Testimony – An Exploration of Our Trust in Others

by

Emma McLarens

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Cognitive Science

in

Cognitive Science

Carleton University

Ottawa, Ontario

Abstract

Whether it's in-person, over email, text or call, people constantly receive information from others. Epistemology views such information, referred to as testimony, as a crucial source of rational belief and knowledge. However, reductionists and anti-reductionists disagree on whether it is sufficient as a stand-alone source. This thesis presents three studies examining possible factors that influence our trust in the words of others. Participants were asked to rate how likely they thought the content of various reports were. Ratings differed depending on the type of information being reported, and the source of the report. There also appear to be differences associated with the gender of the source, but further testing should be done to get a better understanding of this result. Finally, participant level of education had no effect on ratings. Overall, the results of this thesis provide support for both the anti-reductionist and reductionist positions.

Acknowledgements

I would first like to thank my thesis advisor Dr. Ida Toivonen of the Institute of Cognitive Science at Carleton University. The door to Dr. Toivonen's office was always open whenever I ran into trouble or had a question about my research or writing. She consistently allowed this paper to be my own work, but steered me in the right direction whenever she thought I needed it.

I would also like to acknowledge Dr. David Matheson of the Department of Philosophy at Carleton University as the second reader of this thesis and Margot Shields from Statistics Canada for all her help with the data analysis of this research. I am gratefully indebted to her for her very valuable comments on this thesis.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Table of Contents

List of Tables

List of Illustrations

List of Appendices

1 Introduction

Epistemology is broadly defined as the theory of knowledge and justification (Audi, 1997; Stroll, 2013; Blackburn, 2016; Steup, 2016). Research in epistemology explores all aspects of belief and knowledge. For example: What is knowledge and how does it differ from belief? How do we come to believe something? Why do some things become knowledge and some things remain beliefs? What are the limits of knowledge (Audi, 1997; Stroll, 2013; Blackburn, 2016; Steup, 2016)?

In studying justification, epistemologists aim to distinguish between 'truth' and 'opinion' by examining how we understand the concept of justification itself; what makes beliefs justified (i.e., acceptable reasoning); and whether justification is internal or external to the mind (Audi, 1997; Stroll, 2013; Blackburn, 2016; Steup, 2016). Key issues of debate include: experiences in generating knowledge, changing forms of knowledge and the possibility of universal skepticism (Audi, 1997; Blackburn, 2016; Steup, 2016).

A key element of the study of justification involves the notion of testimony, which is the main focus of this thesis. Testimony is the intentional transfer of information from one person to another (Lackey, 2008; Pritchard, 2016). This transfer can be verbal, written or communicated in some other way (Coady, 1992; Lackey, 2008; Lackey, 2010). *Natural testimony* is encountered in everyday, social circumstances (Coady, 1992; Lackey 2008). It can be something as simple as giving someone directions, telling someone the weather forecast, or recounting to someone a news story you heard on the way to school. In contrast, *formal testimony* is a statement offered under oath in a courtroom or commission of inquiry (Coady, 1992; Lackey 2008). For the purpose of this thesis, I will be referring to testimony in the 'natural' sense.

For my thesis, I will explore how testimony generates knowledge by examining factors that influence how people (referred to as the 'hearers') come to believe things (i.e. information) on the basis of others (referred to as the 'speakers') reporting these things. In particular, my study will explore factors related to (1) the type of information being reported, (2) who is reporting the information, and (3) the hearer's level of education.

1.1 Testimony

The purpose of testimony is the formation of belief in the hearer. By asserting (or otherwise communicating) a proposition, the speaker aims to get the hearer to believe that proposition (Coady, 1992; Lackey, 2008; Lackey, 2010; Adler, 2012). For example, if I tell you that there is a freezing rain warning for the Ottawa area this afternoon, I am extending an invitation to you to believe this information.

Contemporary epistemology regards testimony as a crucial source of rational belief and knowledge, and is grounded in the 18th century works of Thomas Reid and David Hume (Hardwig 1985; Coady, 1992; Audi, 1997; Elgin, 2002; Lackey, 2006; Lackey, 2008; Origgi, 2008; Lackey, 2010; Adler, 2012; Davies & Matheson, 2013). In 1764, Reid argued that humans receive the majority of their knowledge from the reports of others (Reid, [1764]; Davies & Matheson, 2013). Thirteen years later, Hume argued that not only is human testimony the most common source of reasoning, it is also the most useful (Hume, [1777]; Davies & Matheson, 2013). In addition, Hume argued that eyewitness reports are necessary to human life (Hume, [1777]; Davies & Matheson, 2013). Further to these arguments, if testimony was not considered a source of rational belief then many (if not all) of our beliefs would not be considered rational knowledge (Hardwig 1985; Coady, 1992; Audi, 1997; Elgin, 2002; Lackey, 2006; Adler, 2012; Davies & Matheson, 2013).

A comparison of two opposing views of testimony, anti-reductionist and reductionist, will be provided (Reid, 1764; Hume, 1777; Lackey, 2005; Lackey, 2006; Adler, 2012). Thomas Reid argued for the anti-reductionist position, which postulates that testimony is a basic source of justification and is credible on its own until shown otherwise (Reid, 1764; Lackey, 2005; Adler, 2012). Simply stated, this means that someone's testimony is acceptable without any outside information (such as previous knowledge about the speaker, the topic, personal experience) (Reid, 1764; Lackey, 2005; Adler, 2012). For example, if I tell you that 'The Bay is having a huge sale', under normal conditions it is acceptable for you to believe my assertion unless you have special reason to object (e.g. you were at the Bay yesterday and were told the sale ended last night). Reid (1764) argued that God intended humans to be social creatures and therefore equipped us with a disposition to tell the truth and to believe in the honesty of others and the truthfulness of what they say (Reid, [1764]; Adler, 2012).

In my opinion, this idea is echoed in Paul Grice's proposed conversational rules, known as the Gricean Maxims (Grice, 1975; Levinson, 1983; Crystal, 1985; Grice, 1989). When followed, these four rules ensure effective communication between people. They are known as the Maxims of Quantity, Quality, Relation and Manner (Grice, 1975). The Maxim of Quantity postulates that a speaker's contribution to the conversation must be informative, but must not provide more or less information than needed. The Maxim of Quality postulates that the speaker must try to make their contribution to the conversation one that is true, and they should not say something they believe is false or lack adequate evidence for. The Maxim of Relation postulates that the speaker must only say what is relevant to the conversation. Finally, the Maxim of Manner is composed of several submaxims including be brief, orderly, and clear (Grice, 1975; Levinson, 1983; Crystal, 1985; Grice, 1989).

Further to this, it is argued that as hearers we automatically assume that speakers are following the maxims while we are engaged in conversation (Grice, 1975; Levinson, 1983; Grice, 1989). If a speaker blatantly disobeys these rules the hearer will find ways to make what is said fit within the confines of the Maxims. For example, when taken literally, if you ask me "Where are the cookies?" and I respond with "I was hungry", my response clearly violates the Maxims of Quantity and Relevance. While it appears that my response is unco-operative (I am refraining from answering your question by changing the subject), you will assume that I am co-operating and search for possible connections between what you asked and what I said. While people do not always adhere to these maxims, the hearers will interpret what is said as conforming to the maxims on some level whenever possible (Grice, 1975; Levinson, 1983; Grice, 1989). In the cookie example, the hearer will likely conclude that because I was hungry I ate the cookies. Origgi (2008) argues that in conversation speakers adjust their language to give the hearer access to only certain information, while hearers adjust their interpretation of that information to meet their pragmatic expectations and cognitive needs. Similar to Reid's view, Grice's Maxims suggest that humans believe other people are honest and trustworthy.

