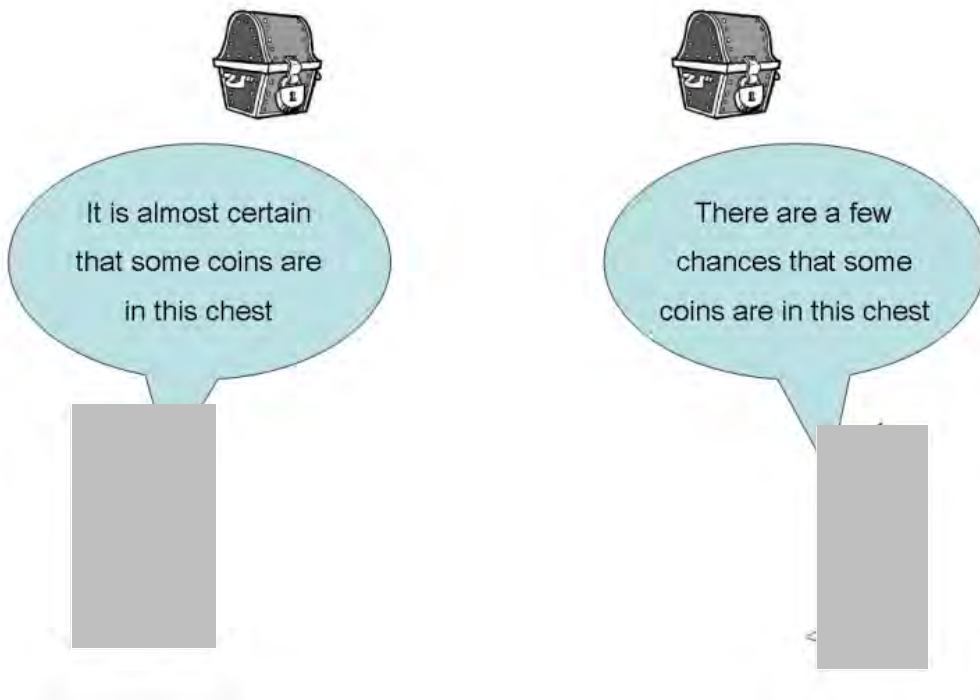


Figure 18: Example of Stimuli Used in Experiment 3b (Benevolent Speaker)

4.3.2. Results

As in Chapter 3 and in Experiment 3a, participants were presented with each verbal probability 36 times, having the opportunity to familiarize themselves with the materials. Hence each analysis was conducted twice, the second time including the experimental order as additional factor to exclude familiarity and/or length of the pairs as confounding variables.

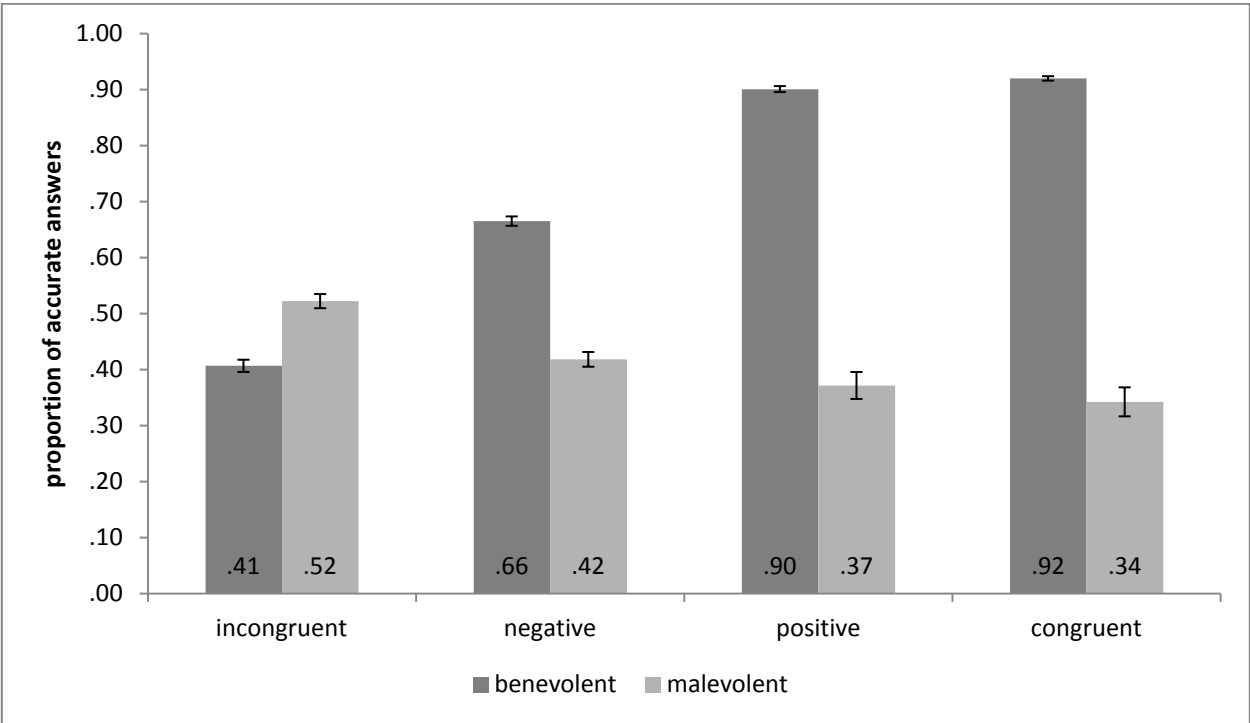
4.3.2.1. Accuracy

In order to test if participants were more accurate when the speaker was benevolent, I compared the performance on the four conditions where a correct answer existed (positive,

negative, incongruent and congruent conditions), under the two speaker conditions. The correct answer was always the option with the higher likelihood, whoever was speaking. As in

Experiment 3a, the accuracy score could range from 0 to 1, 1 representing a normative answer (that is, a preference for the higher likelihood). Mean proportions of accurate answers according to the comparison condition (when accuracy is a relevant measure) and the time condition are presented in Figure 19.

A 4 (comparison condition: incongruent, negative, positive or congruent) by 2 (speaker: benevolent or malevolent) repeated-measures analysis of variance (ANOVA) revealed a main



^a error bars represent standard errors of the mean

Figure 19: Mean Proportion of Accurate Answers as Function of the Comparison Condition and the Speaker (Experiment 3b)

effect of comparison condition on accuracy both for subjects, $F_1(3,54) = 10.73, p = .0041, \eta^2_p = .37$ (lower bound adjustment), and items, $F_2(3,32) = 12.28, p < .001, \eta^2_p = .54$. Planned repeated contrasts applied with Bonferroni correction (with $\alpha = .0166$) indicated that accuracy in the incongruent condition was significantly lower than in the negative condition, $F_1(1,18) = 7.76, p = .012, \eta^2_p = .30$. Accuracy in the negative condition was significantly lower than in the positive condition, $F_1(1,18) = 14.60, p = .001, \eta^2_p = .45$. Accuracy in the positive condition was not different from accuracy in the congruent condition, $F_1(1,18) < 1, p = .672$.

The ANOVA also revealed a main effect of speaker. Accuracy was significantly lower when Captain Hook provided the information than when it was Peter Pan, $F_1(1,18) = 31.21, p < .001, \eta^2_p = .63$, and for items, $F_2(1,32) = 44.47, p < .001, \eta^2_p = .58$.

There was an interaction between comparison condition and speaker both for subjects, $F_1(3,54) = 32.49, p < .001, \eta^2_p = .64$, and for items, $F_2(3,32) = 11.49, p < .001, \eta^2_p = .52$. When Peter Pan was the speaker, the comparison condition had a main effect, $F_1(3,54) = 142.30, p < .001, \eta^2_p = .89$. With Peter Pan, accuracy in the incongruent condition was significantly lower than in the negative condition, $F_1(1,18) = 149.77, p < .001, \eta^2_p = .89$. Accuracy in the negative condition was significantly lower than in the positive condition, $F_1(1,18) = 63.74, p < .001, \eta^2_p = .78$. Accuracy in the positive condition and the congruent condition were similar, $F_1(1,18) = 1.23, p = .282$. When Captain Hook was the speaker, the comparison condition had no effect, $F_1(3,54) = 2.59, p = .125$ (lower bound adjustment).

A second analysis was conducted taking into account the experimental order. The main effect of comparison condition and the interaction of comparison condition and speaker

remained for subjects and for items (lowest $F = 7.04$, highest $p = .004$). The main effect of speaker remained the same for items, but was marginal for subjects (lowest $F = 3.87$, highest $p = .065$). There was an effect of the experimental order for subjects, $F_1(1,18) = 31.45$, $p < .001$, $\eta^2_p = .64$, but not for items, $F_2(1,32) < 1$, $p = .459$. Participants were more accurate in the second half than in the first half.

There was an interaction between comparison condition and experimental order for subjects, $F_1(1,18) = 28.99$, $p < .001$, $\eta^2_p = .62$, but not for items, $F_2(3,32) < 1$, $p = .818$. The effect of the comparison condition was significant in the first half, $F_1(3,54) = 12.48$, $p = .002$, $\eta^2_p = .41$ (lower bound adjustment), as well as in the second half, $F_1(3,54) = 7.67$, $p = .013$, $\eta^2_p = .30$ (lower bound adjustment; Bonferroni correction; $\alpha = .025$). There was no difference in accuracy between the first and the second halves in the incongruent condition, the negative condition, the positive condition, or the congruent condition, $t(18) = 0.19$, $p = .849$, $t(18) = 0.28$, $p = .784$ and $t(18) = 0.78$, $p = .447$, $t(18) = 1.90$, $p = .073$, $r = .06$ respectively (Bonferroni correction; $\alpha = .0125$).

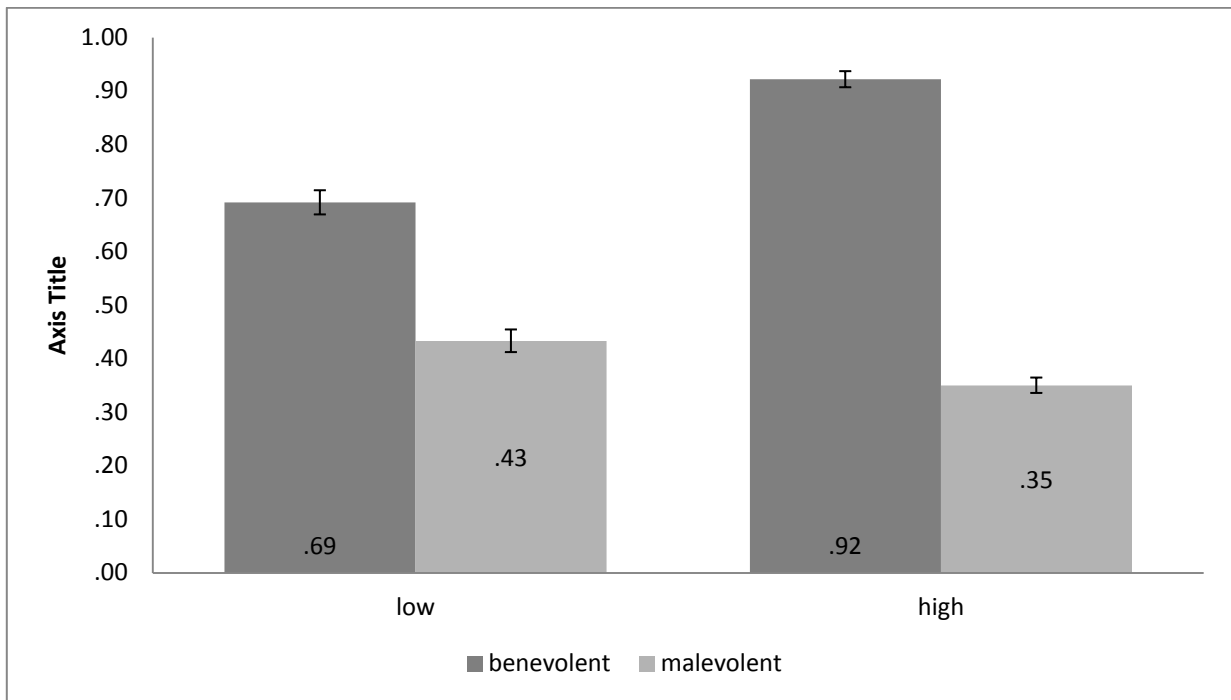
There was also an interaction between experimental order and speaker for subjects, $F_1(3,54) = 37.91$, $p < .001$, $\eta^2_p = .68$, but not for items, $F_2(1,32) < 1$, $p = .333$. There was no difference in accuracy between the first and the second half, whether when Peter Pan was the speaker, $t(18) = 0.01$, $p = .992$, or when Captain Hook was, $t(18) = 1.55$, $p = .138$. Participants were more accurate with Peter Pan as speaker than with Captain Hook, both in the first, $t(18) = 5.66$, $p < .001$, $r = .63$, and the second half, $t(18) = 5.40$, $p < .001$, $r = .62$.

Finally, there was an interaction between speaker, experimental order and comparison condition, for subjects, $F_1(3,54) = 17.04$, $p < .001$, $\eta^2_p = .49$, but not for items, $F_2(3,32) = 1.95$, $p = .141$. The interaction between speaker and comparison condition was significant in the first half of the task, $F_1(3,54) = 27.00$, $p < .001$, $\eta^2_p = .60$, as well as in the second half, $F_1(3,54) = 31.53$, $p < .001$, $\eta^2_p = .64$. There was no interaction between experimental order and comparison condition when the speaker was malevolent, $F_1(3,54) = 1.84$, $p = .151$, or when he was benevolent, $F_1(3,54) < 1$, $p < .766$. Finally there was no interaction between experimental order and speaker in any of the comparison conditions (highest $F = 2.05$, lowest $p = .170$).

4.3.2.2. Directionality

I compared the choices made under the two speaker conditions and under the two conditions with no correct answer (low and high conditions) in order to check if participants' preference for the option presented with the positive verbal probability could be reduced by a speaker is deemed malevolent. As in Experiment 3a, participants received 0 for each negative expression and 1 for each positive expression chosen, and these were averaged to give a score between 0 and 1. Mean proportions of choices of the positive verbal probabilities in the high and low conditions are presented in Figure 20.

A 2 (comparison condition: low or high) by 2 (speaker condition: benevolent or malevolent) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on the preference for positive verbal probabilities: participants chose the outcome presented with a positive verbal probability significantly more often under the high condition



^a error bars represent standard errors of the mean

Figure 20: Mean Proportion of Choices in Favour of the Positive Verbal Probability (Standard Deviation) as Function of the Comparison Condition and the Speaker Condition (Experiment 3b)

than under the low condition for subjects, $F_1(1,18) = 5.99, p = .025, \eta^2_p = .25$, but not for items, $F_2(1,16) = 2.77, p = .115, \eta^2_p = .15$.

The ANOVA also revealed a main effect of speaker: participants chose the outcome presented with a positive verbal probability significantly more often when Peter Pan was speaking than when Captain Hook was, $F_1(1,18) = 37.13, p < .001, \eta^2_p = .67$, and for items, $F_2(1,16) = 42.13, p < .001, \eta^2_p = .73$.

There was an interaction between comparison condition and speaker both for subjects, $F_1(1,18) = 28.11, p < .001, \eta^2_p = .61$, and for items, $F_2(1,16) = 6.20, p = .024, \eta^2_p = .28$. When

Peter Pan was speaking, participants chose the outcome described by a positive verbal probability more often in the high condition than in the low condition, $t(18) = 10.93, p < .001, r = .82$. This was not the case when Captain Hook was speaking, $t(18) = 1.49, p = .153$.

One-sample t -tests indicated that when Peter Pan was the speaker, participants chose the outcome described by a positive verbal probability significantly more often than chance, both in the low and high conditions, $t(18) = 9.73, p < .001, r = .75$ and $t(18) = 24.79, p < .001, r = .94$ respectively; when Captain Hook was the speaker, participants chose more often the option described by a negative verbal probability. However they did not do this significantly more than they would be expected to by chance, $t(18) = -1.88, p = .076$, in the low condition, and $t(18) = 1.75, p = .097$, in the high condition.

A second analysis was conducted taking into account the experimental order. The main effect of comparison condition remained the same, that is significant for subjects but not for items (lowest $F = 2.77$, highest $p = .115$). The main effect of speaker and the interaction of comparison condition and speaker remained for subjects and for items (lowest $F = 6.20$, highest $p = .024$). There was no effect of experimental order both for subjects, $F_1(1,18) < 1, p = .839$, or for items, $F_2(1,16) < 1, p = .503$. There was no interaction between speaker and experimental order, for subjects, $F_1(1,18) = 2.59, p = .125$, but there was one for items, $F_2(1,16) = 8.44, p = .010$. Participants chose more often the option described by a positive verbal probability when Peter Pan was speaking than when Captain Hook was speaking, both in the first half, $t(17) = 5.20, p < .001, r = .70$, and in the second half, $t(17) = 6.05, p < .001, r = .77$ (Bonferroni correction; $\alpha = .025$). There was no difference between the first and the second halves, when

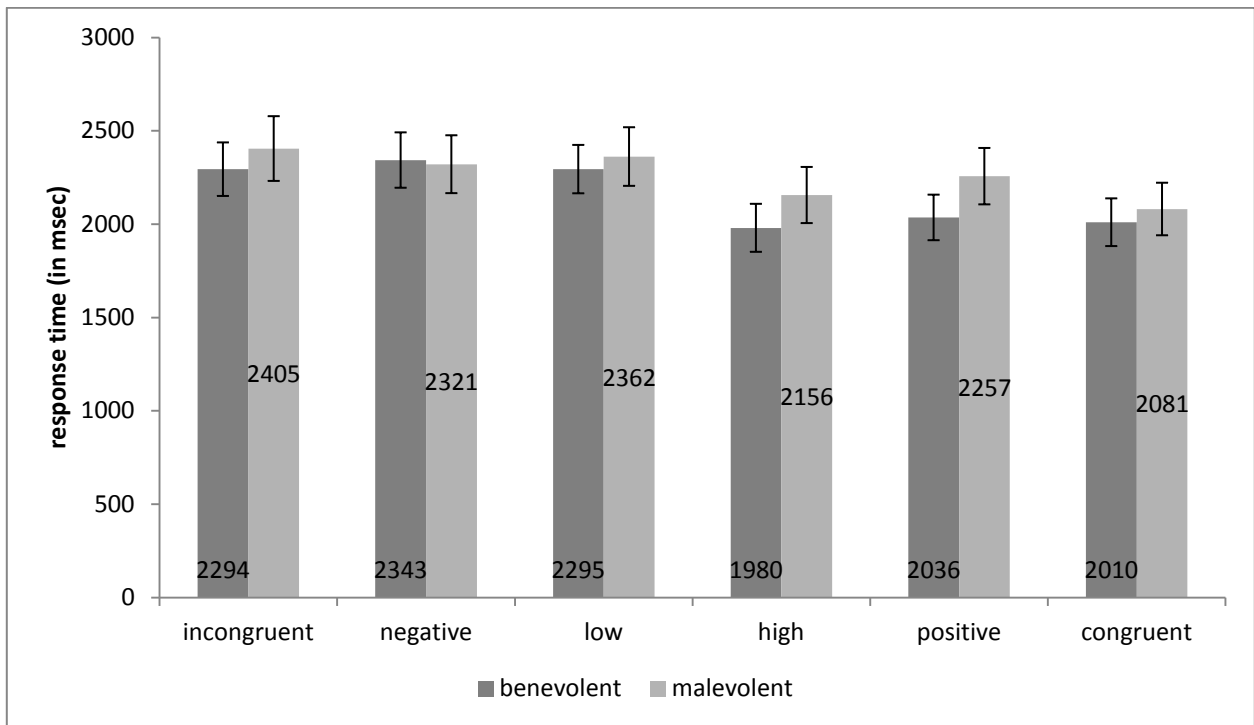
Peter Pan was the speaker, $t(17) = 1.08, p = .297$. However, when Captain Hook was the speaker, participants chose the option described by a positive verbal probability less often in the first half than in the second half, $t(17) = 2.77, p = .013, r = .23$ (Bonferroni correction; $\alpha = .025$).