Reid's view is also echoed by the Knowledge Norm of Assertion. The Knowledge Norm is a proposal that postulates speakers should only say things they know or represent themselves as knowing (Adler, 2012; Benton, 2018). For example, if I tell you that an elephant's oversized ears are used to radiate excess heat away from the body, I am representing that I know this and should only say it if it is true.[1] In the absence of contradictory information, a speaker's assertion of the statement 'P' is sufficient (in and of itself) to generate knowledge for the hearer (Adler,

[1] I learned this information about elephants from an article online written by Noormohamed (2014).

2012). In addition to Reid's view, this echoes Grice's conversation rules and more specifically the Maxim of Quality.

In contrast, Hume's reductionist view holds that testimony by itself is not an acceptable source of knowledge; hearers need sufficiently good evidence to accept the report (Hume, 1777; Lackey, 2005; Lackey, 2006; Adler, 2012; Matheson, 2016). More specifically, an individual's belief in a report depends on supporting information from sources such as perception, memory and induction, as influenced by his/her experience with different communities (e.g. friends, family, co-workers, strangers, etc.) **(**Hume, 1777; Lackey, 2005; Lackey, 2006; Adler, 2012; Matheson, 2016). Generally speaking, Adler (2012) argues that 'hearers' know very little about 'speakers' so offering and accepting testimony cannot be considered a true form of knowledge acquisition.

Similar to this, Matheson (2016) argues that the formation of beliefs is caused by a connection between the content of the report and cognitive events (e.g. memory, induction, perception) of the hearer. This supports the work of Sperber (2010) as he explains that there are 'internal' and 'external' reasons that lead someone to accept a belief. Internal reasons are related to the content of the belief and some form of evidence in favour of the argument (Sperber, 2010). For example, I believe the water has boiled because I can see bubbles and steam coming from the pot and science has shown that this is what happens to water when it heats up. External reasons for accepting a belief are related to its source (Sperber, 2010). For example, I believe that I have strep throat because my doctor told me I do, and I trust my doctor. If a stranger on the bus told me I had strep throat I would have no external reason to believe them. Of note, external reasons can be influenced by internal reasons reflectively or intuitively. For example, I may trust my

doctor because he/she has never misdiagnosed me before, or because my parents say he/she is a good doctor and I trust my parents (Sperber, 2010).

The spread of knowledge through testimony is common (Reid, [1764]; Hume, [1777]; Hardwig 1985; Coady, 1992; Audi, 1997; Goldman, 1999; Elgin, 2002; Lackey 2006; Origgi, 2008; Davies & Matheson, 2013). However, as argued by the reductionist view, it is not always considered effective (Graham, 1997; Lehrer, 2006). The *Vulnerability Problem of Epistemological Testimony* occurs when the hearer's only reason for believing something is the speaker's word (Goldman, 1999; Adler, 2012). By trusting the speaker, the hearer assigns some kind of authority to him/her because they believe that person is more knowledgeable and will convey the relevant information (Goldman, 1999; Origgi, 2008; Adler, 2012). The hearer however may not always have good reason to ascribe authority to the speaker (Goldman, 1999; Origgi, 2008; Adler, 2012). This is primarily an issue for the anti-reductionist view.

As identified by Faulkner (2000, 2002), Siegel (2005) and Adler (2012), and in line with the reductionist perspective, there are several conditions hearers use to determine the trustworthiness of the testimonial setting and the speaker's credibility, including reputations and sanctions, impersonal knowledge of informants, motivation, social-moral bonds, co-operation, and prior plausibility. For example, the reputation of individuals and society's social standards are forceful constraints, especially in small communities and institutional settings. As previously noted, fundamental values, honesty and concern for others are instilled in humans through standard moral development, upbringing, and education. These values are easily learned because of the advantage and mutual benefit of receiving accurate information. As observed by Burge (1993), lying for fun is craziness and, speakers only gain from transmitting false information occasionally. Finally, hearers are equipped with a powerful filtering device that

automatically sorts assertions to challenge or dismiss. This ability is founded by the large amount of well-formed beliefs that hearers bring to the testimonial setting (Faulkner, 2000; 2002; Siegel, 2005; Adler, 2012).

While the disagreement between reductionists and anti-reductionists on whether testimony is sufficient as a stand-alone source of knowledge continues, research on the matter provides evidence for both sides of the argument.

1.2 Factors That Influence our Trust in Testimony

Upon reflection, I believe I am someone who considers multiple factors before trusting the information I receive from others. On the one hand, if I don't know the topic being discussed, I am more likely to trust what the other person is saying. In addition, I am more likely to trust the word of someone I know personally than a stranger. That being said, I am also someone who takes in the information given to me by others, does my own personal research and then makes my own decision about that information. If I am confident in my self-assessment, then I trend towards a reductionist view of testimony. Previous research on the subject has shown that I am not alone.

While not the focus of my thesis, previous work on *formal testimony* has found multiple factors that influence how a juror interprets information being reported by the witness. Some of these factors include: vocal pitch, the posture of the speaker, use of hand gestures, smiling, eye contact, attractiveness, expertise and gender (Boyle, 2014; Nagle, 2016; Gilmore, 2016, 2017). Male witnesses were consistently perceived as friendlier than females (Boyle, 2014; Nagle, 2016), and female witnesses were consistently doubted more than men (Gilmore 2016, 2017), so much so that lawyers prefer male witnesses (Boyle, 2014). In addition, Gilmore (2017) argues that people of colour, especially women of colour, are doubted more than anyone.

In every day conversation, Adler (2012) proposes several cues that hearers use to assess the credibility of a source and guide them to trust (or not trust) the information they are hearing:

1. If the speaker's testimony is a single sentence this reduces any justification the testimony may receive by being part of a group of mutually-supportive assertions.
2. When there is only one speaker any justification coming from corroboration with speakers is eliminated.
3. Whether the context of the report and the environment itself is one where the truthfulness norm holds, and the purpose is to inform.
4. Whether the testimony sustains the corresponding belief of the hearer.
5. Whether the speaker is assumed not to have 'expert knowledge' on the topic.
6. If the hearer has no special knowledge about the speaker.

In addition, it's argued that the biases of the speaker may affect their credibility (Ballantyne, 2015). For example, if it's known that the speaker dislikes a certain actor; his/her opinion of movies containing that actor will be less believable than someone with no strong feelings towards that actor. These conditions are important because they suggest that speakers must fulfill at least some of the criteria before being seen as trustworthy by the hearer. In addition, throughout the exchange, hearers are also using their perception of the environment to determine whether the speaker meets these criteria. If this is the case, the word of the speaker is not a source of knowledge by itself; there are many factors at play. Some of these criteria will be further explored below.

Previous work on testimony not associated with the courtroom setting has argued that individuals rely more on the word of authority (or expert) figures than non-authority figures (Origgi, 2008; Sperber 2010). Blackburn (2006) defines *authority* as a person, institution, or

organization that exercises legitimate power. This power is authorized by some system of norms to which the speaker adheres to (Blackburn, 2016). Walton (1997, 2006) created a list of conditions to determine whether or not an appeal to authority is rational. These conditions include the competence, credibility, sincerity and prestige of the expert; accuracy, evidence and verifiability of the report; and relation of the expert's opinion to the field of expertise (Walton, 1997, 2006; Wagemans, 2011). Similar to Alder's (2012) list of conditions individuals use to isolate dependence on the word of the speaker, Walton (2006) proposes six critical questions for the appeal to expert opinion (Wagemans, 2011):

1. How credible is the speaker as an expert source?
2. Is the speaker an expert in the field that he/she is reporting?
3. What did the speaker assert that implies what they are reporting?
4. Is the speaker reliable as a source?
5. Is what the speaker reporting consistent with what other experts are asserting?
6. Are the speaker's reports based on evidence?

In a study discussed by Origgi (2008), participants were asked to read, and judge information reported by experts and non-experts. This study found that when asked to explain a passage they read, participants recounted very different information depending on the source of the report; when the source of the passage was an expert, participants spent more time and effort to make sense of the information than they did when the source was a non-expert (Origgi, 2008; Sperber, 2010). Sperber (2010) argues that our trust in expert figures is influenced by the confirmation bias; the tendency to interpret new information in a way that it confirms our existing beliefs. This is important because I am trying to experimentally test how speakers evaluate different types of reports. More specifically, I am trying to test whether it makes a difference if someone is an expert or not.

Fricker (2007) extends this idea of authority to her work on gender. She argues that women are often victims of testimonial injustice, a form of epistemic injustice whereby prejudice causes a hearer to give a deflated level of credibility to a speaker's word (Fricker, 2007). For example, if a male and a female tell the exact same story, the female will not always be ascribed the same authority as that male, making her word less believable. This type of injustice is important to the study of testimony because it demonstrates how hearers use social stereotypes in their assessment of the speaker's credibility (Fricker, 2008). Additionally, it is not limited to gender differences. Anyone fitting within a stereotype(s) can experience injustice (Fricker, 2008). For example, border security may not believe someone because they are not fluent in English.

This work is important because I will be exploring the effect of the source of information on the likelihood ratings of reports.

2 Experiments

These studies will explore possible factors that affect our trust in the words of others. More specifically, they will examine factors related to (1) the type of information being reported (i.e. subjective opinions versus objective facts), (2) who is reporting the information (i.e. experts versus non-experts), and (3) the hearer's level of education.

2.1 Pilot Study

The purpose of the pilot study, modeled after Lesage et al. (2015), was to check whether the selected methodology was likely to teach us anything about different types of testimony and testimonial sources. The study was designed to analyse whether people evaluate testimony differently based on (1) the type of testimony provided (i.e. subjective opinion or objective fact), (2) the source of the testimony (i.e. who is reporting this information), and (3) the participant's

education level. I also wanted to make sure there were no obvious problems with the study's instructions or the stimuli.

Method

Participants

Participants (N = 180) were recruited from the Carleton University Cognitive Science SONA system and by personal invitations spread on social media and via email. Before analysis, twenty-four participants were excluded because they were not native speakers of English (N = 16) or because they completed less than half of the survey (N = 8).[2] The remaining participants (N = 156) included male (N = 49) and female (N = 107) students and non-students, 16 years of age or older. Participants with access to the Carleton University Cognitive Science SONA system are Carleton University students who are currently enrolled in the first year cognitive science course CGSC 1001 - Mysteries of the Mind. However, note that this course is not restricted to first year students. Participants recruited directly via this system were compensated a 0.25 grade percent in CGSC 1001 for their participation. Participants recruited via social media and email did not receive any form of compensation.

Materials

Participants were asked to complete an online survey hosted on Qualtrics, a web-based survey tool.[3] The survey consisted of forty scale-based questions (twelve target stimuli and twenty-eight filler stimuli), took approximately fifteen minutes to complete and was completely

[2] Because this is a language based study, non-native speakers of English were excluded from analysis to eliminate potential language-related issues.

[3]The data for this thesis was generated using Qualtrics software, Version 052017 of Qualtrics. Copyright © 2018 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com

anonymous. Stimuli were randomized to prevent ordering effects. The data was collected between January 2017 and April 2017.

At the beginning of the survey, participants answered a background questionnaire about their gender, education level, native language and age (see Appendix C). Next, participants were presented with pairs of sentences and asked to judge how likely it was that the second sentence was true, given that the first sentence was true. The likelihood of the second sentence being true was reported along a seven-point scale, with 1 being neither likely nor unlikely, and 7 being extremely likely. All target sentences were of the form "*Someone told someone…"* (e.g., *Anna told Peter…*) or of the form *"Someone said to someone…"* (e.g., *Anna said to Peter…*).

Twelve target stimuli were divided into three categories: subjective opinion, objective fact, and expert report. Each category comprised four of the twelve target stimuli.

1. *Subjective opinion* examples all involved a person stating their opinion about something that cannot be easily verified as objectively true or false. For example: Assume that sentence (A) is true. Please rate the likelihood that sentence (B) is true on a scale from 1 (neither likely nor unlikely) to 7 (extremely likely).

   (A) Kate said to Bob that she was cold.

   (B) Kate was cold.

2. *Objective fact* examples all involved statements that can be verified as either true or false. For example:

   (A) Jesse said to Ray that the film won a Grammy award.

   (B) The film won a Grammy award.

3. *Expert report* examples involved *objective fact* statements made by a person with expertise on the topic of the statements. For example:

(A) The doctor told Tim that his son has strep throat.

(B) Tim's son has strep throat.

Twenty-eight filler questions which had a similar structure were used. These questions were focused around the five human senses (sight, smell, hearing, taste, and touch).

Example 1:

(A) It smells like Henry has been baking banana bread.

(B) Henry has been baking banana bread.

Example 2:

(A) It looks like Greg has been crying.

(B) Greg has been crying.

For the full list of stimuli (both target and filler pairs), see Appendix D.

Finally, before exiting the survey, an optional comment section was provided for participants to express any questions, concerns or thoughts they may have had while completing the survey.

This research project was cleared by the Carleton University Research Ethics Board-B (Clearance # 106091) before the study began (December 2016).

Procedure

Participants were provided with an online link to the survey either through the Carleton University Cognitive Science SONA System, social media postings or email. When the link was accessed participants were directed to the survey hosted on Qualtrics (Qualtrics, Provo, UT). Participants completed the survey on their own time using any device with internet access (computer, laptop, smart phone, etc.). Upon completion, participants who accessed the survey through SONA were automatically directed back to the Carleton University Cognitive Science

SONA system and awarded a 0.25 percent grade bonus for the course CGSC 1001. All other participants closed the browser upon completion.

Results

I began by testing descriptive statistics on the data to get an idea of whether or not there was an effect of testimony type on participant rating that needed to be investigated. Mean ratings showed that subjective opinion (M = 5.63, SD = 1.68, SE = 0.07) was rated slightly higher than objective fact (M = 5.52, SD = 1.41, SE = 0.06). In addition, when the source of the report was an expert on the topic, the reports were rated higher than both subjective opinion and objective fact (expert report: M = 6.28, SD = 1.11, SE = 0.04). The data for this study displayed a non-normal distribution: subjective opinion had a skewness of -1.31, objective fact had a skewness of -1.26 and expert report had a skewness of -1.78 (see **Figure 1**). The statistical analysis of the Pilot Study will therefore be based on percentiles.[4]

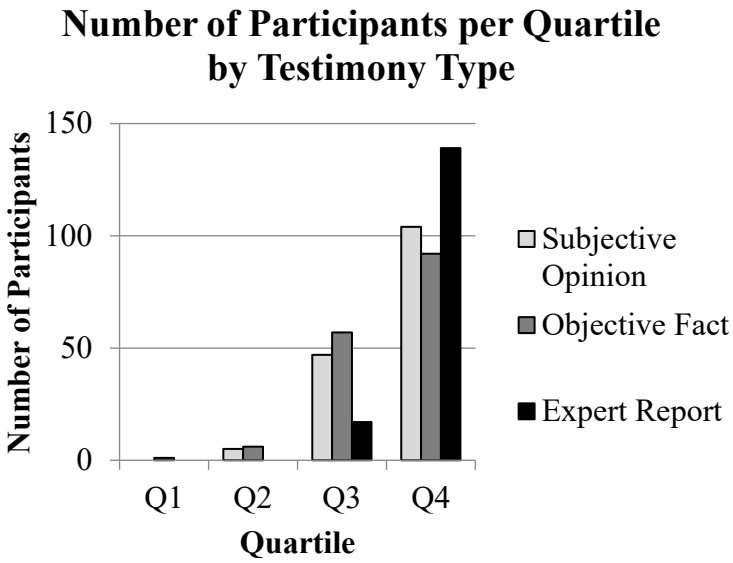

Figure 1. Number of participants within each quartile per type of testimony and source.