Finally, there was no interaction between experimental order and comparison condition both for subjects or for items, nor between time condition, experimental order and comparison condition, for subjects or for items (all $F < 1$, lowest $p = .617$).

4.3.2.3. Response time

In order to investigate if making decisions that take into account speakers' intentions was accompanied by an increase in response time, I compared response times under the two speaker conditions and under each comparison condition (low, high, positive, negative, congruent and incongruent). I used the medians of each participant in each condition. Means of median response times according to the comparison conditions are presented Figure 21.

A 6 (comparison condition: incongruent, low, negative, high, positive or congruent) by 2 (speaker condition: benevolent or malevolent) repeated-measures analysis of variance (ANOVA) revealed a main effect of comparison condition on the response time both for subjects, $F_1(5,90) = 16.31, p < .001, \eta^2_p = .48$, and for items, $F_2(5,48) = 6.27, p < .001, \eta^2_p = .40$. Planned repeated contrasts (order of conditions: incongruent, negative, low, positive, high, congruent) found only the low and positive conditions to be significantly different, $F_1(1,18) = 16.27, p = .001, \eta^2_p = .48$. Therefore I ran a post-hoc special contrast which confirmed that response time was longer under the low, incongruent and negative conditions, compared to the high,



^a error bars represent standard errors of the mean

Figure 21: Mean Response Time (in msec.) as Function of the Comparison Condition and the Speaker Condition (Experiment 3b)

positive and congruent conditions (see Table 6 for F and p values).

The ANOVA also revealed a main effect of speaker. Response time was significantly longer when the speaker was Captain Hook than when it was Peter Pan, both for subjects, $F_1(1,18) = 5.48, p = .031, \eta^2_p = .23$, and for items, $F_2(1,48) = 8.22, p = .006, \eta^2_p = .15$. The ANOVA found no interaction between comparison condition and speaker, both for subjects, $F_1(5,90) = 2.11, p = .072, \eta^2_p = .11$, or for items, $F_2(5,48) = 1.43, p = .229, \eta^2_p = .13$.

Table 6

Post-Hoc Special Contrasts on the Response Time (Experiment 3b)

compared conditions	F values	p values	η^2_p
low vs. high	29.47	< .001 ^a	.62
low vs. positive	16.27	= .001 ^a	.48
low vs. congruent	46.05	< .001 ^a	.72
incongruent vs. high	29.75	< .001 ^a	.62
incongruent vs. positive	10.58	= .004 ^a	.37
incongruent vs. congruent	28.13	< .001 ^a	.61
negative vs. high	22.16	< .001 ^a	.55
negative vs. positive	8.25	= .010	.31
negative vs. congruent	31.46	< .001 ^a	.64

^a significant after Bonferroni correction (for 9 tests, $\alpha = .0056$)

A second analysis was conducted taking into account the experimental order. The main effect of comparison condition and of speaker remained for subjects and for items (lowest $F = 6.27$, highest $p = .006$), and the interaction between comparison condition and speaker remained not significant for subjects and for items (highest $F = 3.65$, lowest $p = .072$). There was a main effect of experimental order both for subjects, $F_1(1,18) = 8.36$, $p = .010$, $\eta^2_p = .32$, and for items, $F_2(1,48) = 293.37$, $p < .001$, $\eta^2_p = .86$.

There was no interaction between speaker and experimental order, for subjects or for items, no interaction between experimental order and comparison condition, for subjects or for items, and no interaction between speaker, experimental order and comparison condition, for subjects or for items (highest $F = 1.69$, lowest $p = .154$).

Additional analyses run using only successful trials yielded similar results. I found a main effect of comparison condition comparing only the four conditions where successful trials could be distinguished from failed ones, $F_1(3,48) = 4.03$, $p = .012$, $\eta^2_p = .20$. I found a marginal effect of speaker, $F_1(1,16) = 4.25$, $p = .056$, $\eta^2_p = .21$. Finally, the interaction between comparison condition and speaker was not significant, $F_1(3,48) = 1.93$, $p = .108$.

4.3.2.4. Individual analyses

During the data analyses, observation of the distributions of accuracy and directionality revealed two patterns of answers. In the congruent condition, some participants ($N = 6$) made as many accurate decisions when the speaker was benevolent ($Md = 34$) and when the speaker was malevolent ($Md = 34$), $T = 3$, $z = -0.76$, $p = .450$. The other participants ($N = 13$) made more accurate decisions when the speaker was benevolent ($Md = 33$) than when he was malevolent ($Md = 3$), $T = 0$, $z = -3.19$, $p = .001$.

I therefore compared the accuracy of each group of participants under each condition, as well as their preference for the positive directionality. The group who did not display differences of accuracy in the congruent condition, between the two speakers, did not display differences either in the other conditions: in the positive condition, $T = 12$, $z = -1.23$, $p = .216$ (for benevolent speaker, $Md = 32$; for malevolent speaker, $Md = 33.5$); in the negative condition, $T = 5$, $z = -1.15$, $p = .249$ (for benevolent speaker, $Md = 22$; for malevolent speaker, $Md = 24$); in the incongruent condition, $T = 14.5$, $z = -0.85$, $p = .398$ (for benevolent speaker, $Md = 13$; for malevolent speaker, $Md = 14$).

The group who displayed differences of accuracy in the congruent condition, between the two speakers, also displayed differences of accuracy in the positive condition, $T = 0$, $z = -3.19$, $p = .001$ (for benevolent speaker, $Md = 34$; for malevolent speaker, $Md = 4$), and in the negative condition, $T = 1$, $z = -3.11$, $p = .002$ (for benevolent speaker, $Md = 22$; for malevolent speaker, $Md = 9.5$). Their accuracy was however not different in the incongruent condition, $T = 16.5$, $z = -2.03$, $p = .043$ (for benevolent speaker, $Md = 14$; for malevolent speaker, $Md = 21$; Bonferroni correction, $\alpha = .0125$).

The group who did not display differences of accuracy in the congruent condition, between the two speakers, did not display differences either in their preference for the positive directionality: in the low condition, $T = 4.5$, $z = -1.26$, $p = .207$ (for benevolent speaker, $Md = 23$; for malevolent speaker, $Md = 23$); in the high condition, $T = 6$, $z = -0.41$, $p = .686$ (for benevolent speaker, $Md = 31$; for malevolent speaker, $Md = 31$). The group who displayed differences of accuracy in the congruent condition, between the two speakers, also displayed differences either in their preference for the positive directionality: in the low condition, $T = 0$, $z = -3.18$, $p = .001$ (for benevolent speaker, $Md = 25$; for malevolent speaker, $Md = 13$); in the high condition, $T = 0$, $z = -3.18$, $p = .001$ (for benevolent speaker, $Md = 35$; for malevolent speaker, $Md = 3$).

4.3.3. Discussion

This experiment drew on previous experiments to test the pragmatic account of directionality further. As expected, comparing and choosing between two verbal probabilities spoken by a benevolent speaker led to similar patterns of accuracy, preference for positive

directionality and response times as I have found before. Crucially when the speaker was malevolent, those patterns all differed. The accuracy in the incongruent and negative conditions was better in the malevolent speaker condition than in the benevolent speaker condition. In the positive and congruent conditions, the accuracy was lower in the malevolent speaker condition than in the benevolent speaker condition. The preference for positive directionality, in the benevolent speaker condition, shifted to a preference for negative directionality in the malevolent speaker condition (low, high, incongruent and congruent conditions). Response times were consistently higher in the malevolent speaker condition than in the benevolent speaker condition. This suggested that it is costly to integrate the intention of a speaker with the information provided by a verbal probability.

This reversed pattern of accuracy when the speaker was malevolent was stronger when considering only the participants who showed an effect of the speaker in the congruent condition. However even within this group, the reverse pattern was only trending in the incongruent condition. This is however consistent with Chapter 3 and Experiment 3a: in the incongruent condition, the directionality and the likelihood are giving opposite cues regarding what to choose, and it seems that a consequence of this is a generally lower performance. In fact, the lower performance in that condition, as I suggested in Chapter 3, highlights that people seem not to treat only the directionality, but to also take into account the likelihood (at least implicitly). Therefore when the malevolent character speaks, participants could decide to use either the directionality or the likelihood as reversed cue (as they did in the positive and negative condition).

The picture formed by this set of changes brought by a malevolent speaker is in line with a pragmatic account of directionality. I suggest that directionality is more than just a feature reflecting the likelihood of a verbal probability, as suggested by Budescu et al. (2003). Directionality allows speakers to convey their intention, and is interpreted as such. Adults' decisions reflected an integration of both directionality and what is known of the speaker's intentions. Children in Experiment 1b dismissed directionality when the speaker was malevolent, instead of reversing it. However, this could also reflect consideration of the intention and directionality, directionality being used or not depending on the speaker's intentions. Presumably considering both fully in order to reverse directionality entails higher demands on cognitive resources. This is reflected in the longer response times observed in the malevolent condition, and could explain why children only partially considered both.

It is noteworthy, however, that about a third of participants did not display different patterns of answers between the benevolent and the malevolent condition. It is possible that the manipulation of the intention simply failed with this subset of participants. However, it might also be that these participants thought of a double-bluff, attributing to the malevolent speaker the intention to trick by making them believe s/he was misleading. The paradigm used here did not allow us to distinguish these two possibilities, but if it were true the latter possibility may suggest individual differences in the use of directionality. That is, higher cognitive resources and/or greater theory of mind abilities could be linked to a greater sensitivity and finer use of this feature of verbal probabilities. As theory of mind is linked to

pragmatic abilities (e.g., Happé, 1993), such a possibility would further support the pragmatic account of verbal probabilities.

4.4. General discussion

In this chapter, I set out to further develop the pragmatic account of directionality. In Chapter 3, I showed that the framing effect of directionality can be overcome when the verbal probabilities are of similar likelihood, if the time pressure is relieved. However in this chapter, I showed that this was not the case when the decision was made in a more conversational context (Experiment 3a). I suggested that directionality is a pragmatic feature of verbal probabilities that listeners interpret as a cue from speakers to signal what they want listeners to choose or what they know to be the correct answer.

In Experiment 3b, I gained further support for this account by manipulating the intention of the speaker. When the speaker was benevolent the results displayed similar patterns to Chapter 3 and Experiment 3a. That is, participants preferred the option described by the positive verbal probability and performed less well when choosing between two negative verbal probabilities or between two incongruent ones. When the speaker was malevolent however, they prefer the option described by the negative verbal probability. They also performed better when comparing two negative verbal probabilities or two incongruent ones, while performing less well when comparing two congruent verbal probabilities or two positive ones.

This suggests that as a feature of verbal probabilities, directionality indicates to listeners what a speaker wants them to do, or know of the correct answer. Note that it could be that listeners make a pragmatic interpretation that is different to the speakers' pragmatic intention.

In both the relevance account of pragmatics (e.g., Wilson & Sperber, 2004), as well as in the Gricean account (Grice, 1975), speakers are required to communicate what they think as true as much as possible. According to the relevance account, this is so that they do not waste listeners' cognitive resources. On that perspective, listeners would behave as if speakers chose their language purposely. They would therefore have to attend to the choices the speaker made. Therefore if a speaker presents two different options with a positive verbal probability for one and a negative verbal probability for the other, the listener would infer that this choice is intended and may deduce that this is because the speaker wants him to choose the option described by the positive verbal probability.

This would explain why choices can be framed by positive verbal probabilities (e.g., Teigen & Brun, 1999; Gourdon & Villejoubert, 2009; this thesis, Experiment 2a). The directionality induces a change in the listener's perspective (e.g., Sanford & Moxey 2003), framing the decision-making. But more importantly the listener interprets it based on other pieces of information, for example social cues. This is supported by my results in Experiment 3b, where changing the intention of the speaker changed the frame of the decision-making, suggesting that listeners reinterpreted the frame in light of speakers' intention. This suggestion is further supported by an increase in response times when responding to the malevolent speaker.

The reinterpretation of directionality by listeners also supports the recent account of epistemic vigilance (Mascaro & Sperber, 2009). In this perspective, human beings always evaluate the competency of speakers, forecasters, etc, and reinterpret their statements accordingly if needed. For example, when a speaker is introduced as being malevolent, children

did the opposite of what the speaker recommended (Mascaro & Sperber, 2009). In this thesis, in Experiment 3b, adult participants reinterpreted verbal probabilities when a known malevolent speaker provided them information. Similarly to children in Mascaro and Sperber, they chose the opposite of what the speaker seemed to suggest them to choose, developing a preference for options described by a negative verbal probability.

Research into how children use information from unreliable speakers suggests that children ignore information from incompetent speakers (e.g., Nurmsoo & Robinson, 2009). However, they use the information in a different way when the speaker is deemed malevolent (Mascaro & Sperber, 2009). In Experiment 3b, I showed that adults also use the information provided in verbal probabilities in a different way when confronted by a malevolent speaker. But little is known about what adults would do if the speaker was judged simply incompetent. Based on children's results, they may dismiss the information and make decisions randomly. To test this directly, one could draw from Koenig and Harris' (2005) testimony paradigm. In this paradigm, participants watch the speaker make mistakes, or not, and build their own representation of the speaker's competence. This could be adapted to intention, with a speaker who states inaccurately what is in a box, even though he just had a look in the box (which excludes that he does not know), suggesting that he is purposely wrong. Another speaker would be presented as simply incompetent, giving incorrect information because, for example, he cannot see in the box as it is too far. Finally, there would also be a benevolent/competent speaker as a control. If incompetence and malevolence are treated differently by adults, the directionality of the verbal probabilities should lead to a preference for the positive in the benevolent condition, and

preference for the negative in the malevolent condition. In the incompetent condition, the directionality should be ignored.

In this chapter I have shown that in a conversational context, participants display a preference for the positive directionality even if they are given unlimited time to choose between two options. I have also shown that the directionality can be used in a reverse way when a speaker is deemed malevolent. That is, a preference for a negative directionality is displayed instead of the preference for a positive one. I suggest that this supports a pragmatic account of directionality, where to listeners, directionality is a cue of what speakers want them to do, or know they should do.

CHAPTER 5: SHOULD THE MESSENGER BE SHOT? RESPONSIBILITY JUDGEMENTS OF SPEAKERS USING VERBAL AND NUMERICAL PROBABILITIES.

5.1. Introduction

On October 22nd, 2012, an Italian tribunal condemned a team of seismologists for manslaughter, after a series of events that had been described both as bad scientific communication (e.g., Ropeik, 2012, October 22) and normal scientific uncertainty (e.g., Leshner, 2010, June 29). On April 6th, 2009, an earthquake at L'Aquila (Abruzzo, Italy) killed almost 300 people, only six days after the *Commissione Grandi Rischi* (Large Risks Committee; own translation) had met to discuss the seismic situation of the region. Without the trial transcripts, it is unclear which words were most held against the scientists, but overall they were accused of understating the risk and over-reassuring the population. The minutes from the Commissione Grandi Rischi (March 31st, 2009) cite one of the scientists saying that '[it was] improbable that there would be, in the short-term, an earthquake like the 1703 one [in the Apennine]' (*improbabile che ci sia a breve una scossa come quella del 1703*; own translation). *Improbable* is a verbal probability representing a low risk, but it does not deny all risk. Although to my knowledge there are no published studies of Italian verbal probabilities, English-speaking populations estimate *improbable* as expressing as much as 30% (e.g., Budescu et al., 2003). But one important feature of *improbable* (or *improbabile* in that case) is its negative directionality. This negative directionality would be expected to focus the receiver on the non-occurrence of the uncertain event (see e.g., Teigen & Brun, 1988). In this chapter, I set out to explore this

phenomenon, that is, how different modes of risk communication, including different directionalities, influence how responsibility is attributed to speakers predicting uncertain events.

According to the so-called communication mode preference paradox (Erev & Cohen, 1990), people prefer being told about the probability of an uncertain event through a numerical format (e.g., a percentage). However, when they have to communicate such a probability to someone, people prefer to use a verbal probability (e.g., *There is a chance*). Asked to justify these two preferences, people mentioned the higher level of precision of numerical probabilities on one hand, and the fact that verbal probabilities are more natural to express on the other (Wallsten, Budescu, Zwick & Kemp, 1993). These reasons were, however, produced explicitly and could have resulted from a post-hoc rationalization.

One alternative account that could also explain the communication mode preference paradox is that people do not want to engage too much in predicting an outcome (Beyth-Marom, 1982). Beyth-Marom argued that the interindividual variability (e.g., Hamm, 1999) in the interpretation of verbal probabilities allows one to avoid having one's predictions judged on their quality. She gave the example that when judged on the quality of predictions, one can still reply 'I said it is possible' (p. 258). That is, one can avoid being held responsible if the uncertain event does not ultimately occur. Such an account is consistent with the Communication Mode Preference paradox: it is easier to hold speakers responsible if they gave precise information that turned out to be wrong, since the predictions can be compared against the outcomes. If listeners received a vague statement, they could, as just said, blame themselves for

misunderstanding the statement. But listeners could also blame themselves simply for having made a decision relying only on vague information.

In this chapter I aimed to test such an explanation of the preference paradox indirectly, by testing whether people judge speakers as more responsible when they use numerical probabilities than when they use verbal probabilities. To my knowledge only Dieckmann investigated the influence of the communication mode upon the responsibility judgements in his PhD dissertation (2007). Dieckmann gave participants a scenario where an intelligence forecast had been delivered a few weeks before a terrorist attack. The intelligence report used three different probability formats: externally framed numerical probability (e.g., *The probability that the event will occur is X%*), internally framed numerical probability (e.g., *We are X% sure...*) and externally framed numerical probability with a confidence interval. Dieckmann also used four different likelihood levels, from 0 to 10%. Participants judged the credibility of the source, the usefulness of the prediction, and how much should the forecaster be blamed ('How much blame do you think should be placed on the analysts that produced the intelligence report?').

Dieckmann (2007) found no difference in attributions of blame between the different modes. However in his scenarios, events had only a low likelihood of occurring and he used only numerical probabilities. Also the judgements were framed as judgements of blame, which impedes de facto holding someone responsible for something positive. I will improve on this design, testing the influence of the communication mode on judgements of responsibility at different levels of likelihood (high or low) and varying the nature of the outcome (the uncertain

event occurred or did not). As outcomes will be negative and positive, I will ask about responsibility rather than blame.

Furthermore, a number of recent studies found the so-called Knobe effect (or side-effect effect, Knobe, 2003), where judgements of intentionality are influenced by the valence of a final outcome. In the original Knobe effect, ultimately negative consequences lead to higher judgements of blame (e.g., Pellizzoni, Siegal & Surian, 2009). Therefore I also manipulated the valence of the uncertain events, using both positive ones (e.g., securing a place for one's first choice university) and negative ones (e.g., getting a permanent scar). This allowed trying to distinguish whether responsibility judgements are based on the consequences or on the accuracy of the prediction. If they are based on the consequences, a negative event ultimately occurring or a positive event ultimately not occurring should yield higher responsibility judgements. If responsibility judgements are based on the accuracy of the prediction, they should be higher for predictions of low likelihood for events which ultimately do occur or predictions of high likelihood for events which ultimately do not occur.