Each testimony type was tested with four stimuli. A total score (with a possible range of 0 to 28) was calculated by adding together the four scores. Higher scores indicated a greater likelihood that the participants judged the statements to be true. Scores for each participant were assigned to quartiles. Quartiles were generated by taking the maximum possible score (28) and multiplying it by 0.25, 0.5 and 0.75. Resulting quartiles were as follows: Q1: 0 to 7, Q2: 8 to 14,

[4] The choice to run my statistics this way was made in consultation with Margot Shields. Mrs. Shields is an expert statistician with 30+ years of experience and 60+ publications with Statistics Canada. For examples of similar methods, see Shields & Wilkins (2009).

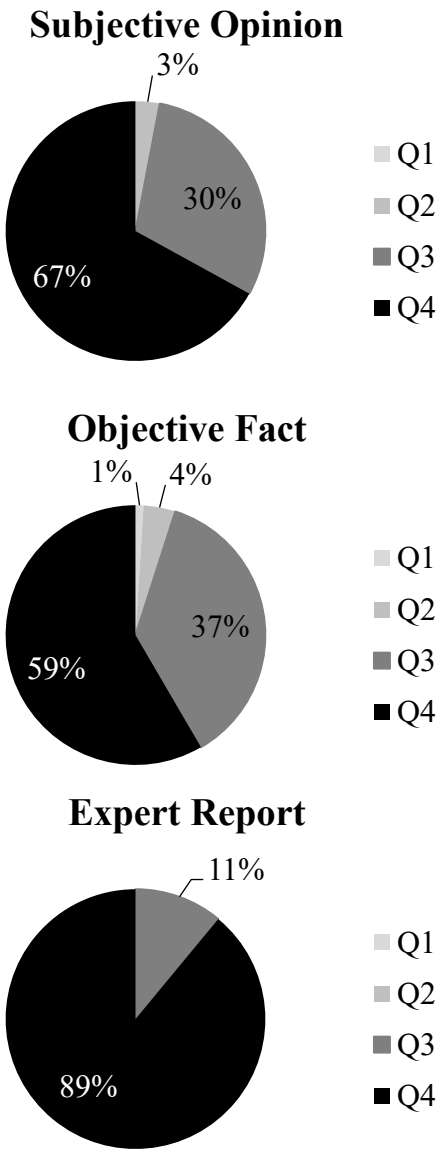

Figure 2. The percentage of participants who fell within each quartile for reports of subjective opinion, objective fact and expert reports.

Q3: 15 to 21, and Q4: 22 to 28. The quartiles can be more easily understood as: Q1 – low likelihood of truth, Q2 – moderate likelihood of truth, Q3 – high likelihood of truth, and Q4 – very high likelihood of truth. This is representative of the 7-point scale used throughout the study. To maximize the number of participants for whom scores were calculated, one response left empty per testimony type was accepted. A score was calculated based on the items with responses and then adjusted by using the mean substitution technique to compensate for the item without a response. The mean substitution method was used for six participants.

Overall, 22% of the participants' scores fell within Q3 and the remaining 78% scored within Q4. This is interesting to note because it supports the anti-reductionist view of testimony by suggesting that participants are more likely to believe a report, regardless of source or subject matter, than to have no opinion at all. I conducted further analyses on the data which demonstrated a possible effect of testimony type (i.e. the type of report provided) on how high (or low) participants rated the likelihood of the report. Approximately 67% of the participants fell within Q4 when it came to reports of subjective opinion and 59% for reports of objective fact

(see **Figure 2**). I conducted t-tests to analyze the rating differences between the two types of testimony and did not find a significant difference ($t(155) = 1.86$, $p = 0.06$). This result suggests that participants do not trust testimony differently depending on the type of information being reported (i.e. opinions versus fact). However, 89% of participants scored within Q4 when the source of the report was an expert on the topic. I conducted t-tests to compare expert reports with the two non-expert types and found that expert reports were rated significantly higher than both subjective opinion ($t(155) = -6.17$, $p < 0.00$) and objective fact ($t(155) = -7.35$, $p < 0.00$). This result suggests that expert reports are believed to be more trustworthy than the reports of non-experts.

Participants were then categorized by their reported education level for further analysis: high school (N = 27), college (N = 37), undergraduate university (N = 78) and graduate school (N = 15). Each education level includes participants who have completed or are currently enrolled in the specified program. For example, 'high school' includes participants who have received ONLY a high school diploma, as well as participants who are working towards completing their diploma. The graduate school category includes participants currently enrolled in Master and Doctorate programs.

Descriptive statistics do not suggest an effect of education level on testimony rating. Graduate students ($M = 5.45$, $SD = 1.89$, $SE = 0.24$) rated the sentences lower than university ($M = 5.66$, $SD = 1.55$, $SE = 0.09$), college ($M = 5.59$, $SD = 1.71$, $SE = 0.14$) and high school ($M = 5.67$, $SD = 1.87$, $SE = 0.18$) students, but the difference is minimal. When looking at this result in terms of quartiles, overall, 85% of high school students, 81% of college students, 74% of university students and 73% of graduate students scored in Q4. As demonstrated by **Table 1**,

when looking within the different types of testimony, the majority of participants fell within the fourth quartile.

Table 1. Percentage of participants by quartile and reported education level, for testimony type.

	Subjective Opinion				Objective Fact				Expert Report			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Education												
High School		4	23	73		4	38	58			4	96
College			40	60	1		35	64			4	96
University		2	30	68		6	34	59			13	87
Grad School		13	13	73			47	53			20	80

Of note, there appears to be an effect of education on the ratings of expert reports. While the percentage of participants in Q4 at each stage of education is similar for subjective opinion and objective fact, there appears to be a gradual reduction in likelihood ratings for the expert reports. University students (graduates and undergraduates) appear to rate expert reports lower than college and high school participants.

For my analysis, I conducted Pearson's product-moment correlations to identify whether a relationship between report ratings and participant level of education exists. More specifically, I wanted to examine whether the report ratings increased or decreased as participants attain higher levels of education. I did not find a significant difference overall ($r = -0.1$, $n = 154$, $p = 0.22$). In addition, no significant correlations were found between education level and acceptance of testimony for any type of report: subjective opinion ($r = -0.03$, $n = 154$, $p = 0.74$), objective fact ($r = -0.01$, $n = 154$, $p = 0.86$), expert report ($r = -0.14$, $n = 154$, $p = 0.08$). Next, I conducted ANOVAs and their results also supported this finding. No effect of participant education level was found for any type of testimonial report: subjective opinion ($F(3, 152) = 0.18$, $p = 0.91$, $n_p^2 = 0.00$), objective fact ($F(3, 152) = 0.04$, $p = 0.99$, $n_p^2 = 0.00$), expert report ($F(3, 152) = 1.06$, $p =$

0.37, $n_p^2 = 0.02$). In this study, I found no evidence that a participant's education level affects how critically participants assess the statements of others.

Discussion

Overall, the results of my Pilot Study indicate that the selected methodology could teach us something about my proposed research questions (i.e. different types of testimony and testimonial sources). In addition, the Pilot Study provided me with information that led to minor modifications to the methods and materials used for my study. For example, there will be changes to the target population, the number of target and filler stimuli and the demographic questions for Study 1.[5]

2.2 Study 1

The purpose of this survey was to collect information on people's attitudes towards different types of testimony and testimonial sources. The study was designed to test whether people evaluate testimony differently based on (1) the type of information being provided (subjective opinion or objective fact), (2) the source of the report (who is reporting this information), and (3) the participants' university year and major field of study. More specifically, this study will aim to shed light on how different types of reports are evaluated, how participants' evaluation of reports is changed based on who is reporting them, and whether university education influences how critically we assess the statements of others.

For Study 1, I hypothesize that (1) objective facts will be rated higher than subjective opinions even though the Pilot suggested participants rated subjective opinions higher than objective facts for the following reasons: the difference in ratings between these report types was not significant in the Pilot, I did not have as many participants in the Pilot as I do in Study 1 and

[5] See Study 1 Methods for a more detailed description of the changes made to the Pilot Study.

from personal experience I tend to trust people's opinions less than information that I can verify. In addition, I hypothesize that, similar to the Pilot Study, (2) expert reports will be rated higher than objective facts. And finally, I hypothesize that (3) as participants progress through university, they will be less likely to trust the given reports. That is to say, participants with more years of university education will rate reports lower than participants with fewer years of education. Although I didn't find a significant effect of education level on report ratings in the Pilot, after analysis of the data it was clear that some participants were answering the education question incorrectly, therefore it did not completely accurately represent the education level of participants.

Method

The methods of Study 1 are almost identical to the Pilot Study; however, a method section is included here to make this section self-contained.

Participants

Participants (N = 810) were recruited from the Carleton University Cognitive Science SONA system. I made the decision to recruit only Carleton University students for the participant pool because the majority of participants in the Pilot Study held some level of post-secondary education. It was also evident from the answers of the demographic questions in the Pilot that some participants were not answering the level of education according to their current level but according to the highest level of education they had completed. While this posed issues for the results of the Pilot Study, this modification should allow for the analysis of individuals as they progress through their undergraduate university studies. In addition, by restricting the target population in this way, I hope to be limiting some possible confounding variables, such as age and experience.

Before analysis, one hundred and eighty-four participants were excluded because they were not native speakers of English (N = 119) or because they completed less than half of the survey (N = 65).[6] The remaining participants (N = 626) included male (N = 186), female (N = 424) and other (N = 16) students, 16 years of age or older. All participants are Carleton University students who are currently enrolled in the first year cognitive science course CGSC 1001 - Mysteries of the Mind and have access to the Carleton University Cognitive Science SONA system. However, note that this course is not restricted to first year students. Participants were compensated a 0.25 grade percent in CGSC 1001 for their participation.

Materials

Participants were asked to complete an online survey hosted on Qualtrics (Qualtrics, Provo, UT). The survey consisted of thirty-six scale-based questions (nine target stimuli and twenty-seven filler stimuli), took approximately fifteen minutes to complete and was completely anonymous. Stimuli were randomized to prevent possible ordering effects. The data were collected between May 2017 and December 2017.

Participants began the survey by answering a background questionnaire about their gender, age, education level and native language. I added two follow-up questions about participants' year of university and program of study in response to changing the target population. This will allow us to study whether a participant's time spent in university has an effect on how they judge the testimony of others and whether different disciplines teach us to think differently about testimony. The question regarding education level was included as a formality and will no longer be used in the analysis of the data.

[6] Because this is a language based study, non-native speakers of English were excluded from analysis to eliminate potential language-related issues.

Next, participants were presented with sentence pairs and asked to judge how likely it was that the second sentence was true, given that the first sentence was true. The likelihood of the second sentence was reported along a seven-point scale, with 1 being neither likely nor unlikely, and 7 being extremely likely.

Of note, the comment section of the Pilot Study highlighted potential issues identified by participants concerning the scale provided (1 = neither likely nor unlikely, 7 = extremely likely). Some participants expressed that they were confused at times and did not understand the scale. Others expressed that there were no options for unlikely, which they would have liked to be able to select for some of the questions. In reviewing the survey scale, it was felt that there are two issues that would arise if I were to switch to a different scale (e.g. 1 = extremely unlikely, 2 = unlikely, 3 = neither likely nor unlikely, 4 = likely, 5 = extremely likely). Firstly, none of the stimuli included examples where an "unlikely" or "extremely unlikely" option was appropriate. Secondly, research has shown that seven-point scales are more likely to reflect a participant's true subjective evaluation than a five-point scale as five-point scales are not sensitive enough to record differences (Finstad, 2010). Seven-point scales have been found to result in stronger correlations when using t-test analysis (Lewis, 1993) and are more suited for electronic distributions (Finstad, 2010). For these reasons, I made the choice to keep my original scale.

The comment section of the Pilot Study also revealed that some participants were trying to guess the purpose of the study. In order to avoid this in Study 1, I reduced the total number of target stimuli from twelve to nine. The nine target stimuli were divided into the same categories as the Pilot Study: subjective opinion, objective fact, and expert report. Each category now comprised three of the nine target stimuli and was of the form "*Someone told someone…"* (e.g., *Anna told Peter…*) or *"Someone said to someone…"* (e.g., *Anna said to Peter…*).

1. The *subjective opinion* examples involved a person stating their opinion about something that cannot be easily verified as objectively true or false. For example: Assume that sentence (A) is true. Please rate the likelihood that sentence (B) is true on a scale from 1 (neither likely nor unlikely) to 7 (extremely likely).

   (A) Karen told Bobby the concert was bad.

   (B) The concert was bad.

2. The *objective fact* examples involved statements that can be verified as either true or false. For example:

   (A) Megan told Tom that the office was closed.

   (B) The office was closed.

3. The *expert report* examples involved *objective fact* statements made by a person with expertise on the topic of the statements. For example:

   (A) The dentist told Mac that he has a cavity.

   (B) Mac has a cavity.

The filler stimuli for Study 1 were the same as the filler stimuli for the Pilot. These questions were focused around the five human senses (sight, smell, hearing, taste, and touch). In addition, sentence pairs where the responses would be clearly 1 or 7 were included to allow us to check whether participants were reading each question and not just clicking through the survey.

Example 1:

(A) It looks like a storm is starting.

(B) A storm is starting.

Example 2:

(A) Chad often has lunch with Judy.

(B) Chad had lunch with Judy yesterday.

Example 3:

(A) It is snowing.

(B) It is snowing.

Study 1 had twenty-seven filler stimuli in total. For the full list of stimuli see Appendix E.

Finally, before exiting the survey, an optional comment section was provided for participants to express any questions, concerns or thoughts they may have had while completing the survey.

For analysis, participants were categorized by their reported program of study using the Statistics Canada (2016) Major Field of Study (MFS) classification structure. This classification structure is composed of eleven major levels wherein each level is composed of several specific categorizations.

1. *Educational, Recreational and Counselling Services.*
2. *Fine and Applied Arts* includes music, theatre, graphic design, esthetics, etc.
3. *Humanities and Related Fields* includes history, languages, linguistics, philosophy, communications, religion, etc.
4. *Social Sciences and Related Fields* includes psychology, sociology, social work, law, political science, etc.
5. *Commerce, Management and Business Administration* includes marketing, administration, finance, and all other business related fields.
6. *Agricultural, Biological, Nutritional, and Food Sciences* includes biochemistry, biology, biophysics, zoology, vet technology, etc.

7. *Engineering and Applied Sciences* includes chemical, mechanical, civil and industrial engineering, forestry and mining.
8. *Applied Science Technologies and Trades* includes all trades at the college level.
9. *Health Professions and Related Technologies* includes nursing, dentistry, optometry, pharmacy and all other medical fields.
10. *Mathematics, Computer and Physical Sciences* includes computer science, mathematics, meteorology, physics, chemistry and actuarial science.
11. *No Specialization.*

I added a twelfth level to this classification system for my study because it is a relatively new field that falls within several of the MFS classifications. It is also my major field of study.

12. *Cognitive Science* is the multidisciplinary study of philosophy, psychology, linguistics, neuroscience, computer science and anthropology, falling within the Humanities, Social Sciences, Mathematics and Computer Science levels of the MFS.

Based on the distribution of the participants' reported university majors, five levels were excluded for having less than forty participants (Educational, Recreational and Counselling Services (N = 0), Fine and Applied Arts (N = 2), Engineering and Applied Science (N = 12), Applied Science Technologies and Trades (N = 0), Health Professions and Related Technologies (N = 1)).