As highlighted in the L'Aquila earthquake case, it was also important to distinguish between positive and negative verbal probabilities. The scientists then used a negative verbal probability. Directionality is thought to focus the attention of a listener on the occurrence (positive) or the non-occurrence (negative) of an uncertain event (see e.g., Moxey and Sanford, 2000). In L'Aquila it would have focused the public on the non-occurrence of an earthquake. If the scientists had said that *it was a very small chance*, they would have focused the public's attention on the occurrence of the earthquake. The public might not have felt that they were

telling them that it would not happen. Therefore, communicating about a risk using a negative verbal probability can be considered as sending a pragmatic cue that the event will not occur. In Chapter 2 and Chapter 4, I already showed that directionality can communicate pragmatic information (e.g., the speaker's intention), and that this information is taken into account. In this chapter I will continue to manipulate the directionality of verbal probabilities. That is, I will present numerical probabilities, positive verbal probabilities and negative verbal probabilities.

In this chapter, I will therefore manipulate the format of communication, to distinguish between numerical and verbal probabilities, and positive and negative verbal probabilities. I will also manipulate the outcome and the valence of the uncertain event, so that the valence of the consequences could be manipulated (determined by the interaction of the outcome and the valence of event). Finally, I will also manipulate the likelihood, which in interaction with the outcome will allow manipulating the accuracy.

In Experiment 4a, it was expected that the final consequences, rather than the accuracy of the prediction, would influence the responsibility judgements, following the Knobe effect (Knobe, 2003). That is, the interaction between the event valence and the outcome will influence responsibility judgements, rather than the interaction between the likelihood and the outcome. Therefore, responsibility judgements should be higher when negative events occurred or when positive events did not occur.

Following the suggestion of Beyth-Marom (1982) that using verbal probabilities mean that speakers can undermine their responsibility, higher responsibility judgements were expected when the format of communication was numerical probabilities. Finally, as positive and

negative verbal probabilities focus the attention on opposite outcomes, it was also expected that the pattern of responsibility judgement would be reversed between positive and negative verbal probabilities. In Experiments 4b to 4e, I will build up on the results of Experiment 4a to consider the potential effects of using precise or round numerical probabilities, both on judgements of responsibility and on the willingness to recommend the speaker.

5.2. Experiment 4a

5.2.1. Method

5.2.1.1. Participants

Fifty undergraduate students enrolled in different courses (mean age = 21.42; range = 18-38; 42 women) at University of Birmingham (UK) took part in this experiment in exchange for course credit or cash (I did not collect data regarding the number of paid participants). Eighteen participants read statements based on negative verbal probabilities, 16 read statements based on positive verbal probabilities and 16 read statements based on numerical probabilities.

5.2.1.2. Materials

I implemented the chance expressions in eight scenarios (see Appendix V, p. 215) relating to student life. Although I did not pilot test the suitability of those scenarios beforehand, I requested feedback at the end of the questionnaire, where participants indicated that they found the scenarios easy to relate to. An example scenario read as follows: 'When he went to Keele University for visit day, Samuel asked a student about the admission rates. The student at the visit day told him that [verbal or numerical probability] more applicants will be admitted this year'. The scenarios were written so that their uncertain event had one of two possible valences

(positive, e.g., being admitted at university; negative, e.g., having a permanent scar after a mole removal).

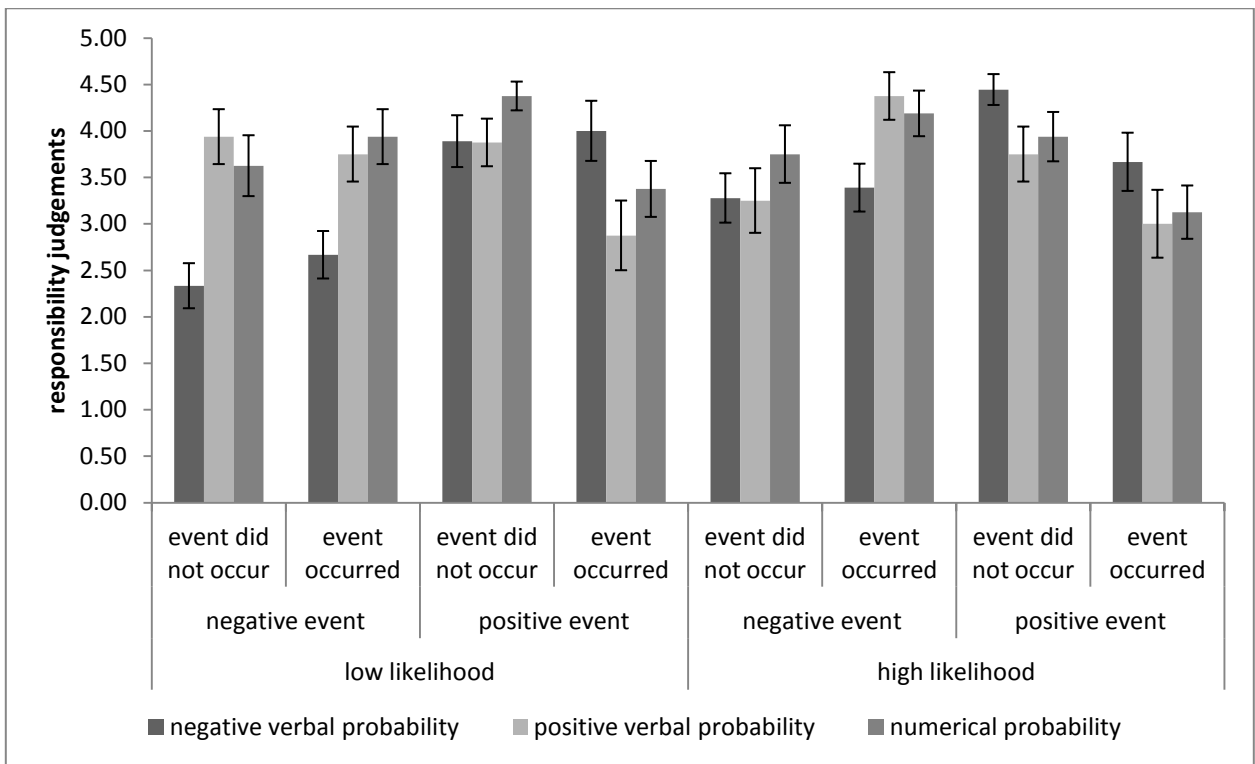
Two levels of likelihood each with two versions (low, i.e. 20 or 40% vs. high, i.e. 60 or 80%) were combined with two possible outcomes (occurrence or non-occurrence). The four conditions thus obtained were counterbalanced across the eight scenarios in order to constitute a factorial design with the event valence (positive or negative) to give 2 levels of likelihood x 2 possible outcomes x 2 event valences. Therefore each scenario was presented with different outcomes and likelihood. The chance expressions were manipulated between participants: a third of them read the scenarios with percentages, another third read them with positive verbal probabilities and the last third read them with negative verbal probabilities. Verbal probabilities were chosen based on a previous study where Psychology students at the University of Birmingham translated verbal probabilities in percentages (Gourdon & Beck, unpublished data), their directionality being determined in a similar manner to Chapters 3 and 4.

5.2.1.3. Procedure

The questionnaire was presented through Medialab[®] software. Each scenario was presented individually on the screen. When participants finished reading it, they clicked to go on and made responsibility judgements on to a scale (5 points, from 'Not at all responsible' to 'Very responsible'). The order of presentation of the scenarios was randomized. After rating responsibility on each of the eight scenarios, participants answered simple demographic questions and had the opportunity to give feedback.

5.2.2. Results

In order to test if speakers were held more responsible when using numerical probabilities and for predicting final negative consequences or for being inaccurate, I compared the responsibility judgements in eight conditions, under three different communication modes (numerical probability, positive verbal probability or negative verbal probability). Responsibility judgements could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Figure 226.



^a error bars represent standard errors of the mean

Figure 22: Mean Responsibility Judgement as Function of the Communication Mode, the Likelihood the Event was Predicted with, the Valence of the Event, and the Outcome (Experiment 4a)

A 2 (likelihood: low or high) by 2 (event valence: negative or positive) by 2 (outcome: non-occurrence or occurrence) by 3 (communication mode: numerical, positive verbal or negative verbal) mixed analysis of variance (ANOVA) found no main effect of communication mode on responsibility judgements, $F(2,47) = 2.30, p = .112$, nor of likelihood, $F(1,47) = 1.12, p = .295$, event valence, $F(1,47) = 1.46, p = .233$, or outcome, $F(1,47) = 2.22, p = .143$. There was also no interaction between likelihood and communication mode, $F(2,47) = 2.22, p = .120$, between outcome and communication mode, $F(2,47) < 1, p = .749$, and between likelihood and outcome, $F(1,47) < 1, p = .573$.

There was an interaction between event valence and communication mode, $F(2,47) = 14.47, p < .001, \eta^2_p = .38$. Post-hoc ANOVAs indicated that there was a main effect of communication mode when the uncertain event was negative, $F(2,47) = 17.13, p < .001, \eta^2_p = .42$. Speakers were judged more responsible for predicting a negative event with a numerical, $p < .001$, or a positive verbal probability, $p < .001$, than with a negative verbal probability, but similarly when using a numerical or a positive verbal probability, $p > .999$ (all three pairwise comparisons applied with Bonferroni correction of the p -value). There was no main effect of communication mode when speakers predicted a positive event, $F(2,47) = 3.25, p = .048$ (Bonferroni correction; $\alpha = .025$).

There was also an interaction between likelihood and valence, $F(1,47) = 4.11, p = .048, \eta^2_p = .08$. Post-hoc paired t -tests applied with Bonferroni correction ($\alpha = .025$) indicated that speakers were not judged more responsible for predicting that positive events had a low likelihood than for predicting that negative events had a low likelihood, $t(49) = 2.04, p = .047$. The valence of the event did not matter either when the prediction was of a high likelihood,

$t(49) = 0.11, p = .911$. There was also no difference between predictions of events having a low likelihood and predictions of events having a high likelihood, both for negative events, $t(49) = 1.91, p = .063$, and for positive events, $t(49) = 0.55, p = .586$ (paired t -tests applied with Bonferroni correction; $\alpha = .025$).

There was a final two-way interaction between event valence and outcome, $F(1,47) = 16.58, p < .001, \eta^2_p = .26$. Post-hoc paired t -tests applied with Bonferroni correction ($\alpha = .025$) indicated that speakers were judged more responsible when positive events did not occur than when negative events did not occur, $t(49) = 4.72, p < .001, r = .43$, while the valence of the event did not matter when the event occurred, $t(49) = 1.34, p = .186$. Speakers were also judged more responsible when positive events did not occur than when they occurred, $t(49) = 3.70, p = .001, r = .33$, but not when negative events occurred rather than not, $t(49) = 2.19, p = .033$ (paired t -tests applied with Bonferroni correction; $\alpha = .025$)

There was a marginal interaction between likelihood, outcome and communication mode, $F(2,47) = 3.04, p = .058, \eta^2_p = .11$. Further ANOVAs indicated that there was no interaction between likelihood and outcome when the communication mode was a negative verbal probability, $F(1,17) = 2.34, p = .229$, when it was a positive verbal probability, $F(1,15) = 2.69, p = .122$, or when it was a numerical probability, $F(1, 15) < 1, p = .620$ (Bonferroni correction; $\alpha = .0167$). When the outcome did not occur, there was no interaction between communication mode and likelihood, $F(2,47) = 3.27, p = .047$, or when the outcome occurred, $F(2,47) < 1, p = .463$ (Bonferroni correction; $\alpha = .0167$). Finally, there was no interaction between outcome

and communication mode when the likelihood was low, $F(2,47) = 2.33, p = .109$, or when it was high, $F(2,47) < 1, p = .414$.

There was no three-way interaction between likelihood, event valence and communication mode, $F(2,47) = 1.27, p = .291$, between event valence, outcome and communication mode, $F(2,47) < 1, p = .382$, or between likelihood, event valence and outcome, $F(1,47) = 2.06, p = .158$. Finally, there was no interaction between likelihood, event valence, outcome and communication mode, $F(2,47) < 1, p = .503$.

5.2.3. Discussion

In Experiment 4a, I asked participants to judge the responsibility of speakers giving uncertain statements about scenarios of daily life. The uncertain events were of positive or negative valence, and they did or did not occur ultimately. Speakers were judged more responsible for predicting a positive event that did not occur, that is when there were negative consequences. However, that was not the case for negative consequences emerging from a negative event that did occur. This is in part similar to the general Knobe effect (Knobe, 2003), where negative consequences are judged as more intentional than positive ones. The fact that this effect is only showing partially might be enlightened by looking at the other effects observed in this experiment.

When the events had a positive valence, speakers were judged more responsible for predicting them with a low likelihood, or for predicting them with a negative directionality. However, this did not interact further with the outcome. This is surprising, as a negative directionality or a low likelihood should focus people's attention on the non-occurrence of the

outcome. Therefore, the possibility that the event will not happen should be more available to listeners, who may be expected to hold speakers more responsible if the event in fact occurred (as this occurrence will contrast with the focus of attention). However, this assumes that responsibility judgements are driven by speakers' accuracy. This was not supported by the results here. I found no interaction between the likelihood and the outcome.

An alternative explanation would be that directionality, and to some extent the likelihood, are taken as pragmatic cue of speakers' intentions, as already observed in Chapters 2 and 4. In the pragmatic perspective, using a negative verbal probability to predict a negative event may signal that the speaker does not want it to happen to the receiver, whereas predicting a positive event with a negative phrase seems to signal a lack of desire to see the receiver enjoy positive consequences. This is supported by the interaction between communication mode and event valence found in Experiment 4a. For negative events, responsibility judgements were higher when a numerical probability or a positive verbal probability was used, and lower when a negative verbal probability was used. For positive events, responsibility judgements were similar between the communication modes. In Experiment 3b, I also found that likelihood was sometimes used in a pragmatic manner. It was apparently used as a reversed cue when there was no difference in directionality and when the speaker was malevolent. That is, if the speaker was malevolent and the decision could only be made based on likelihood, decision makers chose more often the event described as having a low likelihood than the one with a high likelihood. Similarly, it seems that using a low likelihood is to some extent (i.e. only when

predicting positive events) perceived as signalling the speaker's wish that the receiver does not enjoy this positive event.

Finally, it is noteworthy that all communication mode effects were due to negative verbal probabilities eliciting different judgements. This is not consistent with an account of the Communication Mode Preference paradox based on vagueness leading to a 'diffusion' of the responsibility. According to this account I would expect responsibility based on both positive and negative verbal probabilities to differ from percentages. In Experiment 4a, positive verbal probabilities and numerical probabilities led to essentially similar judgements. However, in Experiment 4a, I used only round numerical verbal probabilities (e.g., *there is an 80% chance, there is a 20% chance*), which might seem too artificial and/or might be too vague to elicit strong responsibility inferences. In the same way that speakers could say that they '*only said it was possible*' after using verbal probabilities (Beyth-Marom, 1982, p.258), speakers could argue that they only say *that it was around 80%*. It is also possible that listeners assume that round percentages are the result of speakers averaging different estimations s/he is aware of. While when using a precise percentage, speakers might sound as they are directly quoting one of those estimates. Experiment 4b aimed to test the possibility that round percentages can be perceived as vaguer than precise ones, and therefore still elicit lower responsibility judgements. This was done by simply manipulating the preciseness of numerical probabilities, still in conjunction with the level of likelihood and the outcome.

5.3. Experiment 4b

5.3.1. Method

5.3.1.1. Participants

Sixteen Psychology students from a Birmingham (UK) high school (mean age = 16.94; range = 16-18; 12 girls) took part in this experiment as part of their psychology curriculum. Half the participants (20) read statements based on round percentages, and the other half (20) read statements based on precise percentages.

5.3.1.2. Materials and procedure

I implemented the numerical probabilities in eight scenarios (see Appendix VI, p. 217) relating to student life. All scenarios related to a positive uncertain event. The eight scenarios included the four scenarios which presented a positive event in Experiment 4a, as well as one of the negative event scenarios, but modified. Three new scenarios were therefore created to replace the negative event scenarios that could not plausibly be modified to be positive.

Two levels of likelihood (low, i.e. 20 or 40% vs. high, i.e. 60 or 80%) were combined with two possible outcomes (occurrence or non-occurrence). The four conditions thus obtained were counterbalanced across the eight scenarios. The numerical probabilities were manipulated between participants: half of them read the scenarios with round percentages (e.g., 60%), and the other half read them with precise percentages (e.g., 63%). The procedure was as in Experiment 4a.

5.3.2. Results

In order to test if speakers were held more responsible when using precise numerical probabilities, I compared the responsibility judgements in four conditions, under two different communication modes (precise or round numerical probability). Responsibility judgements could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 7.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) mixed analysis of variance (ANOVA) found no main effect of communication mode on responsibility judgements, $F(1,14) < 1, p = .858$, nor of likelihood, $F(1,14) < 1, p = .603$. There was a main effect of outcome, $F(1,14) = 6.77, p = .021$,

Table 7

Mean Responsibility Judgement (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4b)

		precise percentages	round percentages
low likelihood	event did not occur	3.25 (0.93)	3.88 (0.79)
	event occurred	3.50 (1.07)	2.75 (1.16)
high likelihood	event did not occur	3.19 (1.39)	4.25 (0.60)
	event occurred	3.06 (1.18)	2.38 (1.09)

$\eta^2_p = .33$; speakers were judged more responsible when the uncertain event did not occur than when it occurred.

There was no interaction between likelihood and communication mode, $F(1,14) < 1$, $p = .603$, or between likelihood and outcome, $F(1,14) = 2.54$, $p = .133$. There was an interaction between outcome and communication mode, $F(1,14) = 8.00$, $p = .013$, $\eta^2_p = .36$. Post-hoc paired t -tests applied with Bonferroni correction ($\alpha = .025$) indicated that when using a round percentage, speakers were judged more responsible when an event did not occur than when it occurred, $t(7) = 3.82$, $p = .007$, $r = .82$. This was not the case when using precise percentages, $t(7) = 0.16$, $p = .877$. Speakers who used a precise percentage were not judged as more responsible than speakers who used a round one when the event occurred, $t(14) = 1.51$, $p = .155$, or when it did not occur, $t(14) = 2.10$, $p = .054$ (independent t -tests with Bonferroni correction; $\alpha = .025$). Finally there was no three-way interaction between likelihood, outcome and communication mode, $F(1,14) < 1$, $p = .603$.

5.3.3. Discussion

In Experiment 4b, I aimed to test if Experiment 4a failed to find a difference between verbal and numerical probabilities because the percentages used were too vague. Although Experiment 4b was underpowered, it suggests that round percentages led to differential judgements, still displaying a Knobe effect (e.g., Knobe, 2003) where the non-occurrence of a positive event (i.e., a negative consequence) led to higher responsibility judgements. However, precise percentages seemed to lead to similar judgements whatever the outcome or the likelihood, and more importantly did not overall differ from round percentages. This suggested

that more precise predictions may not lead to infer more responsibility to speakers than predictions using only a round percentage. However it could also be that as the percentages used here were precise, they seemed unrealistic to participants. It might seem odd that one might say *There is a 63% chance that...* For example, Witteman, Zimund-Fisher, Waters, Gavaruzzi and Fagerlin (2011) showed that receivers of numerical probabilities find percentages with decimal places less credible than integer-based percentages.