This research project was cleared by the Carleton University Research Ethics Board-B (clearance # 106091) before the study began (May 2017).

Procedure

Participants were provided with an online link to the survey through the Carleton University Cognitive Science SONA System. When the link was accessed participants were

directed to the survey hosted on Qualtrics (Qualtrics, Provo, UT). Participants completed the survey on their own time using any device with internet access (laptop, smart phone, etc.). Upon completion, participants were automatically directed back to the Carleton University Cognitive Science SONA system and awarded a 0.25 percent grade bonus for the course CGSC 1001.

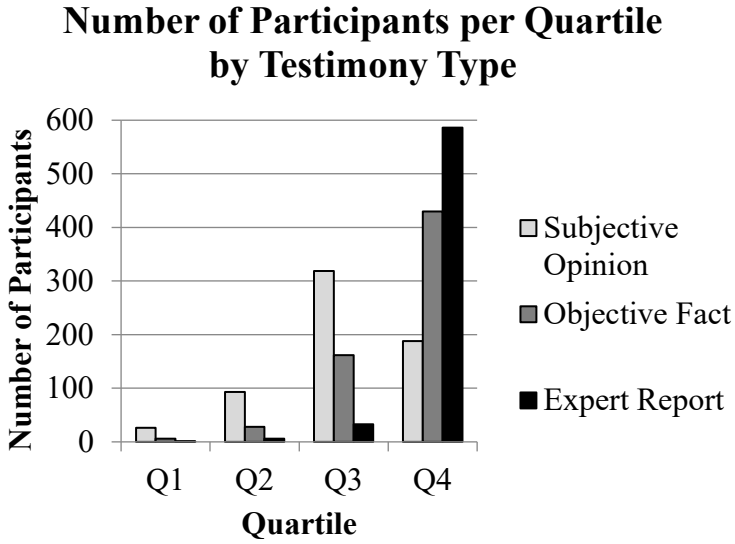

Figure 3. Number of participants within each quartile per type of testimony and source.

Results

On average, reports of subjective opinion (M = 4.53, SD = 1.61, SE = 0.04) were rated lower than reports of objective fact (M = 5.56, SD = 1.43, SE = 0.03). This finding is different from the Pilot Study but supports my hypothesis for Study 1. Similar to the Pilot Study expert reports (M = 6.52, SD = 0.94, SE = 0.02) were rated higher than both non-expert types: subjective opinion and objective fact. The results of this study displayed a non-normal distribution (see **Figure 3**): subjective opinion had a skewness of -0.61, objective fact had a skewness of -1.00 and expert reports had a skewness of -2.62. The statistical analysis of my study will therefore be based on percentiles.[7]

Each testimony type was tested with three stimuli. Scores for each participant were assigned to quartiles. Quartiles were generated by taking the maximum possible score (21) and multiplying it by 0.25, 0.5 and 0.75. Resulting quartiles were as follows: Q1: 0 to 5.25, Q2: 5.26

[7] The choice to run my statistics this way was made in consultation with Margot Shields. Mrs. Shields is an expert statistician with 30+ years of experience and 60+ publications with Statistics Canada. For similar examples of this method, see Shields & Wilkins (2009).

to 10.4, Q3: 10.5 to 15.74, and Q4: 15.75 to 21. The quartiles can be more easily understood as: Q1 – low likelihood that the statements were true, Q2 – moderate likelihood that the statements were true, Q3 – high likelihood that the statements were true, and Q4 – very high likelihood that the statements were true. This is representative of the 7-point scale used throughout the study. To maximize the number of participants for whom scores were calculated, one response left empty per testimony type was accepted. A score was calculated based on the items with responses and then adjusted by using the mean substitution technique to account for the item without a response. The mean substitution method was used for seventeen participants.

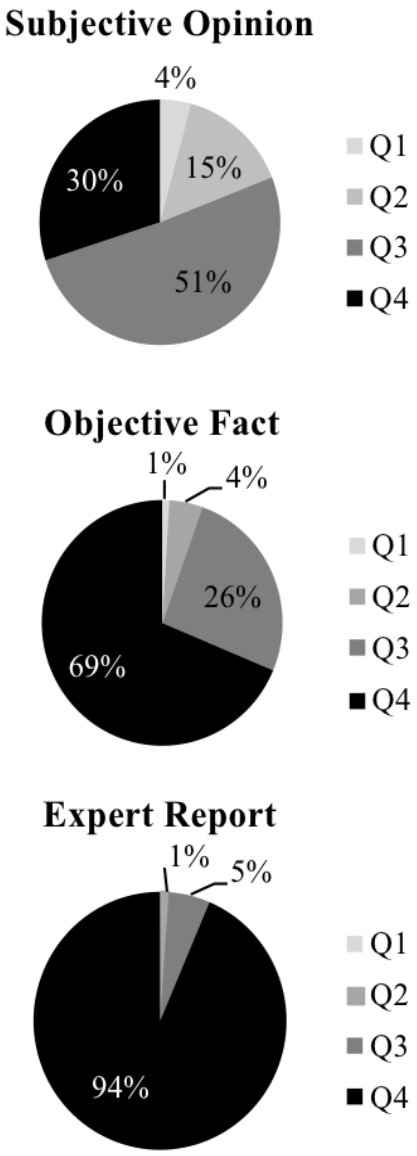

Figure 4. The percentage of participants who fell within each quartile for reports of subjective opinion, objective fact and expert reports.

Overall, 2% of participants fell within Q2, 30% fell within Q3, and 68% fell within Q4. No participants fell within Q1. This result confirms the result of the Pilot Study and suggests that participants are highly likely to believe the testimony of others, regardless of source or subject matter. When comparing across the different types of testimony, the data demonstrates possible effects of testimony type and source on how high (or low) participants rate the likelihood of statements being true. Only 30% of the

population fell within Q4 for subjective opinions, while approximately 69% of participants fell within Q4 when it came to objective facts, and 94% of the population fell within Q4 when it came to expert reports. I ran Welch's two sample t-tests to test the difference in ratings between testimony types and found a significant difference between reports of subjective opinion and objective fact ($t(625) = -16.79$, $p < 2.2^{-16}$). Recall that in the Pilot Study, subjective opinion was rated very slightly higher than objective fact but t-tests showed that the difference in rating was not significant. Study 1 however shows that participants rate objective facts significantly higher than subjective opinions. This result suggests that participants ranked the likelihood of statements differently according to the type of information being reported (i.e. participants ranked opinions lower than facts).

As in the Pilot Study, when the source of the report was an expert on the topic, the reports were ranked significantly higher than both types of non-expert reports; subjective opinion ($t(625) = -25.97$, $p < 2.2^{-16}$), objective fact ($t(625) = -11.99$, $p < 2.2^{-16}$). These results are interesting because they suggest that not only are expert reports considered more believable, but also people's trust is greater when the report is something that can be verified as true or false. Overall, these results indicate that (1) participants are more likely to trust certain types of testimony more than others and (2) participants trust reports differently depending on who is reporting the information.

While this was not one of my original research questions, upon analysis of Study 1 data, there appears to be a difference in testimony rating depending on the gender of who is reporting the information (i.e. the gender of the speaker). The following are the three subjective opinion stimuli used in Study 1:

1. (A) Karen told Bobby the concert was bad. (B) The concert was bad.

2. (A) Kyle told Sue the new restaurant was tasty. (B) The new restaurant was tasty.

3. (A) Scotty said to John that the exams were confusing. (B) The exams were confusing.

Descriptive statistics of the previous sentence pairs showed that the ratings for the sentence with Karen (M = 4.24, SD = 1.66, SE = 0.07) was lower than Kyle (M = 4.86, SD = 1.53, SE = 0.06) and Scotty (M = 4.5, SD = 1.57, SE = 0.06). It's unclear whether it makes sense to conduct inferential statistics on this data because there are many factors that could be at play (e.g. the content of the sentences, the type of the report, etc.), however I conducted them as a primary analysis that should be followed up on. Welch's t-tests between sentences 1 and 2 and between sentences 1 and 3 found that Karen was rated significantly lower than Kyle ($t(625) = -9.3$, $p < 2.2^{-16}$) , as well as Scotty ($t(625) = -4.22$, $p < 2.8^{-5}$). This result suggests that males are more likely to be trusted than females. However, t-tests also showed a significant difference between Kyle and Scotty ($t(625) = 5.55$, $p = 4.26^{-8}$). These results lead me to question whether the effect of gender shown is caused by who the speaker is or by the nature of the report (i.e. exams may be more relevant to my study population than new restaurants or concerts). I will examine this further in Study 2.

Keeping with the idea of gender differences, when the data was examined with respect to the participant's gender, there does not appear to be an effect on the level of trust in testimonial reports. Overall, 67% of males fell within Q4, while 69% of females fell within Q4. I ran an ANOVA on this result and did not find a significant difference, suggesting that male and female participants judge reports similarly ($F(2, 623) = 1.22$, $p = 0.29$, $n_p^2 = 0.00$). Additional ANOVAs were conducted to examine whether males and females rated the three types of testimony differently. Results of these tests found no significant effect of gender across testimony types: subjective opinion ($F(2, 623) = 0.12$, $p = 0.89$, $n_p^2 = 0.00$), objective fact ($F(2, 623) = 2.1$, $p =$

0.12, $n_p^2 = 0.01$), expert report (F(2, 623) = 0.86, p = 0.42, $n_p^2 = 0.00$). This result reinforces the finding that males and females rate reports similarly.

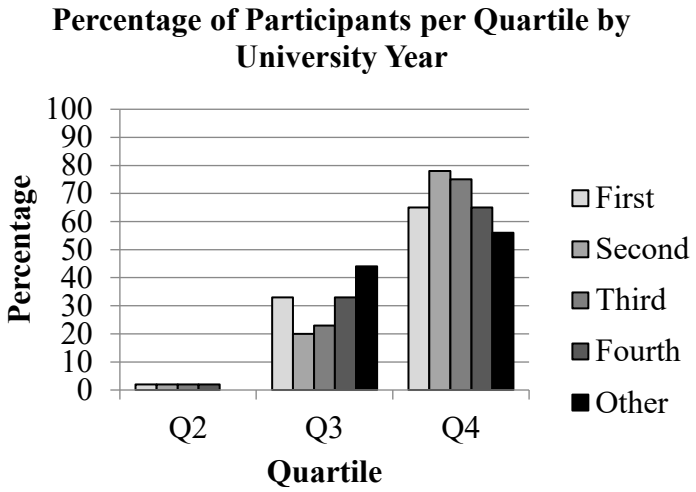

Figure 5. Percentage of participants by quartile and reported university year, for overall testimony ratings.

Next, I divided participants by their year of university study. Approximately 65% of first year students scored in Q4 while 79% of second year students, 75% of third year students, 65% of fourth year students and 56% of other students scored within Q4. Preliminary results from Pearson's product-moment correlation tests did not show a relationship between university year and testimony rating (r = 0.01, n = 624, p = 0.72). No significant correlations were found between university year and acceptance of testimony for subjective opinion (r = -0.03, n = 624, p = 0.42) or objective fact (r = 0.05, n = 624, p = 0.2). However, a positive correlation was found for expert reports (r = 0.1, n = 624, p = 0.00). I then used ANOVAs to test whether students at different levels of university (first year (N = 320), second year (N = 107), third year (N = 89), fourth year (N = 69)) judge reports differently. I removed participants who indicated that they had 'other' years of study (N = 41) from this part of the analysis because 'other' could mean many different levels/types of university education (e.g. fifth year student, special student, etc.) and I wanted to prevent any kind of effect this unknown may cause.[8] No significant effects of university level were found for any type of testimony: subjective opinion (F(3, 581) = 1.11, p = 0.34, $n_p^2 = 0.01$), objective fact

[8] I ran the same statistics with 'other' students included in the dataset and the results were only slightly different. No significant results were found for any type of report: subjective opinion (F(4, 621) = 1.66, p = 0.12, $n_p^2 = 0.01$), objective fact (F(4, 621) = 0.84, p = 0.5, $n_p^2 = 0.01$), expert report (F(4, 621) = 0.75, p = 0.56, $n_p^2 = 0.00$).

($F(3, 581) = 0.83$, $p = 0.48$, $n_p^2 = 0.00$), expert report ($F(3, 581) = 0.57$, $p = 0.63$, $n_p^2 = 0.00$). This result suggests that participants with higher levels of university do not approach reports more critically than participants with lower levels (see **Figure 5**). An additional break down of the percentage of participants for each quartile by university level is provided in **Table 2**.

Participants were then categorized using the MFS for further statistical testing (Humanities and Related Fields (N = 48), Social Sciences (N = 296), Commerce, Management and Business (N = 46), Agricultural, Biological, Nutritional and Food Sciences (N = 50), Mathematics, Computer and Physical Sciences (N = 65), No Specialization/Undeclared (N = 65), Cognitive Science (N = 41). Again, ANOVAs did not find any significant differences between university majors when it came to rating the likelihood of a report ($F(6, 604) = 1.08$, $p = 0.38$, $n_p^2 = 0.01$). Even when tested by individual types of testimony, the ANOVAs did not find any significant differences in report ratings among university majors: subjective opinion ($F(6, 604) = 0.68$, $p = 0.66$, $n_p^2 = 0.01$), objective fact ($F(6, 604) = 1.34$, $p = 0.24$, $n_p^2 = 0.01$), expert report ($F(6, 604) = 0.5$, $p = 0.81$, $n_p^2 = 0.00$). This suggests that regardless of differences in their major fields of study, participants rank the testimony similar to one another.

Table 2. Percentage of participants by quartile and reported university year, for testimony type.

	Subjective Opinion				**Objective Fact**				**Expert Report**			
University Year	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
First	5	15	50	30	1	4	28	67		1	7	92
Second	4	12	50	34	1	5	18	77	1	1		98
Third	1	10	57	31	1	2	26	71		2	6	92
Fourth	3	19	48	30	1	6	25	68			7	93
Other	7	22	49	22		7	32	61			2	98

In addition to these ANOVAs, I ran ANOVAs to test whether participants with the same major rated reports differently depending on their university level. No statistical significances were found for the subjective opinion reports; Humanities and Related Fields ($F(3, 42) = 1.39$, $p = 0.26$,