Therefore Experiment 4c aimed to explore this. Participants were asked to what extent they would recommend the speakers for advice based on their predictions. If precise percentages are less credible (Witteman et al., 2011), speakers using them should be perceived as less trustworthy. Therefore, they should be less recommended as forecasters. The Knobe effect (Knobe, 2003) leads to more responsibility being attributed when consequences are negative. As they are judged more responsible, speakers should be less recommended. That is, following the Knobe effect, when consequences are negative, speakers should be less likely to be recommended.

5.4. Experiment 4c

5.4.1. Method

5.4.1.1. Participants

Forty Psychology students (mean age = 19.33; range = 18-33; 34 women) at University of Birmingham (UK) took part in this experiment in exchange for course credits. Half the participants (20) read statements based on round percentages, and the other half (20) read statements based on precise percentages.

5.4.1.2. Materials and procedure

The materials were identical to Experiment 4b, with the addition of a supplementary scale. Participants had to imagine that somebody else asked the main character of the scenario to recommend the forecaster (e.g., 'Imagine somebody else wanted to ask Peter's uncle for information. How likely would Peter be to recommend asking him?'). Participants then indicated on a 5 point-scale how likely they would be to recommend the speaker to a friend, from 'Not likely at all' to 'Very likely'.

The questionnaire was implemented through the Kwik Survey[®] software, so it could be filled online by participants. As it was not possible to randomize the order of presentation of the scenarios with Kwik Survey, counterbalancing was used to choose the order of display. The order of presentation of the responsibility and recommendation scales was also counterbalanced. The procedure was otherwise similar to Experiments 4a and 4b.

5.4.2. Results

5.4.2.1. Responsibility judgements

As in Experiment 4b, I tested whether speakers were held more responsible when using precise numerical probabilities by comparing the responsibility judgements in four conditions, under two different communication modes (precise or round numerical probability). Responsibility judgements could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 8.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) mixed analysis of variance (ANOVA) found a main

Table 8

Mean Responsibility Judgement (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4c)

		precise percentages	round percentages
low likelihood	event did not occur	1.90 (0.70)	2.48 (0.95)
	event occurred	3.00 (0.95)	3.10 (0.79)
high likelihood	event did not occur	2.50 (0.84)	2.95 (1.04)
	event occurred	3.08 (0.99)	3.45 (0.76)

effect of communication mode on responsibility judgements, $F(1,38) = 4.80$, $p = .035$, $\eta^2_p = .11$.

Speakers were judged more responsible when using round percentages than when using precise ones, in contrast with Experiment 4b (where the communication mode had no effect).

There was also a main effect of likelihood, $F(1,38) = 7.45$, $p = .010$, $\eta^2_p = .16$; responsibility judgements were higher when the uncertain event was predicted to have a high likelihood than when it was predicted with a low one. Finally, there was also a main effect of outcome, $F(1,38) = 25.89$, $p < .001$, $\eta^2_p = .41$; speakers were judged more responsible when the uncertain event occurred than when it did not occur.

There was no interaction between likelihood and communication mode, $F(1,38) < 1$, $p = .786$, between likelihood and outcome, $F(1,38) = 2.37$, $p = .132$, or between outcome and

communication mode, $F(1,38) < 1, p = .324$. Finally there was no three-way interaction between likelihood, outcome and communication mode, $F(1,38) < 1, p = .349$.

5.4.2.2. Likelihood of recommending

In order to test if the precise percentages seemed plausible to the participants, I tested whether participants would recommend the speakers. I compared the likelihood of recommending to a friend in four conditions, under two different communication modes (precise or round numerical probability). The likelihood of recommending could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 9.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) mixed analysis of variance (ANOVA) found no main effect of communication mode on likelihood of recommending, $F(1,38) < 1, p = .397$, nor of likelihood, $F(1,38) = 2.58, p = .116$. However, there was a main effect of outcome, $F(1,38) = 267.45, p < .001, \eta^2_p = .88$; participants were more likely to recommend the speakers when the uncertain event occurred than when it did not occur.

There was no interaction between likelihood and communication mode, $F(1,38) < 1, p > .999$, or between communication mode and outcome, $F(1,38) = 1.97, p = .168$. There was an interaction between outcome and likelihood, $F(1,38) = 7.09, p = .011, \eta^2_p = .16$. Post-hoc paired *t*-tests applied with Bonferroni correction ($\alpha = .025$) indicated that when the event did not occur, speakers were more likely to be recommended if they had predicted a low likelihood than if they had predicted a high one, $t(39) = 2.80, p = .008, r = .31$. This was not the case when

Table 9

Mean Likelihood of recommending (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4c)

		precise percentages	round percentages
low likelihood	event did not occur	2.35 (0.88)	2.50 (0.93)
	event occurred	4.13 (0.56)	4.18 (0.75)
high likelihood	event did not occur	2.05 (0.71)	1.75 (0.73)
	event occurred	4.00 (0.90)	4.50 (0.69)

the uncertain event occurred, $t(39) = 0.60, p = .555$. Finally there was no three-ways interaction between likelihood, outcome and communication mode, $F(1,38) = 3.67, p = .063$.

5.4.3. Discussion

In the Experiment 4b, I tested if round percentages are perceived as a vague mode of risk communication, by comparing round percentages and precise ones (without decimals). While round percentages displayed an effect of the outcome in the manner of the Knobe effect, precise percentages showed no effect of either the likelihood or the outcome. This raised the possibility that precise percentages were simply dismissed because they seemed unrealistic. Thus in Experiment 4c, I also measured how likely speakers were to be recommended, next to responsibility judgements.

The results from 4c were rather different to 4b: the communication mode had an effect, that is using more precise percentages leading to speakers being judged as less responsible. This contrasts with my earlier suggestion that using more vague predictions diffuses the responsibility. Instead it suggested that the more precise, the less responsible, as if people judged speakers based on their effort. This was not apparent in Experiment 4b. This may indicate that answering the recommendation scale led participants to consider speakers' performance in more depth. The 'more precise, less responsible' finding is also consistent with the likelihood of recommending findings. There was an interaction between the likelihood and the outcome found in the likelihood of recommending: in the case of events not occurring, the likelihood of recommending was higher if the speaker had predicted a low likelihood, that is, if the speaker was accurate.

One could argue that speakers may have been perceived as simply quoting percentages they would have read or heard previously, which would have lessen their responsibility. However in that case, the recommendation judgements would not be expected to reflect some accuracy of the predictions, since they would not bear any responsibility in the forecasting. Although it is possible also that they are recommended for being able to quote accurate forecast, that is for choosing sources of information which are of quality.

The likelihood of recommending was higher when the uncertain event had occurred, that is, when consequences were positive since all events were positive. This is consistent with the Knobe effect, where blame judgements are higher for negative consequences. If speakers are less to blame for positive consequences, they may be expected to be recommended more.

However, the responsibility judgements reflected the opposite of a Knobe effect, responsibility judgements being higher when the positive event occurred or when it was predicted with a high likelihood. This contrasts with Experiments 4a and 4b, where responsibility judgements were higher when the consequences were negative. One possibility is that being asked about recommendation shifted the meaning of responsibility towards a more positive one. That is, in Experiments 4a and 4b, responsibility could have been interpreted as blame. But by asking about recommendation as well, I might have brought the focus on trustworthiness. That is I may have asked participants to evaluate speakers according to a positive characteristic instead of a negative one (blame) in Experiments 4a and 4b.

Experiment 4d aimed to test for this by asking about responsibility and about recommendation to different participants. As I suspected that the effect of the communication mode found in Experiment 4c resulted from having to judge recommendations as well as responsibility, it was needed to measure the same type of judgements in a between design.

5.5. Experiment 4d

5.5.1. Method

5.5.1.1. Participants

Eighty Psychology students (mean age = 19.61; range = 18-39; 72 women) at University of Birmingham (UK) took part in this experiment in exchange for course credits. Half the participants (40) made responsibility judgements, and the other half (40) expressed how likely they were to recommend speakers. In each of those two groups, half the participants (20) read

statements based on round percentages, and the other half (20) read statements based on precise percentages.

5.5.1.2. Materials and procedure

The materials used in this Experiment were identical to the materials in Experiment 4c.

The procedure was as in Experiment 4c, with the only difference that the two different scales were administered between participants.

5.5.2. Results

5.5.2.1. Responsibility judgements

As in Experiments 4b and 4c, I tested whether speakers were held more responsible when using precise numerical probabilities by comparing the responsibility judgements in four conditions, under two different communication modes (precise or round numerical probability). Responsibility judgements could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 10.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) mixed analysis of variance (ANOVA) found no main effect of communication mode on responsibility judgements, $F(1,38) < 1, p = .415$. There was a main effect of likelihood, $F(1,38) = 24.37, p = .010, \eta^2_p = .39$; responsibility judgements were higher when the uncertain event was predicted with a high likelihood than when it was

Table 10

Mean Responsibility Judgement (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4d)

		precise percentages	round percentages
low likelihood	event did not occur	2.10 (0.75)	2.15 (0.83)
	event occurred	2.55 (0.71)	2.28 (0.75)
high likelihood	event did not occur	2.63 (0.90)	2.63 (0.79)
	event occurred	3.23 (0.85)	2.90 (0.91)

predicted with a low one. Finally, there was also a main effect of outcome, $F(1,38) = 11.21$, $p = .002$, $\eta^2_p = .23$; speakers were judged more responsible when the uncertain event occurred than when it did not occur.

There was no interaction between likelihood and communication mode, $F(1,38) < 1$, $p = .831$, between likelihood and outcome, $F(1,38) < 1$, $p = .519$, or between outcome and communication mode, $F(1,38) = 2.25$, $p = .142$. Finally, there was no three-way interaction between likelihood, outcome and communication mode, $F(1,38) < 1$, $p > .999$.

5.5.2.2. Likelihood of recommending

In order to test if the precise percentages seemed plausible, I tested whether participants would recommend the speakers by comparing the likelihood of recommending to a friend in

four conditions, under two different communication modes (precise or round numerical likelihood). The likelihood of recommending could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 11.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) mixed analysis of variance (ANOVA) found no main effect of communication mode on the likelihood of recommending, $F(1,38) < 1$, $p = .715$. However there was a main effect of likelihood, $F(1,38) = 21.94$, $p < .001$, $\eta^2_p = .37$; speakers were more likely to be recommended when they predicted a high likelihood than when they predicted a low one. There was also a main effect of outcome, $F(1,38) = 89.60$, $p < .001$, $\eta^2_p = .70$; participants were more likely to recommend the speakers when the uncertain event occurred than when it did not occur.

There was no interaction between likelihood and communication mode, $F(1,38) < 1$, $p = .552$, or between likelihood and outcome, $F(1,38) < 1$, $p = .545$. There was an interaction between outcome and communication mode, $F(1,38) = 4.27$, $p = .046$, $\eta^2_p = .10$. Post-hoc paired t -tests applied with Bonferroni correction ($\alpha = .025$) indicated that when speakers used round percentages, they were more likely to be recommended if the event ultimately occurred than if it did not, $t(19) = 8.55$, $p < .001$, $r = .82$. This was also the case when speakers used precise percentages, although with a smaller effect, $t(19) = 5.01$, $p < .001$, $r = .55$. Finally there

Table 11

Mean Likelihood of Recommending (Standard Deviation) as Function of the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4d)

		precise percentages	round percentages
low likelihood	event did not occur	2.10 (0.75)	2.15 (0.83)
	event occurred	2.55 (0.71)	2.28 (0.75)
high likelihood	event did not occur	2.63 (0.90)	2.63 (0.79)
	event occurred	3.23 (0.85)	2.90 (0.91)

was no three-way interaction between the likelihood, the outcome and the communication mode, $F(1,38) < 1, p = .545$.

5.5.3. Discussion

In Experiment 4d, participants either judged the responsibility of speakers giving uncertain statements about scenarios in daily life, or judged the likelihood that they would recommend these speakers. As in Experiment 4c, the likelihood of recommending showed results consistent with the Knobe effect. That is, negative consequences (outcome not occurring) led to lower likelihood to be recommended. I suggest that this is because negative consequences lead to higher responsibility judgements.

Speakers were also more likely to be recommended when they predicted a high likelihood for the uncertain event. Crucially, the effect of the outcome did not interact with the predicted likelihood, showing an indifference to the accuracy in the recommendations. As in Experiment 4a, it seems that listeners interpreted that speakers wanted the best outcome for them, or rather what is perceived as wishful thinking, matters more than accurately predicting when evaluating speakers. If accuracy mattered, I should have found higher likelihood of recommending when an event predicted with a high likelihood did not occur, or an event predicted with a low likelihood did occur. It seems instead that predicting a high likelihood is perceived as speakers signalling that they wish the event (always positive) will happen.

Also similarly to Experiment 4c, the responsibility judgements displayed an opposite pattern to what the Knobe effect would predict: speakers were judged as more responsible when the positive event occurred or when it was predicted with a high likelihood. I suggested after Experiment 4c that being asked about recommendation might shift the meaning of responsibility towards a more positive one, but here this possibility could be excluded, as participants only gave responsibility judgements. It is noteworthy, however, that throughout this chapter, the consequences have always had an effect on responsibility judgements, but not in a consistent direction. In Experiments 4a and 4b, responsibility judgements were higher when a positive event did not occur. In Experiments 4c and 4b, responsibility judgements were higher when a positive event occurred. One possibility is therefore that the question used in Experiments 4a to 4d was too vague to elicit consistent judgements. Responsibility is defined as 'the state or fact of being accountable or to blame for something' (*Oxford Online Dictionary*,

2012). Accountability is a neutral concept likely to encompass responsibility for positive and negative events. But blame is a negative concept, likely to encompass only responsibility for negative events. It is possible therefore that in Experiments 4a and 4b, participants treated the responsibility question as a blame one, but in Experiments 4c and 4d, they treated it as an accountability one.

Experiment 4e was thus set up to compare judgements of responsibility as tested in Experiments 4a to 4d and judgements of responsibility in answer to causality-framed questions. That is participants were asked how much they thought speakers are responsible for the consequence, or how much they think that speaker's predictions were a cause of the consequences. Causality judgements were used as a proxy for accountability. This would allow me to exclude the possibility that the responsibility question used so far elicit causality judgements. If the responsibility-framed question did not also elicit judgements of causality, participants' responsibility answers to the responsibility-framed questions should be different from their answers to the causality-framed questions.

5.6. Experiment 4e

5.6.1. Method

5.6.1.1. Participants

Twenty-one Psychology students from a Birmingham (UK) high school (mean age = 16.89; range = 16-18; 17 girls, 1 undeclared gender) took part in this experiment as part of their psychology curriculum. Eleven participants read statements based on precise percentages, and 10 read statements based on round percentages. In the group which read precise percentages,

five read the causality-framed question, and six read the responsibility-framed question. In the group which read round percentages, half of participants (five) read the causality-framed question, and the other half (five) read the responsibility-framed question.

5.6.1.2. Materials and procedure

The materials used in this Experiment were similar to the materials in Experiment 4c, except that half the participants judged to what extent the characters of each scenario attributed causality to the speaker's prediction (e.g., *How much does Ewan think that his friend's advice is a cause of him winning his bet?*). The other half of the participants judged responsibility by answering to the same question as in the previous experiments of this chapter. The procedure was as in Experiment 4c. Participants answered the causality-framed or the responsibility-framed question, followed by the recommendation question.

5.6.2. Results

5.6.2.1. Responsibility judgements

I tested whether speakers were held more responsible when using precise numerical probabilities by comparing the responsibility judgements in four conditions, under two different communication modes (precise or round numerical probability) and when judging responsibility by answering two different questions (responsibility-framed or causality-framed). Responsibility judgements could range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 12.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) by 2 (question frame: responsibility or causality)

mixed analysis of variance (ANOVA) found no main effect of communication mode on responsibility judgements, $F(1,17) < 1, p = .808$. There was a main effect of question frame, $F(1,17) = 9.88, p = .006, \eta^2_p = .37$; responsibility judgements were higher when they were given as answer to the responsibility-framed question than when given through the causality-framed question. There was also a main effect of outcome, $F(1,17) = 4.44, p = .050, \eta^2_p = .21$; speakers were judged more responsible when the uncertain event did not occur than when it occurred. Finally, there was also a main effect of likelihood, $F(1,17) = 5.16, p = .036, \eta^2_p = .23$; speakers were judged as more responsible when the uncertain event was predicted with a low likelihood than when it was predicted with a high one.

Table 12

Mean Responsibility Judgement (Standard Deviation) as Function of the Question Frame, the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4e)

		causality-framed question		responsibility-framed question	
		precise percentages	round percentages	precise percentages	round percentages
low likelihood	event did not occur	3.20 (0.57)	2.70 (1.26)	3.83 (0.98)	3.90 (1.14)
	event occurred	2.00 (0.50)	2.90 (0.96)	3.25 (1.13)	2.90 (1.14)
high likelihood	event did not occur	2.90 (1.08)	2.00 (0.79)	2.25 (0.98)	4.00 (1.46)
	event occurred	2.60 (0.85)	1.50 (0.50)	2.50 (1.41)	3.10 (0.42)

There was no interaction between: likelihood and communication mode, $F(1,17) < 1$, $p = .892$; likelihood and question frame, $F(1,17) < 1$, $p = .892$; outcome and communication mode, $F(1,17) < 1$, $p = .850$; outcome and question frame, $F(1,17) < 1$, $p = .824$; outcome and likelihood, $F(1,17) < 1$, $p = .469$; communication mode and question frame, $F(1,17) = 3.77$, $p = .069$. There was no three-way interaction between likelihood, outcome and communication mode, $F(1,17) = 2.32$, $p = .146$, nor between likelihood, outcome and question frame, $F(1,17) < 1$, $p = .638$, or outcome, communication mode and question format, $F(1,17) = 2.09$, $p = .167$.

There was a three-way interaction between likelihood, communication mode and question format, $F(1,17) = 8.89$, $p = .008$, $\eta^2_p = .34$. There was no interaction between communication mode and likelihood when the question was framed in terms of responsibility, $F(1,9) = 3.69$, $p = .087$, or when it was framed in terms of causality, $F(1,8) = 6.78$, $p = .031$ (Bonferroni correction; $\alpha = .025$). There was no interaction between question frame and likelihood when the communication mode was round, $F(1,8) = 3.84$, $p = .086$, or when it was precise, $F(1,9) = 5.14$, $p = .050$ (Bonferroni correction; $\alpha = .025$).

There was no interaction between communication mode and question frame when the uncertain event was predicted with a low likelihood, $F(1,17) < 1$, $p = .628$. However, when it was predicted with a high likelihood, there was a significant interaction between communication mode and question frame, $F(1,17) = 14.68$, $p = .001$, $\eta^2_p = .46$. Round percentages led to higher responsibility judgements than precise percentages when the question was framed in terms of responsibility, for events predicted with a high likelihood, $t(9) = 2.82$, $p = .020$, $r = .69$, but not

when it was framed in terms of causality, $t(8) = 2.64, p = .030$ (Bonferroni correction; $\alpha = .025$). The responsibility-framed question led to higher judgements of responsibility when the communication mode was round percentages, for events predicted with a high likelihood, $t(8) = 3.94, p = .004, r = .81$, but not when the communication mode was precise percentages, $t(9) = 1.08, p = .309$ (Bonferroni correction; $\alpha = .025$). Finally, there was no interaction between likelihood, outcome, communication mode and question frame, $F(1,17) < 1, p = .579$.

5.6.2.2. Likelihood of recommending

I tested whether speakers were more likely to be recommended when using precise numerical probabilities by comparing the responsibility judgements in four conditions, under two different communication modes (precise or round numerical probability) and when judging responsibility by answering two different questions (responsibility-framed or causality-framed). Likelihood of recommending judgements were given on a range from 1 to 5. Mean judgements according to the conditions and the communication modes are presented in Table 13.

A 2 (likelihood: low or high) by 2 (outcome: non-occurrence or occurrence) by 2 (communication mode: round or precise) by 2 (question frame: responsibility or causality) mixed analysis of variance (ANOVA) found a main effect of communication mode on the likelihood of recommending, $F(1,15) = 6.42, p = .023, \eta^2_p = .30$. Speakers were more likely to be recommended if they used round percentages than if they used precise ones. There was also a main effect of outcome, $F(1,15) = 23.61, p < .001, \eta^2_p = .61$; speakers were more likely to be recommended when the uncertain event did not occur than when it occurred. There was no

Table 13

Mean Likelihood of Recommending (Standard Deviation) as Function of the Question Frame, the Likelihood the Event was Predicted to Have, the Outcome and the Communication Mode (Experiment 4e)

		causality-framed question		responsibility-framed question	
		precise percentages	round percentages	precise percentages	round percentages
low likelihood	event did not occur	3.20 (1.10)	4.20 (0.67)	3.75 (0.94)	3.33 (0.76)
	event occurred	1.90 (0.55)	2.30 (0.45)	1.75 (0.82)	2.33 (0.29)
high likelihood	event did not occur	3.50 (1.00)	4.50 (0.35)	3.58 (2.01)	4.33 (0.29)
	event occurred	2.10 (0.65)	1.80 (0.76)	1.83 (1.03)	2.00 (0.00)

main effect of question frame, $F(1,15) < 1$, $p = .649$. Finally, there was no main effect of likelihood, $F(1,15) < 1$, $p = .369$.

There was no interaction between: likelihood and communication mode, $F(1,15) < 1$, $p = .959$; likelihood and question frame, $F(1,15) < 1$, $p = .771$; outcome and communication mode, $F(1,15) < 1$, $p = .624$; outcome and question frame, $F(1,15) < 1$, $p = .943$; outcome and likelihood, $F(1,15) = 3.27$, $p = .091$; communication mode and question frame, $F(1,15) < 1$, $p = .431$. There was no three-way interaction between: likelihood, outcome and question frame, $F(1,15) < 1$, $p = .869$; outcome, communication mode and question format, $F(1,15) < 1$, $p = .446$; likelihood, communication mode and question format, $F(1,15) = 2.31$, $p = .149$. There was a marginal interaction between likelihood, outcome and communication mode, $F(1,15) = 4.34$,

$p = .055$, $\eta^2_p = .22$. Finally, there was no interaction between likelihood, outcome, communication mode and question frame, $F(1,15) < 1$, $p = .433$.

5.6.3. Discussion

In Experiments 4a and 4b, speakers were judged as more responsible when the ultimate consequences were negative, in a consistent manner with the Knobe effect (e.g., Knobe, 2003.) However in Experiments 4c and 4d, participants judged speakers as more responsible after ultimately positive consequences. This was the case when participants were asked to judge the likelihood that speakers would be recommended (Experiment 4c), but also when they only judged responsibility (Experiment 4d). One possibility for this shift was that the responsibility question was vague. Experiment 4e was therefore set up to verify what happens when the responsibility judgements are framed more precisely, in this case in terms of causality.

Similarly to Experiments 4a and 4b, in Experiment 4e judgements of responsibility were higher when consequences were negative (outcome not occurring). Although judgements were higher when the question was framed in terms of responsibility than in terms of causality, there was crucially no interaction between the question frame and the outcome, suggesting that in this experiment, participants judged responsibility and causality in a similar way. However, it may also be that the small sample in this experiment did not allow detecting such an interaction.

Interestingly however, judgements of responsibility in Experiment 4e were higher under the responsibility frame only when speakers used round percentages. This can be related to the results of Experiment 4b, where I suggested that speakers using precise percentages could have

been considered to have done the best they could, and therefore they were judged less responsible. In Experiments 4c and 4d, I found no effect of the preciseness on the likelihood of recommending, and this supported this possibility by excluding that speakers were considered not to be trusted for being too precise.

Participants in Experiment 4e were also judged more responsible for forecasting using low probabilities. This can be linked to my suggestion that speakers are rewarded for displaying that they wish the best outcome for listeners, whether it is showed through using a negative directionality (for negative events; Experiment 4a) or a high likelihood (for positive events; Experiments 4d). It is as if by forecasting with a high likelihood, speakers were signalling to listeners that they wish the event (always positive in Experiment 4e) to happen. However, the likelihood of recommending did not seem to reflect a reward for wishing the best for listeners. Speakers were more likely to be recommended when the predicted outcome did not occur (i.e., when ultimate consequences were negative). This result would rather suggest a reward for accuracy that is not fully efficient (i.e., failing to reward for accuracy in the case of positive consequences). The interaction of communication mode with likelihood and outcome suggested that the former is more likely. When the percentages were round, the likelihood of recommending was higher in case of a low predicted likelihood, which could have been taken as pessimism. It may be that pessimism was taken as cautiousness, therefore leading to a different way of evaluating speakers' trustworthiness. In Experiment 4e this may be further supported by the likelihood of recommending being influenced by the communication mode, with higher ratings when the speakers used round percentages. Using round percentages might be

considered as a sign of cautiousness. Witteman et al. (2011) found that speakers using decimal points in a percentage were less credible than speakers using no decimal points. Thus, speakers using precise percentages might be judged as not cautious enough for giving non-credible predictions. If this is the case, it would also suggest that responsibility judgements are driven by what speakers seem to wish for receivers, while the likelihood of recommending would be judged through a different process.

5.7. General discussion

In this chapter, I tested if vaguer predictions (based on verbal probabilities) elicit higher judgements that speakers are responsible for the ultimate consequences, than using precise statements relying, for example, on numerical probabilities. The results are summarized in Table 14. In Experiment 4a, positive verbal probabilities and numerical probabilities did not elicit different responsibility judgements. Since there was no difference between positive verbal probabilities and numerical probabilities in Experiment 4a, I tested if round percentages are in fact considered as vague too, by comparing responsibility judgements elicited by round and precise percentages. Responsibility judgements in Experiment 4b showed an effect of the preciseness, such that they were influenced by the final outcome only when the percentages were round. Experiment 4b was underpowered, but in Experiment 4c and 4d, the addition of the likelihood of recommending judgements allowed me to verify that predictions based on precise percentages were not simply dismissed because they were judged too precise to be credible (see Witteman et al., 2011). However in Experiment 4c, an effect of the communication mode was also found, supporting the claim that using vaguer statements allows diffusion of

Table 14

Summary of the Main Effects and Interactions Found in Experiments 4a to 4e

	Main effects	Interaction effects
Experiment 4a (responsibility judgements)	none	Event valence x Communication mode Likelihood x Valence Event valence x outcome
Experiment 4b (responsibility judgements)	Outcome	Outcome x Communication mode
Experiment 4c (responsibility judgements)	Outcome Communication mode Likelihood	none
Experiment 4c (likelihood of recommending)	Outcome	Likelihood x Outcome
Experiment 4d (responsibility judgements)	Likelihood Outcome	none
Experiment 4d (likelihood of recommending)	Likelihood Outcome	Outcome x Communication mode
Experiment 4e (responsibility judgements)	Question frame Outcome Likelihood	Likelihood x Communication mode x Question format
Experiment 4e (likelihood of recommending)	Communication mode Outcome	

responsibility. But in Experiment 4e round percentages led to an effect of the question frame (in terms of causality or responsibility), while precise ones showed no such effect. Experiment 4e was underpowered, but this was similar to Experiment 4b (also underpowered), when precise percentages yield no effect of the outcome (occurred or did not occur), but round percentage

did. Overall, it seems more that using precise percentages let speakers off the hook, as they have tried their best.

In this chapter, I also tested if responsibility judgements are influenced by the valence of the consequences (as in the Knobe effect; e.g., Pellizzoni, Siegal & Surian, 2009), or more normatively, by the accuracy of the prediction. The results did not allow us to draw a clear picture about what influences the responsibility judgements. However, accuracy did not influence the responsibility judgements. Throughout this chapter, it was never found that the interaction of the likelihood and outcome mattered in judging speakers' responsibility. That is, responsibility judgements were not higher when speakers predicted a low likelihood for an event that ultimately did occur, or when they predicted a high likelihood for an event that did not occur. However, the valence of the consequences was found to be a consistent factor, but this was not in a consistent direction.

In Experiments 4a, 4b and 4e, speakers were judged as more responsible when the consequences were negative (when a positive event did not occur or a negative one occurred, in Experiment 4a; when the event, always positive, did not occur, in Experiment 4b and 4e). This is in line with the Knobe effect (Knobe, 2003), where responsibility judgements are higher when negative side-effects occur than when positive side-effects occur. However in Experiments 4c and 4d, they were judged as more responsible when the event (always positive) occurred, in contradiction with the Knobe effect. It should also be considered that both Experiments 4b and 4e were underpowered, which suggests that the evidence is stronger for an effect that is opposite to the Knobe effect (Knobe, 2003). That is, responsibility judgements seem to be

higher when the outcome is positive. Although this is surprising given the literature on the Knobe effect, it may be that responsibility was understood in a different meaning.

Different experimental manipulations were made to identify reasons for these different results, but none was conclusive. In Experiment 4e, the judgements were framed either in terms of responsibility or in terms of causality. But I did not find an interaction between the question framing and the valence of the event, which would have indicated that the consideration of the outcome was different depending on the meaning given to responsibility. If one sets aside Experiment 4a (as it was not comparing different type of numerical probabilities), another possibility remains to consider to explain the inconsistency in the effect if the outcome. In Experiments 4b and 4e, when negative outcomes lead to higher responsibility judgements, participants were younger students, at pre-undergraduate level. Conversely in Experiments 4c and 4d, participants were undergraduate students. It is possible, consistently with the suggestion that the meaning of responsibility may vary, that younger participants have a definition of responsibility which is more blame-based. It is also possible that high-school students and undergraduate students perceive differently the severity of the events used in the scenarios. Further research could aim to pilot the scenario beforehand in order to select scenarios which are known to elicit the same severity judgment in the population being tested.

As mentioned before, what participants understood by responsibility remains unclear, and it cannot be excluded that this was one cause of the inconsistent effects in these studies. Actually the simple fact that the valence of the consequences always had an effect (although inconsistent), but the accuracy did not, seems to suggest that participants' concept of

responsibility relies at least partially on intentionality. Further research should aim to clarify this before looking again at the effect of the consequences. A possible manipulation would be to define different meanings of responsibility to participants at the start of the experiment. This would allow investigation of which concept of responsibility leads to higher judgements. For example, Wright and Bengson (2009) suggested that the Knobe effect is driven by the intentionality rather than the valence of the consequences. How direct the causal relationship is could also be manipulated. In the scenarios used in this chapter, forecasters had little control over the outcome. This could also explain why the meaning of responsibility seemed to change over the five experiments. In light of the trial of L'Aquila's seismologists, investigating the particular liability attributed to forecasters seems even more crucial, and refining paradigms to reach an understanding of responsibility judgements is essential.

Within the pragmatic account of verbal probabilities that I developed in the previous chapters, the last pattern of results found in this chapter regarding responsibility is the most relevant. In Experiment 4a, using a negative directionality to predict a negative event led to lower responsibility judgements, suggesting that speakers were rewarded for indicating their wish that this negative event would not happen. This is consistent with the results I discussed in Chapters 2 and 4, where directionality was used by decision makers as a cue of speakers' intention. This conversational account of risk communication was further supported by some of the experiments of this chapter where predicting a high likelihood led to lower responsibility judgements. This is consistent with Experiment 3b where, if directionality could not be used as a cue to the speaker's intention because the speaker was known to be malevolent, decision

makers used the likelihood in a reversed manner, supporting that allegedly more objective features can also be interpreted differently in a conversational context (e.g., Sirota & Juanchich, 2012). This explanation is however not fully satisfactory, as being wishful/optimistic means sometimes being inaccurate, and therefore potentially more misleading.

As for the responses to the recommendation questions, they displayed no effect of the communication mode in Experiments 4c and 4d. In Experiment 4e, the communication mode had an effect on the likelihood of recommending. Speakers using precise percentages were less likely to be recommended than speakers using round percentages. This is consistent with the results of Chapter 5 regarding responsibility judgements. If speakers are not judged more responsible when they use a vague format of communication, it is not surprising that they are also not likely to be recommended. Most importantly, this is in opposition to self-reported data that suggest that people prefer to receive numerical probabilities because they are more precise (Wallsten et al., 1993). Rather this is consistent with the results of Witteman et al. (2011), who found that percentages with decimal points were judged as less credible than percentages without decimal points.

However in Wallsten et al. (1993), the sample of participants was made of different groups. For example, one sub-sample was of naïve undergraduate students, but two others were of postgraduate students. What is more, the postgraduate students were mostly studying in fields where one is more likely to receive uncertain information from experts (i.e., nursing and business). It is therefore possible that these different samples reported preferences based on their experience of receiving uncertain information. Chapter 6 will aim to explore this possibility

by manipulating speakers' expertise and asking participants which of a numerical probability and a verbal probability they would prefer, for each scenario.

CHAPTER 6: RECEIVING UNCERTAIN INFORMATION: PRECISION IS ONLY AS GOOD AS THE SPEAKER'S EXPERTISE

6.1. Introduction

The likelihood of an event occurring can be described in different ways to a decision-maker. The probability can be communicated directly or as a percentage. Both are referred to as numerical probabilities. Or it can be expressed with a verbal probability (e.g., in the case of an 80% chance, one could say *It is very likely*), and thus given without using numbers. While people prefer to give uncertain information verbally, they prefer to receive it numerically (Erev & Cohen, 1990). This phenomenon is referred to as the Communication Mode Preference paradox. Although the paradox is thought to be generalizable across individuals, in Erev and Cohen's research, the uncertain information was only ever given by experts, whereas the recipients of the information were non-experts. Yet, we know that the source's expertise leads to different expectations regarding the strength of the argument communicated (Bohner, Ruder & Erb, 2002). It is possible, therefore that the preference for receiving numerical information is specific to it coming from an expert, rather than a general paradox. In this chapter, I investigated whether this apparent preference is in fact the result of expectations about expert speakers.

To investigate people's preference for giving and receiving verbal and numerical probabilities, Erev and Cohen (1990) asked a group of four expert sports commentators to make predictions of events that might happen during basket-ball games. They translated these

predictions so that each had a verbal and numerical version, i.e. when experts predicted an event numerically, they translated it into a verbal probability, and *vice versa*. Non-expert students were offered both types of prediction and indicated on which prediction they would prefer to bet. Three quarters of the students chose to bet based on a numerical prediction, which Erev and Cohen interpreted as showing that people preferred to receive uncertain information numerically. However, as the students were explicitly informed that the predictions came from experts we do not know whether this is a general preference for a type of information, or something related to our expectations about experts.

Investigating general preferences for receiving uncertain information, Wallsten, Budescu, Zwick and Kemp (1993) further explored Erev and Cohen's claims (1990). They used self-report measures rather than production and decision tasks. Sixty-nine percent of their participants expressed a preference for receiving uncertain information numerically. Erev & Cohen and Wallsten et al. both tested English speaking participants. Xu, Ye and Li (2009) found a similar self-reported preference in a sample of Chinese-speaking undergraduates: 63% to 77% declared that they preferred to receive numerical predictions. To some extent these self-report measures should reassure us that the preference paradox is not restricted to communications with experts. However, it is a concern that in the Wallsten et al. study, only the more experienced students (e.g., students who participated previously in experiments involving verbal predictions) were this likely to show the preference (64 to 82% of various samples). Only 37% of the less experienced undergraduates preferred numerical predictions. Is the preference paradox really as well established as we have thought?

Furthermore, in the studies to date the speaker was either clearly identified as an expert (Erev & Cohen, 1990), or unspecified (Wallsten & al., 1993; Xu & al., 2009). Perhaps in the cases where the nature of the speaker was unspecified, participants inferred that s/he was an expert. In Wallsten et al.'s study (1993), most of the participants who preferred to receive numerical predictions (all but the group of less experienced undergraduates, which showed the weakest preference) were studying topics where they were likely to often interact with experts (e.g., nursing). It remains open to question, therefore, whether a speaker's expertise influences people's preferences regarding receiving uncertain information.

Bohner et al. (2002) manipulated the apparent expertise of communicators through descriptions of their level of education. They found that an expert (i.e., a professor) was expected to develop more valid and more convincing arguments than a non-expert (i.e., a high school student). Verbal probabilities are not as precise as numerical ones, as evidenced by the large interindividual variability in the numerical translations people give of those expressions (e.g., Budescu & Wallsten, 1985). As such, they could be thought to represent weaker arguments. Thus, experts may be expected to use a numerical format when communicating uncertain information, more so than lay speakers. Bohner et al. also observed that an expert using information to a lower standard than the one expected leads the communication to backfire. It is therefore important to confirm how people prefer to receive uncertain information that comes from experts.

I addressed this question by asking participants to choose between verbal probabilities and numerical predictions (percentages) when the speaker was explicitly either an expert (defined

by higher or vocational education) or a lay speaker (a friend with some experience of the situation). Following Bohner et al. (2002), I expected that expert speakers would elicit higher preferences for numerical probabilities than would lay speakers.

6.2. Method

6.2.1. Participants

Forty-three psychology students (mean age = 19.44; range = 18-22; 35 females) at University of Birmingham (UK) participated in this experiment in exchange for course credits.

6.2.2. Materials

Eight scenarios were built where a character was given a prediction regarding an uncertain event. Scenarios were such that either a friend who had some experience with the same event or an expert was giving the prediction. For example, when the uncertain event was a car passing the Ministry of Transport (MOT) safety certification, the prediction was given either by a friend who had the same problem on his own car or by a mechanic.

I chose the verbal probabilities from a set of data collected previously from another group of psychology students from the same university (Gourdon & Beck, under revision). All the verbal probabilities predicted a high likelihood that the event would occur (e.g., *It is almost certain*, usually translated as around 90%). Directionality of verbal probabilities (i.e., their positive or negative linguistic nature) is known to frame decisions (e.g., Teigen & Brun, 1999), therefore I used only positive verbal probabilities. The percentages used in the predictions were determined by the mean numerical translation those previous participants made of the verbal probabilities.

6.2.3. Procedure

Each participant read eight scenarios. Scenarios are presented in Appendix VII (p. 219) where they are coded A to H. In half of the scenarios the prediction was given by an expert, and in the other half it was given by a friend. Four versions of the questionnaire were used: this allowed us to present each scenario with the expert or the friend and also to combine the scenarios into two different sets. Thus, in questionnaire 1 for scenarios A, B, C and D, participants read the expert version; in questionnaire 2 (the complement to questionnaire 1), they read scenarios E, F, G and H as expert scenarios; in questionnaire 3, they read scenarios A, B, E and G as expert scenarios; and in questionnaire 4 (the complement to questionnaire 3), participants read scenarios B, D, F and H as expert scenarios. The order of presentation of the eight scenarios within each questionnaire was randomized, and questionnaires were presented using Medialab[®].

The instructions informed the participants that in each scenario they would be presented with two different ways of saying something. Therefore scenarios were immediately followed by the two possibilities, as in the example of expert scenario below. Participants indicated which sentence they would prefer to be told if they were the main character, by pressing key A or key B on the computer keyboard.

Jonathan had back pain. The chemist told him to take some Kebucid.

A. He told Jonathan that there was an 80% chance that this drug would work.

B. He told Jonathan that it was very possible that this drug would work.

6.3. Results

For each scenario participants chose between verbal and numerical probabilities. They were given a score of 1 for each choice of the numerical format and these were summed to give a score between 0 and 4 for each condition (expert / friend). A paired samples *t*-test indicated that participants chose more numerical predictions in the expert condition ($M = 2.37, SD = .95$) than in the friend condition ($M = 1.96, SD = .99$), $t(42) = 2.23, p = .031, r = .21$. One-sample *t*-tests showed that in the expert condition participants preferred to receive the numerical prediction more often than would be expected if they chose between the two options by chance, $t(42) = 2.56, p = .014, r = .37$. In the friend condition however, participants did not show a preference for one mode or the other, $t(42) = -.26, p = .799, r = .04$.

6.4. Discussion

Erev and Cohen (1990), Wallsten et al. (1993), and Xu et al. (2009) provided evidence that there is a general preference to receive uncertain information in numerical form but give it in verbal form, the communication mode preference paradox. However, I suspected that this was conflated with an expectation of people to receive particular types of information from expert speakers (Bohner et al., 2002). I speculated that people may show the preference for the numerical format only when the information came from an expert speaker. In line with this hypothesis, I found a greater preference for numerical probabilities (percentages) when the speaker was an expert (defined by higher or vocational education) than when s/he was a friend who had experience of the same situation. Furthermore, participants chose the numerical mode more often than chance when the speaker was an expert, but not when s/he was a friend.

These results were consistent with Bohner et al.'s findings (2002) that experts are expected to make strong arguments. In the case of communication of uncertainty, my results seemed to indicate that being an expert leads listeners to expect speakers to know enough to produce a more precise estimate to communicate. Following Grice's maxim of quantity (1975), listeners want to receive as much relevant information as possible and it seems likely that percentages are perceived as more informative. Thus, listeners prefer experts to provide them. But when the speaker was not an expert, s/he might not be expected to know enough to produce such precise numerical estimates. Using a verbal probability would then not be a violation of the maxim of quantity and would be pragmatically acceptable. Note that participants did not completely reject hearing numerical probabilities from non-experts. They showed no preference for either mode.

While consistent with Erev and Cohen's results (1990) that people prefer to receive numerical probabilities from speakers, my results challenge the generalizability of this claim. It is important to consider listeners' social expectations when investigating communication of uncertain information. Speaker expertise is an important contributor to social expectations, and should be controlled in future studies. In line with other studies, I found that communication about uncertainty is influenced by pragmatic factors. For example, verbal probabilities are interpreted in light of the conversational expectations of the listener and of the severity of the uncertain event (Bonnefon & Villejoubert, 2006): when participants thought that the speaker used verbal probabilities to be tactful (rather than to communicate likelihood), a severity bias (i.e., the overestimation of the likelihood of more severe events) was found. That is,

expectations that the speaker was hedging changed the numerical interpretation of the verbal probability. Another pragmatic factor, directionality of verbal probabilities is also used by speakers to fulfil an argumentative function (Juanchich, Teigen & Villejoubert, 2010). Positive verbal probabilities were used to revise predictions upwards, while negative verbal probabilities were used to revise predictions downwards.

These results are relevant to professions where practice relies on communicating uncertain information. My participants preferred experts to give them percentages. One possibility is that medical professionals and forecasters should consider making additional efforts to go against their natural tendency to use verbal probabilities (Erev & Cohen, 1990). But of course, based on my findings, I cannot say whether people's decision making is improved if given the type of information they prefer.

If using percentages makes a stronger argument, following Bohner et al. (2002), speakers could then expect to deliver a more convincing message. In other words, it is expected that an expert delivering uncertain information with percentages will be more convincing and the information seem more valid than if the expert uses verbal probabilities. However, Karmarkar and Tormala (2010) found the effect of expertise to backfire when associated with a high confidence of the speaker. A highly expert speaker was more convincing when expressing uncertainty than when being certain. Through their higher preciseness, percentages may make speakers sound more confident. In light of Karmakar's and Tormala's backfire effect, by using numerical probabilities, speakers could sound more confident and thus be less convincing. Further research investigating the effect of preciseness on persuasion would help disentangling

those two opposite accounts. If experts are less convincing numerical probabilities, it could suggest that numerical probabilities signal confidence.

I showed that people prefer to be told about uncertain events with percentages only when the speaker is an expert. This challenged the communication mode preference paradox (e.g., Erev & Cohen, 1990), where one generally prefers to give uncertain information as verbal probabilities, but to receive it as percentages. Future research will explore if, through this preference, a speaker can also deliver a more persuasive message using numerical probabilities.

CHAPTER 7: GENERAL DISCUSSION

7.1. Summary

I aimed in this thesis to provide an account of how children and adults understand and use verbal probabilities, within a pragmatic framework. This was done in three ways. First, I looked at how children used the different features of verbal probabilities (directionality and likelihood), and how this is influenced by their knowledge of speakers' intentions (Experiments 1a and 1b). These experiments were motivated by a need to develop the literature on children's understanding of verbal probabilities. While a few studies had explored the question, it was only in Gourdon and Villejoubert (2009) that this question was investigated in light of the research on the directionality of verbal probabilities. This study had found that 8 year-old children used only directionality in their likelihood judgements, but used both directionality and the likelihood when making decisions. This difference between the two tasks was explained by the concrete consequences of being wrong in the decision making task. However, it could have been explained by a practice effect.

Similarly to Gourdon and Villejoubert (2009), I found in Experiment 1a that 8-year-old children used only directionality to judge likelihood. I also found only directionality to be used in decision-making, suggesting that the effect of likelihood found in Gourdon and Villejoubert was indeed due to practice allowed by the task order. In Experiment 1b, I replicated these results. I also found evidence that directionality is used by children in combination with speakers' intentions. Therefore children use directionality to make decisions only when the speaker is benevolent, and dismissed it when the speaker is malevolent. This was consistent with a pragmatic account of verbal probabilities, where its

different features allow the speaker to convey further information than a simple estimate of likelihood (Teigen & Brun, 1999). This was only partially consistent with the recent development of the epistemic vigilance account (e.g., Mascaro & Sperber, 2009), where inferences are based on both speakers' intention and on their answers, from 5 years old. In Mascaro and Sperber's experiment, children made decisions that were opposite to the malevolent speakers' answer. That is, they reversed statements from malevolent speakers to infer the correct answer. In Experiment 1b, children decided randomly when speakers were malevolent, instead of making opposite decisions. Therefore I suggested that the integration of the information provided by directionality and of speakers' intentions was too costly for children.

In a second series of experiments, I endeavoured to show that directionality and the likelihood entail different levels of cognitive demands. This was expected to account for the framing effects of directionality (Experiments 2a and 2b). If directionality had lower levels of cognitive demands, it would explain why adults base their decisions on this feature rather than on the likelihood, as it would allow faster and less costly decisions. Then adults' decisions would get framed because directionality focused their attention (e.g., Teigen & Brun, 1988). These two experiments drew on the children's use only of directionality in Gourdon and Villejoubert (2009) and Experiments 1a and 1b, which could suggest that this feature requires fewer resources. I suggested that if directionality requires less cognitive resources, it should take less time to make decisions when directionality is the only feature that can be used, than when likelihood is the only feature that can be used. However, I found that decisions made only on the basis of directionality could take as much time as

deciding only on the basis of the likelihood, in some conditions. So my results suggested that both directionality and likelihood are taken into account during decision making.

In Experiment 2a, I also aimed to test if the framing effect of directionality remains when participants have the opportunity to compare verbal probabilities, as previous studies most often used designs where participants considered only one verbal probability at a time (e.g., Teigen & Brun, 1999). I found that this was the case, and in Experiment 2b I tested if this framing effect could be reduced by giving participants more time to make decisions. I observed that the framing effect of directionality could be overcome when the time pressure was relieved. However, this was only the case when the likelihood of the two compared verbal probabilities was similar. Then positive and negative verbal probabilities were chosen at chance level, if participants were given unlimited time. When the likelihood was different but incongruent with directionality (e.g., a positive verbal probability with a low likelihood), participants still (although at a lower extent) preferred the positive verbal probability half of the time, even if this was normatively the wrong answer. I suggested that this was consistent with a pragmatic account of verbal probability. If the speaker chose to use a directionality that is incongruent with the likelihood, it has to be because it is relevant (in the Relevance Theory framework; e.g., Wilson & Sperber, 2004): Speakers are communicating their intention that decision makers choose the option described by the positive directionality, and decision makers pick up on this.

To test the pragmatic account further, in Experiments 3a and 3b I used a more conversational context. In Experiment 3a, the framing effect of directionality was found as expected when the two verbal probabilities were uttered within a conversation. This was the case even when the time pressure was relieved. In Experiment 3b directionality was used in

combination with the benevolence or malevolence of the speakers. This supported the claim that directionality is used as an indicator of speakers' intentions (as suggested by Experiment 1b). Unlike the children in Experiment 1b who simply disregarded information from the malevolent speaker, adults used directionality in a reversed manner when the speaker was malevolent. Under these conditions they preferred the option predicted with a negative verbal probability. Likewise, in Mascaro and Sperber's study (2009), children used speakers' statements in a reverse manner when they were malevolent. This suggests that the usual answer to malevolent speakers is to use their statement to infer the truth. This therefore supported my suggestion that in Experiment 1b, children might have lacked the resources to integrate fully directionality and the intention, leading them simply to dismiss the information given by malevolent speakers. More importantly, this supported the claim that by choosing a positive or a negative directionality, speakers provide more than simple numerical, probabilistic information, in line with Teigen and Brun's suggestion (1999).

In Chapters 5 and 6, the last experimental set of this thesis drew on Beyth-Marom's suggestion (1982) that verbal probabilities are preferred by speakers because they allow speakers to diffuse their responsibility. The aim of Chapter 5 was to test this proposition indirectly, by investigating to what extent speakers are judged responsible when using verbal probabilities or numerical probabilities. The format chosen to communicate probabilities had an inconsistent effect across Experiments 4a to 4e. In Experiment 4a, there was no difference between numerical probabilities and positive verbal probabilities. But negative verbal probabilities yielded an effect of the valence of the uncertain event, with higher responsibility judgements when the uncertain event was positive. In Experiment 4b, round percentages yielded an effect of the outcome, with higher responsibility judgements when

the uncertain event did not occur. Precise percentages did not yield such an effect. In Experiment 4c, speakers were judge as more responsible when they used round percentages. Finally in Experiments 4d and 4e, the communication mode had no main or interaction effect. However, when an effect was found it was always in the direction that using more precise statements led to being held less responsible. One interpretation of this is that speakers were considered as having done their best. As speakers had provided precise percentages, participants may have considered that they had done as much as they could to inform.

What was more consistent was an effect of the consequences. Similar to what is known as the Knobe effect or side-effect effect (e.g., Knobe, 2003), in Experiments 4a, 4b and 4e, speakers were held more responsible if consequences were negative. Conversely to the Knobe effect however, in Experiments 4c and 4d, they were held more responsible when ultimate consequences were positive. Further Experiment 4c and 4d had higher power, which supports more an effect opposite to the Knobe effect. I suggested that these differences could be due to the meaning of responsibility not being defined enough. That is, responsibility could have been understood as blame on the one hand (i.e., with a negative value), and as accountability on the other hand (i.e., with a neutral value). Experiment 4e was set up to test if an overlap between the concepts of causality and responsibility could explain the inconsistency in my results. However, the results showed that the effect of the ultimate consequences went in the same direction whether the judgement was framed in terms of causality or in terms of responsibility, bringing no support to this possibility. However, the power may have been insufficient here as well.

Nevertheless, the focus on the ultimate consequences rather than on the accuracy of the prediction remained the most interesting set of results in Chapter 5. According to a normative perspective, listeners should be interested in a forecaster who can provide good predictions. The results of that chapter suggested that, as often found in judgement and decision making, what ought to be is not what is (e.g., Kwan, Wojcick, Miron-Shatz, Votruba & Olivola, 2012). From a rational point of view, speakers provided predictions and should only be held responsible for the quality of these predictions. That is, if they predicted an event with a high likelihood and the event did not occur, the quality of their prediction is low, and their responsibility should be engaged. The same should be the case if they had predicted an event with a low likelihood and in fact it occurred. However, when participants were also asked to indicate how likely they would be to recommend the forecaster, their answers reflected more attention to the accuracy. It seemed therefore that it is not that participants could not evaluate the accuracy, but rather that it does not matter in the responsibility judgements.

Finally, in Chapter 6, I looked at adults' preferences for receiving numerical probabilities in light of their knowledge of the speakers' expertise. Previous research had found that people prefer to receive predictions as numerical probabilities. This is despite the fact that most people cannot judge their likelihood accurately (e.g., Gonzalez & Wu, 1999). However, to my knowledge, it had not been investigated if this preference depends on speakers' characteristics. According to a Gricean pragmatic perspective, expert speakers are expected to use numerical probabilities more often, as they should give the most relevant information they possess under the maxim of quantity (Grice, 1975). Non-expert speakers, being expected to know less, should avoid violating the maxim of quantity and instead use vague,

verbal probabilities. Therefore, they may be expected to be less likely to use numerical probabilities. Results were consistent with these expectations. They also supported previous research showing that experts are expected to use stronger arguments (Bohner et al., 2002). This was also in line with the previous chapters of this thesis, indicating that the use and interpretation of verbal probabilities is influenced by pragmatic factors beyond the likelihood they convey.

7.2. Links to literature

After Teigen and Brun (1999) started investigating whether verbal probabilities conveyed information about speakers' intentions, Budescu et al. (2003) objected that directionality was merely a reflection of the likelihood of verbal probabilities. That is, according to Teigen and Brun's perspective, directionality hints at the speaker's intention by framing the listener's perspective. Whereas according to Budescu et al. directionality is only a function of the likelihood, and therefore supports a normative decision-making by framing the perspective in the correct direction. Gourdon and Villejoubert (2009), however, showed that a positive directionality is not always associated with a high likelihood, or a negative directionality with a low likelihood. Furthermore, Juanchich et al. (2009) showed that directionality can be used as an argumentative tool, in which case it is chosen according to the direction in which a speaker needs to revise a statement of uncertainty (positive for upwards, negative for downwards). In Experiment 2b, participants were equally likely to choose a positive or a negative verbal probability when both expressed either a high or a low likelihood. This showed that positive directionality is not confined to high likelihoods and the negative one to low likelihoods, conversely to Budescu et al.'s claim.

Drawing on Teigen and Brun's idea that directionality is indicating something more about what the speaker wants, in Experiment 3b, participants used directionality in different ways if the speaker was known to be malevolent. Consistent with this, in Experiment 4a, judgements of responsibility were a function of directionality and the valence of the ultimate consequences. Thus, when the consequences were negative, speakers were judged as more responsible when they had used a positive directionality than when they had used a negative directionality. It seemed that speakers were therefore rewarded for expressing their wish that the negative consequences would not occur. This further supported the claim that directionality is a pragmatic feature of verbal probabilities, which conveys information about intentions as well as about simple likelihood.

These results are consistent with the argumentative perspective of language (Anscombe & Ducrot, 1983), in which language both informs and argues, and the argumentation depends on the structure of language. Under this perspective, directionality of expressions of quantity (verbal probabilities or simple quantifiers) influences the perspective that the listener takes when interpreting those expressions (Sanford & Moxey, 2003). This is what leads to framing effects such as those observed in Chapters 3 and 4. Listeners seem to interpret the choice of directionality made by speakers as deliberate, leading them to struggle to make a choice when directionality contradicts the likelihood. This definition of directionality as a feature which signals a speaker's intention is also consistent with Relevance Theory (e.g., Wilson & Sperber, 2004). Within the Relevance Theory framework, utterances should provide relevant information, so that the cognitive cost of processing them does not outweigh the benefits. Therefore if a speaker chooses to use a negative directionality over a positive, it is not meaningless, and the listener should use this as a cue.

As I showed in Experiment 3b, the perspective provided by directionality is interpreted according to what is known about the speaker's intention. This supports the claim that listeners consider the choice of directionality as deliberate. This is also consistent with the recent development of the concept of epistemic vigilance (Sperber et al., 2010). Because language also fulfils an argumentative function as well as an informative one (Anscombe & Ducrot, 1983), communication entails a risk of being misinformed. That is, speakers can use the argumentative function to frame listeners' perspective towards one particular possibility (e.g., Sanford & Moxey, 2003). In that case, listeners would pay less attention to the other possibilities. Misinformation is costly, from the cognitive cost of understanding utterances that are ultimately not useful, to the potential negative consequences of decisions made based on incorrect information. Sperber et al. (2010) argued, however, that human beings have evolved a set of mechanisms tailored to detect the worthiness of a communicator, e.g., epistemic vigilance. Epistemic vigilance is evidenced by developmental studies, where children show an appropriate appreciation of both the past accuracy (e.g., Nurmsoo & Robinson, 2009) and the intention of speakers (e.g., Mascaro & Sperber, 2009). For example, in Mascaro and Sperber's study, children used what a malevolent speaker was saying to make inferences regarding the truth.

In Experiment 1b, I provided further support for the development of epistemic vigilance, as children used directionality differently when the speaker was known to have bad intentions. However, I found that children dismissed the speaker's prediction in that case, instead of using it to infer the correct answer (i.e., the most likely), as did adults in Experiment 3b. This is slightly different from Mascaro and Sperber's claim (2009), but could be explained by higher cognitive demands of the task used in Experiment 1b. Mascaro and

Sperber's task used actual puppets and boxes, instead of children having to maintain the characters and choices in working memory. It is nevertheless the case that directionality can be used differently after the source had been evaluated through to epistemic vigilance. This supports the suggestion that directionality is the feature that fulfils the argumentative function in verbal probabilities.

The understanding of verbal probabilities could also be accounted by Mercier and Sperber's (2011) argumentative account of reasoning. Mercier and Sperber proposed that reasoning serves an argumentative function. Under this perspective, framing effects can occur because the choices of language elicit different inferences. That is, they influence how new representations are produced. For example McKenzie and Nelson (2003) found that listeners assume that a glass used to be full if a speaker describes it as half empty. That is, listeners can infer speakers' reference point from the framing of their statement. I made reference to some evidence that verbal probabilities also elicit different inferences depending on their locus of uncertainty (Juanchich et al., 2010), or that their directionality is chosen differently when speaker's reference points differ (Juanchich et al., 2009). Further I have presented evidence suggesting that different directionalities elicit different inferences on what the speakers know and want listeners to do. By framing perspectives, as suggested by Sanford and Moxey (2003), verbal statements of quantity highlight some reasons more than others and frame the reasoning itself.

An argumentative reasoning account of verbal probabilities could also lead one to reject Windschitl and Wells' (1996) suggestion that verbal probabilities elicit Type 1 reasoning, which is intuitive, fast and cognitively cheap. Based on children's use of directionality only, my own suggestion in Chapter 2 was that making decisions on the basis of directionality

alone was less costly. This would mean that using the directionality would be Type 1 reasoning, and that using the likelihood would be a form of Type 2 reasoning. However, I found in Chapters 3 and 4 that directionality-driven decisions were not less costly than likelihood-driven ones. Deciding between two high verbal probabilities, only differing in directionality, took as much time as deciding between two positive verbal probabilities, only differing in likelihood. It does not seem to be that there are heuristic answers to verbal probabilities, using directionality, and also normative ones, using the likelihood.

It could however be the case that people implicitly know the normative answer, and this answer conflicts with the heuristic answer which uses directionality. This would be consistent with what De Neys (2012) has suggested for other reasoning tasks. This would explain the longer reaction time when the likelihood is similar, both verbal probabilities expressing low likelihood: in that case the normative answer of not choosing an unlikely option would conflict with the heuristic answer of choosing the positive verbal probability. When both verbal probabilities were incongruent (i.e., a low likelihood expressed by a positive verbal probability, and vice-versa), the normative answer of choosing the high likelihood would conflict with the heuristic of not choosing a negative directionality. Finally, when both verbal probabilities had a negative directionality, the normative answer (choosing the higher likelihood) would also be in conflict with the heuristic 'Do not pick a negative directionality'.

It could seem that a heuristic/normative (or Type 1/Type 2) perspective on the features of verbal probabilities does not allow one to take into account the pragmatic/argumentative account of directionality. In a heuristic/normative account, directionality would be used preferentially by adults because it is a fast and less costly tool. In an argumentative account,

directionality would be used because it is framing listeners' inferences about what speakers know. However, the epistemic vigilance account could accommodate the two accounts as epistemic vigilance could allow dealing with the cognitive limits of human rationality. It is indeed costly to be misinformed, but it would also be costly to evaluate each received statement and to systematically make inferences. Epistemic vigilance allows one not only to avoid the cost of misinformation, but also to rely on speakers once they have been evaluated as trustworthy and knowledgeable, in order to avoid the cost of reasoning when it is not needed. Using directionality as a pragmatic cue to what speakers think/argue should be done *would be* the heuristic in this perspective.

However as seen in Experiment 3b, adults adopt another heuristic once speakers have been evaluated as malevolent: they use directionality systematically as an inverse cue. This is opposite to what would be predicted by an integrated account of epistemic vigilance and heuristic and normative reasoning. Once a speaker is deemed malevolent, the normative answer would be in line with my findings in Experiment 1b with children. That is to dismiss the answer and choose randomly. If a more normative answer can be reached (for example using the likelihood), it may be *reasonable* to use it. Whereas children's interpretations of verbal probabilities do not incorporate the likelihood because they do not yet fully understand it (as shown in Experiment 1a), adults should do so. Therefore they should answer randomly when verbal probabilities differ only by their directionality, since the likelihood cannot provide them with an answer. But when only the likelihood is different, this should be used to reach a decision in a normative way. This is consistent with Experiment 3b, where participants faced with a malevolent speaker chose the negative

directionality more often when likelihoods were similar, and the lower likelihood when directionalities were the same.

7.3. Future research

In Chapter 2, I showed that at 8-9-year-old English-speaking children use only directionality to judge and make decisions, as also found in French-speaking children (Gourdon & Villejoubert, 2009). There were, however, some trends in Experiment 1b suggesting that 8- to 9-year-old children might have also started developing their interpretation of likelihood information. The first step in investigating further the understanding of verbal probabilities in children should therefore extend the age range and try to describe the development of the use of both features. Given that 7- to 8-year-old children in this thesis used directionality, extending the age range downwards would be informative to find at which age children start using directionality. Based on Champaud and Bassano's results (1987) with quantifiers, children can display some sensitivity to argumentative function from 6 years old. Therefore it could be expected that directionality is used to make decisions as early as 6 years old.

Children studying for the GCSE exam in the UK (15- to 16-year-olds) are specifically taught which likelihood meanings are to be associated to simple verbal probabilities (BBC, 2012; see Appendix IV, p. 214, for an example of exercise with answers). It would therefore be interesting to compare children following the British curriculum, i.e., taught how to use verbal probabilities, to children following a curriculum where this is not taught (e.g., the American curriculum, as it is delivered in the same language). First, the development of the use of the likelihood might be later if verbal probabilities are not explicitly part of the curriculum. Second, if teaching the use of verbal probabilities has any effect on their

understanding, adolescents who have received this curriculum should interpret verbal probabilities in a way that matches the content of this teaching. Crucially, they should also display less intraindividual variability. This last point would further inform us if it is possible to shape adults' use of verbal probabilities. The question of prescribing the use of verbal probabilities is not new. For example, Hamm (1991) suggested that a lexicon of verbal probabilities should be established to standardize their use. But the question has resurged recently in the high stakes context of communicating about climate change (Budescu, Broomell & Por, 2009). For example, current research is being done across several cultures and languages, where participants are told beforehand of the meaning of each verbal probability used (D. Budescu, personal communication, March 16, 2010), and then asked to evaluate statements on climate change. The purpose of the project is to try to recommend verbal probabilities that organizations should use in their communication. Therefore, testing if teaching of the meaning of verbal probabilities influences their interpretation would inform this question as well.

I highlighted that recently research on verbal probabilities has also explored the distinction between internal and external verbal probabilities (e.g., Fox & Ülkümen, 2011). Although this research is still developing and results are not consistent, the question of how children take into account the locus of uncertainty should be asked. Robinson et al. (2009) showed that children make different choices when uncertainty is epistemic (internal) or physical (external), preferring to bet when the uncertainty is internal. Children are therefore able to make the distinction between internal and external uncertainty. Such a distinction should be reflected in children making different interpretations of internal and external verbal probabilities.

Research on the locus of uncertainty in verbal probabilities has also found that the external/internal nature of phrases informs adults on the source of knowledge (Juanchich et al., 2010). That is, adults are more likely to interpret that speakers based their prediction on statistical information, when speakers use an external verbal probability. Therefore the locus of uncertainty could be used to manipulate the speaker's competence. For example, speakers' competence could be demonstrated by the use of predictions based on external verbal probabilities, and their lower competence could be demonstrated by the use of predictions based on internal verbal probabilities. In the same way that intention influences the use of directionality by adults and children, according to an epistemic vigilance perspective (Sperber et al., 2010), listeners' perception of competence could change the way adults use directionality.

I argued earlier that verbal probabilities convey more than simple likelihood information, and that directionality in particular is a pragmatic feature of verbal probabilities. A further test of this account would be to investigate if individual differences in pragmatic competence have an effect on the decisions made by listeners (directionality-driven or likelihood-driven). Nieuwland, Ditman and Kuperberg (2010) used for example the Autism Quotient Communication Subscale in a general population. They found the scale to be able to distinguish participants based on their sensitivity to violations of pragmatic rules (e.g., that communication is meant to be informative enough). That is, participants categorized by the scale as having higher pragmatic competence were more sensitive to low informativeness than participants categorized as having lower pragmatic competence. The Autism Quotient Communication Subscale method could therefore be used in the context of verbal probabilities. It would be expected that participants with higher pragmatic

competence as defined by the scale would use more often directionality than participants with lower pragmatic competence.

Finally, a third line of further research should develop the results of Chapter 6. In Experiment 5, I have shown that only when speakers are experts, are they expected to use numerical probabilities. When the speakers were friends who have some experience of the situation, there was no preference for numerical or verbal probabilities. I suggested that Gricean pragmatics could explain this result. Under the maxim of quantity, speakers are expected to provide the maximum relevant information, which an expert will have in a higher quantity. Therefore, experts would be expected to be more informative. Numerical probabilities are less vague, so are likely to seem more informative. Therefore, they are expected from experts as expectations of informativeness are higher.

One could note however that the non-expert speakers were also friends. It might be that expectations of informativeness were not higher for experts compared to non-experts, but rather that vagueness is more permissible or less unexpected from friends compared to strangers. Under a pragmatic perspective, verbal probabilities can be perceived as hedging device (Juanchich et al., 2012). Hedging can allow speakers to save face, allowing them to avoid admitting that they lack of information. It can also allow listeners to save face: the vagueness of verbal probabilities allows speakers to avoiding predicting negative events too concretely, that is bad news can be broken gently. Friends therefore might be expected to hedge more, in order to protect listeners. This could lead to people being less likely to expect to receive numerical probabilities from friends. Further research should test this possibility, by running a follow on experiment to Experiment 5, with an additional speaker condition where the speaker is neither an expert nor a friend.

7.4. Conclusion

In this thesis I have investigated the understanding of verbal probabilities in children and adults, with a particular interest in directionality of verbal probabilities. I have also examined the influence of different formats of communication of probabilities on judgements by adults. I have argued that verbal probabilities convey more than simple probabilistic information, following Teigen and Brun (1999). I have provided evidence that directionality is a pragmatic feature of verbal probabilities, which signals to listeners what speakers want and/or know. This was particularly shown when both children and adults, confronted by a malevolent speaker, changed the way they relied on directionality to make decisions. In a second phase of research, I suggested that forecasters are not judged responsible based on the format they use to communicate likelihood, nor on their accuracy. Rather the consequences matter, as in the Knobe effect (e.g., Knobe, 2003), although not always in the same direction. I also showed that forecasters are, however, expected to be more precise when they are experts than when they are non-expert friends. I then suggested that this might have consequences for the level of persuasion of a speaker. In this thesis I have therefore advanced a pragmatic account of verbal probabilities and risk communication. This supports that verbal probabilities have pragmatic features and communicated likelihood is interpreted by listeners on the basis of pragmatic assumptions. Future research should focus on developing evidence that the pragmatic features of verbal probabilities are multiple, for example by casting light on the locus of uncertainty.

Appendices

Appendix I: Verbal probabilities used in the pilot study of Experiment 1a

It is almost impossible	It is a little doubtful
It is very unlikely	It is quite probable
It is unlikely	It is quite improbable
It is not certain	It is quite likely
There are a few chances	There is some possibility
There are few chances	It is somewhat likely
It is possible	It is almost certain
It is rather certain	It is doubtful
It is very likely	It is uncertain
It is almost certain	It is quite unlikely
There are a few doubts	It is quite doubtful
It is a little unlikely	It is not entirely sure
It is probable	There is a small possibility
It is improbable	It is not entirely certain
There is a very poor chance	It is very possible
There is a good chance	There is a small chance
It is very doubtful	

Appendix II: Scenarios used in Experiment 1a

Example of training scenario



Anna found a clue saying that it is absolutely sure that the treasure is in the chest.

Experimental scenarios



Julie had been invited to play at her friend's home on Saturday. When she asks her parents if she can go, they tell her that **there is some possibility that her cousin will come from Scotland that day**. However they say that it is up to her to decide whether she will go to her friend's or she will stay at home in case her cousin comes.



Marc is at his friend's birthday party. He decides to have some cake and to wait to have some sweets later. His friend's mum tells him that there are a few chances that there are some sweets left by this time.

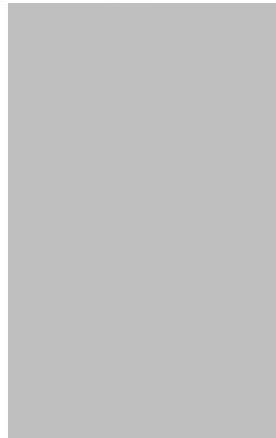


Some days Stephanie gets the bus to school, some days she walks. Today she just missed her bus for school. Before deciding to walk to school she asks her neighbour who's waiting also when the next bus will come. Her neighbour tells her that it is not entirely certain that it comes soon.



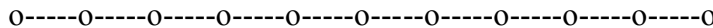
Karl is watching his favourite TV show when his mum offers him to play a board game. Karl asks if he can finish watching his TV show. His mother tells that it's OK and that it is a little unlikely that she will have time to play later.

Appendix III: Example page of the scenario booklet used in Experiment 1a

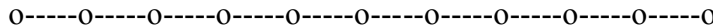


Captain Hook told Julie that **there is some possibility that the treasure is in the chest.**

When Captain Hook says **there is some possibility that the treasure is in the chest,** how much do you think that this means that the treasure is in the chest?



As **there is some possibility that the treasure is in the chest,** how happy do you think Julie will be if she chooses to open the chest?



As **there is some possibility that the treasure is in the chest,** if you were Julie, what would you choose to do?

not open the chest



open the chest



Appendix IV: Screenshot of an exercise proposed to students preparing GCSEs (BBC, 2012)



Appendix V: Materials used in Experiment 4a

Scenarios

Positive uncertain events

A. When he went to Keele University for visit day, Samuel asked a student about the admission rates. The student at the visit day told him that [insert percentage or verbal probability] that more applicants would be admitted this year. Samuel put Keele as his first choice. [insert outcome]

B. Peter was looking for a summer job. He asked his uncle about one of his friends who owns a company. His uncle told him that if he volunteered for his friend for a week, [insert percentage or verbal probability] he would get a job afterwards in the company. Peter volunteered for his uncle's friend. [insert outcome]

C. Hannah wanted to study Italian at University, but was still hesitating a little. When she met her school counsellor, this latter told her that if she does, there was a X% chance she would find a job within a year from graduation. Hannah decided to study Italian. [insert outcome]

D. James was meant to meet his last date for the second time tonight, in the city centre. When he arrived near there his car got stuck in the traffic, so he asked a policeman which alternative route he could take. The policeman told him that by the South [insert percentage or verbal probability] it would be faster. James took the alternative route. [insert outcome]

Negative uncertain events

E. Patrick asked the chemist assistant advice for his stomach ache. He recommended a drug and said that [insert percentage or verbal probability] to give him a rash. Patrick took the drug. [insert outcome]

F. Lauren was told by her friend who works at the airport that [insert percentage or verbal probability] that flight crews would go on strike at Easter. Lauren booked her hotel in Spain. [insert outcome]

G. Helena had a big mole on her neck. The doctor said she could have an operation to have it removed. [insert percentage or verbal probability] it would leave a permanent scar. Helena had the operation. [insert outcome]

H. Emily's car had a problem with the brake fluid. Having it repaired would cost £200 and the mechanic told her that after the repair [insert percentage or verbal probability] she would have to buy a new car. Emily made it repair. [insert outcome]

Probabilistic expressions used in Experiment 4a

	numerical probabilities	positive verbal probabilities	negative verbal probabilities
low likelihood	there was a 20% chance	there was a small possibility	it was quite improbable
	there was a 30% chance	there were a few chances	it was quite unlikely
high likelihood	there was a 70% chance	it was quite likely	it was not guaranteed
	there was a 80% chance	it was rather certain	it was not definite

Outcomes used in Experiment 4a

Scenario	Uncertain event occurred	Uncertain event did not occur
A	He got a place there.	He did not get a place there.
B	He got a Summer job in his company.	He did not get a Summer job in his company.
C	She found a job 6 months after graduating.	She found a job 15 months after graduating.
D	He arrived on time for his date which then went good.	He did not arrive on time and his date was already gone.
E	He got a rash.	He did not get a rash.
F	The strike went on and she had to take a long ferry to Spain.	The strike was cancelled and she got a short flight to Spain.
G	It left her with a permanent scar.	She did not get a permanent scar.
H	It didn't work and she had to buy a new car.	It worked and Emily did not have to buy a new car.

Appendix VI: Scenarios used from Experiment 4b to Experiment 4e

Scenarios

A. Ewan wanted to place a small bet on the Football World Cup. He asked a friend who is a big football fan about it. His friend told him that there was a [insert numerical probability]% chance that England would qualify from their group. [insert outcome]

B. Peter was looking for a summer job. He asked his uncle about one of his friends who owns a company. His uncle told him that if he volunteered for his friend for a week, there was a [insert numerical probability]% chance he would get a job afterwards in the company. Peter volunteered for his uncle's friend. [insert outcome]

C. When Tom was at the bus stop this morning, he asked a commuter who takes this route daily if the bus should come soon. The person said that there was a [insert numerical probability]% chance that the bus was coming pretty soon. Tom waited for the bus instead of walking. [insert outcome]

D. James was meant to meet his last date for the second time tonight, in the city centre. When he arrived near there his car got stuck in the traffic, so he asked a policeman which alternative route he could take. The policeman told him that by the South there was a [insert numerical probability]% chance that the bus was coming pretty soon. James took the alternative route. [insert outcome]

E. When he went to Keele University for visit day, Samuel asked a student about the admission rates. The student at the visit day told him that there was a [insert numerical probability]% chance that more applicants would be admitted this year. Samuel put Keele as his first choice. [insert outcome]

F. Helena had a big mole on her neck. The doctor said she could have an operation to have it removed. There was a [insert numerical probability]% chance it would not leave any scar. Helena had the operation. [insert outcome]

G. Eve saw a coat she liked very much in a shop. The shop assistant told her that there was a [insert numerical probability]% chance that the coat would be in the sales. Eve waited until the sales started. [insert outcome]

H. Hannah wanted to study Italian at University, but was still hesitating a little. When she met her school counsellor, this latter told her that if she does, there was a [insert numerical probability]% chance she would find a job within a year from graduation. Hannah decided to study Italian. [insert outcome]

Probabilistic expressions used from Experiment 4b to Experiment 4e

	round numerical probabilities	precise numerical probabilities
low likelihood	there was a 20% chance	there was a 23% chance
	there was a 30% chance	there was a 32% chance
high likelihood	there was a 70% chance	there was a 67% chance
	there was a 80% chance	there was a 84% chance

Outcomes used from Experiment 4b to Experiment 4e

Scenario	Uncertain event occurred	Uncertain event did not occur
A	England qualified and Ewan won his bet.	England did not qualify and Ewan lost his bet.
B	He got a Summer job in his company.	He did not get a Summer job in his company.
C	The bus came in the five minutes and Tom was on time.	The bus did not come for 15 minutes and Tom was late.
D	He arrived on time for his date which then went good.	He did not arrive on time and his date was already gone.
E	He got a place there.	He did not get a place there.
F	It left her with a permanent scar.	She did not get a permanent scar.
G	When she came back, she got it for half the initial price.	When she came back, all the coats were gone.
H	She found a job 6 months after graduating.	She found a job 15 months after graduating.

Appendix VII: Material in Experiment 5 (in experimental order)

General instructions: 'You are going to read short stories where people make decisions. In each story, one character will say something in two ways. Your task is to select how you would prefer to hear this information if you were the main character in the story.'

expert scenarios

- A** Jonathan had back pain. The chemist told him to take some Kebucid.
He told Jonathan that there was an 80% chance that this drug would work.
He told Jonathan that it was very possible that this drug would work.
How would you prefer the chemist to tell you what he thinks?
- B** Linda wanted to invest in the Rigobel stock.
Her financial consultant told her that there was a 68% chance that this investment would be profitable.
Her financial consultant told her that it was quite likely that this investment would be profitable.
How would you prefer the financial consultant to tell you what she thinks?
- C** Ben needed to take his car for an MOT but it was making a weird noise.
The mechanic told Ben that there was a 73% chance that his car would pass the MOT.
The mechanic told Ben that there was a good chance that his car would pass the MOT.
How would you prefer the mechanic to tell you what he thinks?
- D** Gina had been trying to get pregnant for a while and was considering having IVF.
Her obstetrician told Gina that there was an 84% chance that she would get pregnant with IVF.
Her obstetrician told Gina that it was very likely that she would get pregnant with IVF.
How would you prefer the obstetrician to tell you what she thinks?
- E** Christopher was going on holiday to the Isle of Wight.
The Met Office meteorologist said that there was a 69% chance that the weather would be sunny.
The Met Office meteorologist said that it was quite probable that the weather would be sunny.
How would you prefer the meteorologist to tell you what she thinks?
- F** Emily bought some orchids.
The florist told her that if she cut them in a certain way, there was a 92% chance that the orchids would last longer.
The florist told her that if she cut them in a certain way, it was almost certain that the orchids would last longer.
How would you prefer the florist to tell you what he thinks?
- G** Giles wanted to upgrade his TV to an Elga-1210.
The journalist from "Which?" said that there was a 61% chance that the Elga-1210 would be very efficient to run.
The journalist from "Which?" said that it was somewhat likely that the Elga-1210 would be very efficient to run.
How would you prefer the journalist to tell you what he thinks?
- H** Christine wanted to do a Masters degree in Engineering.
The programme director told her that there was a 71% chance that she would get an engineering job after this Masters.
The programme director told her that it was probable that she would get an engineering job after this Masters.
How would you prefer the program director to tell you what she thinks?
-

non-expert scenarios

- A** Jonathan had back pain. His friend had had the same pain and taken Kebucid.
He told Jonathan that there was an 80% chance that this drug would work.
He told Jonathan that it was very possible that this drug would work.
How would you prefer your friend to tell you what he thinks?
- B** Linda wanted to invest in the Rigobel stock. Her friend invested in a similar stock recently.
She told Linda that there was a 68% chance that this investment would be profitable.
She told Linda that it was quite likely that this investment would be profitable.
How would you prefer your friend to tell you what she thinks?
- C** Ben needed to take his car for an MOT but it was making a weird noise. His friend drives the same car and had the same problem.
He told Ben that there was a 73% chance that his car would pass the MOT.
He told Ben that there was a good chance that his car would pass the MOT.
How would you prefer your friend to tell you what he thinks?
- D** Gina had been trying to get pregnant for a while and was considering having IVF.
Her friend who had IVF told her that there was an 84% chance that she would get pregnant with IVF.
Her friend who had IVF told her that it was very likely that she would get pregnant with IVF.
How would you prefer your friend to tell you what she thinks?
- E** Christopher was going on holiday on the Isle of Wight. His friend went on holiday there at the same time last year.
She told Christopher that there was a 69% chance that the weather would be sunny.
She told Christopher that it was quite probable that the weather would be sunny.
How would you prefer your friend to tell you what she thinks?
- F** Emily bought some orchids. Her friend had some orchids in the past and she cut hers in a certain way.
He told Emily that if she cut them in this way, there was a 92% chance that the orchids would last longer.
He told Emily that if she cut them in this way, it was almost certain that the orchids would last longer.
How would you prefer your friend to tell you what he thinks?
- G** Giles wanted to upgrade his TV to an Elga-1210. His friend had also bought this model.
He told Giles that there was a 61% chance that the TV Elga-1210 would be every very efficient to run.
He told Giles that it was somewhat likely that the Elga-1210 would be every very efficient to run.
How would you prefer your friend to tell you what he thinks?
- H** Christine wanted to do a Masters in Engineering. Her friend did the same Masters last year.
She told Christine that there was a 71% chance that she would get an engineering job after this Masters.
She told Christine that it was probable that she would get an engineering job after this Masters.
How would you prefer your friend to tell you what she thinks?
-

preference questions

How do you usually prefer people to give you their judgements of uncertainty?

Numerically (e.g., "There is a 50% chance")

Verbally (e.g. "There is a fifty-fifty chance")

How do you usually prefer to give your judgements of uncertainty to others?

Numerically (e.g., "There is a 50% chance")

Verbally (e.g. "There is a fifty-fifty chance")

References

- Anscombe, J.-C (1985). Théorie de l'argumentation, topoi et structuration argumentative (Argumentation theory, topoi and argumentative structuration). *Revue de Linguistique Québécoise*, 18, 13-65. DOI : 10.7202/602639ar
- Anscombe, J., & Ducrot, O. (1983). *L'argumentation dans la langue* (Argumentation in language). Bruxelles, Belgium: Pierre Mardaga.
- Baron, J. (2008). *Thinking and deciding* (4th edition). Cambridge, UK: Cambridge University Press.
- BBC (2012). GCSE Bitesize: Using words to describe probability. Retrieved from <http://www.bbc.co.uk/schools/gcsebitesize/maths/statistics/probability1rev1.shtml>
- Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting*, 1, 257-269.
DOI: 10.1002/for.3980010305
- Bismark, M. M., Gogos, A. J., Clark, R. B., Gruen, R. L., Gawande, A. A., & Studdert, D. M. (2012). Legal disputes over duties to disclose treatment risks to patients: A review of negligence claims and complaints in Australia. *PLoS Medecine*, 9, e100128. doi: 10.1371/journal.pmed.1001283
- Bohner, G., Ruder, M., & Erb, H. (2002). When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology*, 41, 495-519. doi: 10.1348/014466602321149858

- Bonnefon, J.-F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science, 17*, 747-751. doi: 10.1111/j.1467-9280.2006.01776.x
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British national corpus*. London: Longman. Frequency lists retrieved from <http://ucrel.lancs.ac.uk/bncfreq/flists.html>
- Budescu, D. V., Karelitz, T. M., & Wallsten, T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making, 16*, 159-180. doi: 10.1002/bdm.440
- Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes, 36*, 391-405. doi: 10.1016/0749-5978(85)90007-X
- Champaud, C., & Bassano D. (1987). Argumentative and informative functions of French intensity modifiers “presque” (almost), “à peine” (just, barely) and “à peu près” (about): An experimental study of children and adults. *European Bulletin of Cognitive Psychology, 7*, 605-631
- Chance (2012). In *Oxford Dictionaries Online*. Retrieved from <http://oxforddictionaries.com/definition/english/chance?q=chance>
- Cohen, J., Dearnley, E. J., & Hansel, C. E. M. (1958). A quantitative study of meaning. *British Journal of Educational Psychology, 28*, 141-148. DOI: 10.1111/j.2044-8279.1958.tb01435.x
- Davies, M. (2008-2012). *The Corpus of Contemporary American English: 450 million words, 1990-present*. Retrieved from <http://corpus.byu.edu/coca/>

- Das, E. H. H. J., de Wit, J. B. F., & Stroebe, W. (2003). Fear appeals motivate acceptance of action recommendations: Evidence for a positive bias in the processing of persuasive messages. *Personality and Social Psychology Bulletin, 29*, 650-664. doi: 10.1177/0146167203029005009
- Dhami, M. K., & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities. *Memory & Cognition, 33*, 1057-1068. doi: 10.3758/BF03193213
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7*, 28-38. doi: 10.1177/1745691611429354
- Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting 'ought' from 'is': Descriptivism versus normativism in the study of the human thinking. *Behavioral and Brain Sciences, 34*, 233-248. doi: 10.1017/S0140525X1100001X
- Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes, 45*, 1-18. doi: 10.1016/0749-5978(90)90002-Q
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255-278. DOI: 10.1146/annurev.psych.59.103006.093629
- Evans, J. St. B. T. (2010). Intuition and reasoning: A dual process perspective. *Psychological Inquiry, 21*, 313-326. DOI: 10.1080/1047840X.2010.521057
- Fagley, N., & Miller, P. (1990). The effect of framing on choice: Interactions with risk-taking propensity, cognitive style, and sex. *Personality and Social Psychology Bulletin, 16*, 496-510. doi: 10.1177/0146167290163008

- Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkebøen & H. Montgomery (Eds.), *Perspectives on thinking, judging, and Decision Making* (pp. 21-35). Oslo, Norway: Universitetsforlaget.
- Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. D. Cummins & C. Allen (Eds.), *The evolution of mind* (pp. 9-29). New York, NY: Oxford University Press.
- Gigerenzer, G. (2003). *Reckoning with risk. Learning to live with uncertainty*. London, UK: Penguin Books.
- Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, 25, 623-629. doi: 10.1111/j.1539-6924.2005.00608.x
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129-166. doi: 10.1006/cogp.1998.0710
- Gough, P. B. (1966). The verification of sentences: The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, 5, 492-496. doi: 10.1016/S0022-5371(66)80067-1
- Gourdon, A., & Beck, S. R. (unpublished). [Likelihood and plausibility judgements based on verbal probabilities in adult participants]. Unpublished raw data.
- Gourdon, A., & Villejoubert, G. (2009). A comparison of children and adults' judgements and decisions based on verbal uncertainty statements. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

- Grice, H. P. (1975). Further notes on logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York, NY: Academic Press.
- Hacking, I. (1975). *The emergence of probability*. London: Cambridge University Press.
- Harper, D. (2001-2012). *Online etymology dictionary*. Retrieved from <http://www.etymonline.com/>
- Honda, H., & Yamagishi, K. (2009). Perceived certainty based on verbal probability phrases: Effect of directionality and its dependence on method. *Japanese Psychological Research, 51*, 266-273. doi: 10.1111/j.1468-5884.2009.00409.x
- Huang, Y., & Wang, L. (2010). Sex differences in framing effects across task domain. *Personality and Individual Differences, 48*, 649-653. doi: 10.1016/j.paid.2010.01.005
- Juanchich, M., Sirota, M., & Butler, C. L. (2012). The perceived functions of linguistic risk quantifiers and their effect on risk, negativity perception and decision making. *Organizational Behavior and Human Decision Processes, 118*, 72-81. doi: 10.1016/j.obhdp.2012.01.002
- Juanchich, M., Teigen, K. H., & Gourdon, A. (2010, November). *Variants of uncertainty*. Paper presented at the 31st Annual Conference of the Society for Judgment and Decision Making, St. Louis, MO.
- Juanchich, M., Teigen, K. H., & Villejoubert, G. (2010). Is guilt 'likely' or 'not certain'? Contrast with previous probabilities determines choice of probability phrases. *Acta Psychologica, 135*, 267-277. doi: 10.1016/j.actpsy.2010.04.016
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697-720. DOI: 10.1037/0003-066X.58.9.697

- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 263–291. doi:10.2307/1914185
- Karmarkar, U., & Tormala, Z. (2010). Believe me, I have no idea what I'm talking about: The effects of source certainty on consumer involvement and persuasion. *The Journal of Consumer Research*, *36*, 1033-1049. doi: 10.1086/648381
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*, 190-193. DOI: 10.1111/1467-8284.00419
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers *Child Development*, *76*, 1261-1277. DOI: 10.1111/j.1467-8624.2005.00849.x
- Kwan, V. S. Y., Wojcik, S. P., Miron-shatz, T., Votruba, A. M., & Olivola, C. Y. (2012). Effects of symptom presentation order on perceived disease risk. *Psychological Science*, *23*, 381-385 . doi:10.1177/0956797611432177
- Leshner, A. (2010, June 29). [Letter to Giorgio Napolitano, President of the Republic of Italy]. Retrieved from http://www.aaas.org/news/releases/2010/media/0630italy_letter.pdf
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX [Data file]. Retrieved from <http://leadserv.u-bourgogne.fr/bases/manulex/manulexbase/index.htm>
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, *36*, 156-166. DOI: 10.3758/BF03195560
- Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, *9*, 563-564.

Likelihood (2012). In *Oxford Dictionaries Online*. Retrieved from

<http://oxforddictionaries.com/definition/english/likelihood?q=likelihood>

Mascaro, O., & Sperber, D. (2009). The moral, epistemic and mindreading components of children's vigilance towards deception. *Cognition*, 112, 367-380.

doi:10.1016/j.cognition.2009.05.012

McGlone, M. S., & Reed, A. B. (1998). Anchoring in the interpretation of probability expressions. *Journal of Pragmatics*, 30, 723-733. doi: 10.1016/S0378-

2166(98)00011-3

McKenzie, C. R., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, 10,

596-602. doi: 10.3758/BF03196520

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57-73. doi: 10.1017/S0140525X10000968

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a

representational theory of mind. *Child Development*, 61, 722-730.

doi: 10.1111/j.1467-8624.1990.tb02815.x

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, 5, 2-12. doi:10.1214/ss/1177012242

Moxey, L. M., & Sanford, A. J. (1986). Quantifiers and focus. *Journal of Semantics*, 5, 189-206. doi: 10.1093/jos/5.3.189

Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*,

14, 237-255. DOI: 10.1002/(SICI)1099-0720(200005/06)14:3<237::AID-ACP641>3.0.CO;2-R

Mullet, E., & Rivet, I. (1991). Comprehension of verbal probability expressions in children and adolescents. *Language & Communication, 1991*, 217-225. doi: 10.1016/0271-5309(91)90007-I

Nakao, M., & Axelrod, S. (1983). Numbers are better than words. Verbal specifications of frequency have no place in medicine. *American Journal of Medicine, 74*, 1061–1065. doi: 10.1016/0002-9343(83)90819-7

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language, 63*, 324-346. doi: 10.1016/j.jml.2010.06.005

Nurmsoo, E., and Robinson, E. J. (2009). Children's trust in previously inaccurate informants who were well or poorly informed: When past errors can be excused. *Child Development, 80*, 23–27. doi: 10.1111/j.1467-8624.2008.01243.x

Odds (2012). In *Oxford Dictionaries Online*. Retrieved from <http://oxforddictionaries.com/definition/english/odds?q=odds>

Office National for Statistics (2011). *Deaths registered in England and Wales in 2010, by cause*. Retrieved from <http://www.ons.gov.uk/ons/rel/vsob1/mortality-statistics--deaths-registered-in-england-and-wales--series-dr-/2010/stb-deaths-by-cause-2010.html>

Pellizzoni, S., Siegal, M., & Surian, L. (2009). Foreknowledge, caring, and the side-effect effect in young children. *Developmental Psychology, 45*, 289-295 . doi: 10.1037/a0014165

- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*, 407–413. doi: 10.1111/j.1467-9280.2006.01720.x
- Piaget, J., & Inhelder, B. (1951). *La genèse de l'idée de hasard chez l'enfant* (The origin of the idea of chance in children). Paris: Presses Universitaires de France.
- Piaget, J., & Inhelder, B. (1966). *La psychologie de l'enfant* (Child psychology). Paris: Presses Universitaires de France.
- Probability (2012). In *Collins English Thesaurus*. Retrieved from <http://www.collinsdictionary.com/dictionary/english-thesaurus/probability>
- Probability (2012). In *Oxford Dictionaries Online*. Retrieved from <http://oxforddictionaries.com/definition/english/probability?q=probability>
- Quattrone, G. A., & Tversky, A. (1988). Contrasting rational and psychological analyses of political choice. *The American Political Science Review, 82*, 719-736.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510-532. doi: 10.1037/0033-2909.114.3.510
- Reagan, R. T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology, 74*, 433. doi: 10.1037/0021-9010.74.3.433
- Reyna, V. F. (1981). The language of possibility and probability: Effects of negation on meaning. *Memory and Cognition, 9*, 642-650. doi: 10.3758/BF03202359
- Robinson, E. J., Pendle, J. E. C., Rowley, M. G., Beck, S. R., & McColgan, K. L. T. (2009). Guessing imagined and live chance events: Adults behave like children with live

events. *British Journal of Psychology*, 100, 645-659.

doi: 10.1348/000712608X386810

Risk (2012). In *Oxford Dictionaries Online*. Retrieved from

<http://oxforddictionaries.com/definition/english/risk?q=risk>

Ropeick, D. (2012, October 22). The L'Aquila verdict: A judgment not against science, but

against a failure of science communication [Web log message]. Retrieved from

[http://blogs.scientificamerican.com/guest-blog/2012/10/22/the-laquila-verdict-a-](http://blogs.scientificamerican.com/guest-blog/2012/10/22/the-laquila-verdict-a-judgment-not-against-science-but-against-a-failure-of-science-communication)

[judgment-not-against-science-but-against-a-failure-of-science-communication](http://blogs.scientificamerican.com/guest-blog/2012/10/22/the-laquila-verdict-a-judgment-not-against-science-but-against-a-failure-of-science-communication)

Sanford, A. J., Fay, N., Stewart, A., & Moxey, L. (2002). Perspective in statements of quantity,

with implications for consumer psychology. *Psychological Science*, 13, 130-134. doi:

10.1111/1467-9280.00424

Sanford, A. J., & Moxey, L. M. (2003). New perspectives on the expression of quantity.

Current Directions in Psychological Science, 12, 240-243. doi: 10.1046/j.0963-

7214.2003.01270.x

Savage, L. J. (1972). *The foundations of statistics* (2nd ed.). New York: Dover.

Sirota, M., & Juanchich, M. (2012). To what extent do politeness expectations shape risk

perception? Even numerical probabilities are under their spell! *Acta Psychologica*,

141, 391-399. doi: 10.1016/j.actpsy.2012.09.004

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010).

Epistemic vigilance. *Mind & Language*, 25, 359-393. doi: 10.1111/j.1468-

0017.2010.01394.x

Sperber, D., & Noveck, I. (2004). Introduction. In I. Noveck & D. Sperber (Eds.), *Experimental*

pragmatics (pp. 1-22). Basingstoke, United Kingdom: Palgrave Macmillan.

- Stupple, E. J., & Ball, L. J. (2011). Normative benchmarks are useful for studying individual differences in reasoning. *Behavioral and Brain Sciences*, *34*, 270-271. doi: 10.1017/S0140525X11000562
- Sutherland, H. J., Lockwood, G. A., Tritchler, D. L., Sem, F., Brooks, L., & Till, J. E. (1991). Communicating probabilistic information to cancer patients: Is there 'noise' on the line? *Social Science & Medicine*, *32*, 725-731. doi: 10.1016/0277-9536(91)90152-3
- Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*, 19156-19159. doi: 10.1073/pnas.0700271104
- Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*, 27-38. doi: 10.1016/0001-6918(88)90043-1
- Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, *88*, 233-258. doi: 0001-6918(93)E0071-9
- Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, *80*, 155-190. doi: 10.1006/obhd.1999.2857
- Teigen, K. H., & Brun, W. (2003). Verbal expressions of uncertainty and probability. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgement and decision making* (pp. 125-145). Chichester, UK: John Wiley & Sons.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232. doi: 10.1016/0010-0285(73)90033-9

- Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.
doi: 10.1037/0033-295X.90.4.293
- Utility (2013). In *Oxford Dictionaries Online*. Retrieved from
<http://oxforddictionaries.com/definition/english/utility?q=utility>
- Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*, 348. doi: 10.1037/0096-3445.115.4.348
- Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, *31*, 135-138.
- Watson, J. M., & Moritz, J. B. (2003). The development of comprehension of chance language: Evaluation and interpretation. *School Science and Mathematics*, *103*, 65-81. doi: 10.1111/j.1949-8594.2003.tb18222.x
- Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 781-789. doi:10.1037/0096-1523.16.4.781
- Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn, & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607-632). Oxford, United Kingdom: Blackwell.
- Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, *2*, 343-364. doi:10.1037/1076-898X.2.4.343

- Witteaman, H. O., Zikmund-Fisher, B. J., Waters, E. A., Gavaruzzi, T., & Fagerlin, A. (2011). Risk estimates from an online risk calculator are more believable and recalled better when expressed as integers. *Journal of Medical Internet Research, 13*, e54. doi:10.2196/jmir.1656
- Wright, J. C., & Bengson, J. (2009). Asymmetries in judgments of responsibility and intentional action. *Mind & Language, 24*, 24-50. doi: 10.1111/j.1468-0017.2008.01352.x
- Xu, J. H., Ye, X. B., & Li, S. (2009). Communication mode preference paradox among native Chinese speakers. *The Journal of Social Psychology, 149*, 125-130. doi: 10.3200/SOCP.149.1.125-130