$n_p^2 = 0.09$), Social Sciences ($F(3, 276) = 0.34$, $p = 0.8$, $n_p^2 = 0.00$), Commerce, Management and Business ($F(3, 39) = 1.7$, $p = 0.18$, $n_p^2 = 0.12$), Agricultural, Biological, Nutritional and Food Sciences ($F(3, 43) = 0.34$, $p = 0.8$, $n_p^2 = 0.02$), Mathematics, Computer and Physical Sciences ($F(3, 59) = 0.53$, $p = 0.66$, $n_p^2 = 0.03$), No Specialization/Undeclared ($F(3, 48) = 0.21$, $p = 0.89$, $n_p^2 = 0.01$), Cognitive Science ($F(3, 35) = 0.66$, $p = 0.58$, $n_p^2 = 0.05$); the objective fact reports; Humanities and Related Fields ($F(3, 42) = 1.28$, $p = 0.3$, $n_p^2 = 0.08$), Social Sciences ($F(3, 276) = 0.27$, $p = 0.84$, $n_p^2 = 0.00$), Commerce, Management and Business ($F(3, 39) = 1.34$, $p = 0.28$, $n_p^2 = 0.09$), Agricultural, Biological, Nutritional and Food Sciences ($F(3, 43) = 1.03$, $p = 0.39$, $n_p^2 = 0.07$), Mathematics, Computer and Physical Sciences ($F(3, 59) = 1.4$, $p = 0.25$, $n_p^2 = 0.07$), No Specialization/Undeclared ($F(3, 48) = 0.46$, $p = 0.71$, $n_p^2 = 0.03$), Cognitive Science ($F(3, 35) = 0.87$, $p = 0.47$, $n_p^2 = 0.07$); or expert reports; Humanities and Related Fields ($F(3, 42) = 1.41$, $p = 0.25$, $n_p^2 = 0.09$), Social Sciences ($F(3, 276) = 1.45$, $p = 0.23$, $n_p^2 = 0.02$), Commerce, Management and Business ($F(3, 39) = 0.42$, $p = 0.74$, $n_p^2 = 0.03$), Agricultural, Biological, Nutritional and Food Sciences ($F(3, 43) = 1.65$, $p = 0.19$, $n_p^2 = 0.1$), Mathematics, Computer and Physical Sciences ($F(3, 59) = 0.46$, $p = 0.71$, $n_p^2 = 0.02$), No Specialization/Undeclared ($F(3, 48) = 0.7$, $p = 0.56$, $n_p^2 = 0.04$), Cognitive Science ($F(3, 35) = 0.16$, $p = 0.93$, $n_p^2 = 0.01$).[9] This result suggests that there is no interaction between major and university level. See **Table 3** for a more detailed comparison of rating averages for each type of testimony by participant major and university year.

Discussion

Overall, Study 1 showed that participants were more likely to trust reports than to have no opinion at all. Approximately 98% of participants scored in Q3 or Q4 for all types of testimony. This result supports the anti-reductionist view that humans are equipped with a

[9] I also conducted ANOVAs on the dataset as a whole to test whether there were interactions between any major and any university year. No significant differences were found for any type of report: subjective opinion ($F(27, 572) = 0.68$, $p = 0.89$), objective fact ($F(27, 572) = 0.71$, $p = 0.86$), expert reports ($F(27, 527) = 1.02$, $p = 0.44$).

Table 3. Mean ratings for each testimony type by participant major and year of study.

	Subjective Opinion	Objective Fact	Expert Report
Humanities			
First Year	M = 4.63	M = 5.87	M = 6.81
Second Year	M = 4.78	M = 5.67	M = 6.11
Third Year	M = 4.83	M = 6.33	M = 6.89
Fourth Year	M = 5.73	M = 6	M = 6.27
Social Sciences			
First Year	M = 4.45	M = 5.59	M = 6.48
Second Year	M = 4.57	M = 5.56	M = 6.72
Third Year	M = 4.58	M = 5.47	M = 6.47
Fourth Year	M = 4.33	M = 5.26	M = 6.59
Commerce/Business			
First Year	M = 4.32	M = 5.72	M = 6.51
Second Year	M = 5.04	M = 6.15	M = 6.48
Third Year	M = 5.33	M = 6.11	M = 6.77
Fourth Year	M = 4.89	M = 5.61	M = 6.67
Biological Sciences			
First Year	M = 4.35	M = 5.38	M = 6.65
Second Year	M = 4.43	M = 5.7	M = 6.4
Third Year	M = 5	M = 5.81	M = 6
Fourth Year	M = 5	M = 6.67	M = 6.83
Math/Computer & Physical Science			
First Year	M = 4.27	M = 4.98	M = 6.28
Second Year	M = 4.79	M = 5.64	M = 6.71
Third Year	M = 4.61	M = 5.81	M = 6.42
Fourth Year	M = 4.67	M = 5.38	M = 6.54
Cognitive Science			
First Year	M = 4.64	M = 5.31	M = 6.59
Second Year	M = 4.67	M = 6.25	M = 6.58
Third Year	M = 4.58	M = 5.5	M = 6.67
Fourth Year	M = 5.33	M = 5	M = 6.67
No Specialization			
First Year	M = 4.65	M = 5.6	M = 6.52
Second Year	M = 4.41	M = 6.04	M = 6.89
Third Year	M = 4.27	M = 5.67	M = 6.48
Fourth Year	M = 4.76	M = 5.66	M = 5.95

disposition to trust others because it demonstrates people's trust in what others say regardless of the source or type of report. However, the results might in general be argued to provide stronger support for the reductionist view because there was notable variation in the number of responses between Q3 and Q4 (30% versus 68%). This shows that participants assigned more or less credibility to sources based on the small amount of information they received (e.g. a name or profession) and in turn made a choice to trust or not trust the information.

As hypothesized, Study 1 found that participants are significantly more likely to trust objective facts than subjective opinions. Moreover, participants are more likely to trust reports from experts than non-experts. In my opinion, this is a believable result because children are brought up (for the most part) to respect expert figures. This finding is similar to our Pilot Study and supports the positions of Origgi (2008) and Sperber (2010). These results are important because they demonstrate the importance of the expert's role in our lives and the ability to assign authority to speakers based on the testimonial setting and the credibility of what is asserted (Adler, 2012).

These findings show more support for the reductionist view of testimony and demonstrate our reluctance to believe everything that is said. Realistically speaking, these results show that hearers use information from the environment to judge the credibility of a report before believing the claims. More specifically, based on the little amount of information that people are given, they have the ability to distinguish what they believe to be a credible and a non-credible source of information. In this study, it was shown that participants favour reports that can be verified as either true or false. This is especially true when the report comes from an expert.

In line with the ideas of Fricker (2007), this study suggests a possible effect of a speaker's gender on how believable hearers deem the information he/she reports to be. In Study 1, subjective opinion sentences where the subjects were males were rated significantly higher

than sentences where the subjects were females. While prejudicial and unreasonable (there is no reason the word of a male should be seen as more trustworthy than a female), this result also supports the reductionist view of testimony by demonstrating the effect of source of information on how people interpret the credibility of that information. This result, along with the findings on expert reports, was echoed in the comment section of the study where participants commented on the subject/speaker of the sentence playing a role in how they evaluated the sentence pairings. However, this study only compared these differences for subjective opinions because the other types of testimony did not include both male and female subjects. We will further investigate this finding in Study 2.

In contrast to my hypothesis, I did not find that participants with higher levels of university trusted the testimony of others any less than participants with lower levels. The fact that I didn't find a significant difference between participants with fewer years of university education compared to those with more raises question about the generalization that higher levels of education make individuals more inclined to critical thinking (McMillan, 1987). In addition, this study did not find a significant effect of participant major on how they rated the given reports. This was surprising to me because I have generally thought of the arts and social sciences as having an approach that focuses on thinking critically about issues; while the maths and sciences are a memorize-and-do type of approach to concepts, with little room to be critical in the undergraduate study years. This result is perhaps influenced by the institution in which I am testing participants (Karplus, 1977; McMillan, 1987; McLeod, 2015).

Looking forward, the results of this study highlight several potential issues and opportunities for future study. In a follow up study, I will be focusing on sources of reports by expanding on the ideas of expert testimony and gender. The previous study examined sentence pairs where experts reported information within their field of expertise. Study 2 will examine

whether experts are trusted more than non-experts generally, or only when discussing matters within their field-of-expertise and whether the gender of the speaker has an effect on the trustworthiness of his/her report.

2.3 Study 2

The purpose of this survey was to further explore people's attitudes towards expert testimony and the gender of the speaker. More specifically, Study 2 aims to shed light on differences between peoples' trust in (1) different types of expert testimony (i.e. non-related expert reports versus expert reports) and, (2) differences in people's trust of male versus female speakers.

For this study, I hypothesized that (1) expert reports will be rated similarly, regardless of whether they are experts on the topic or not. This is based on the work of Sperber (2010) and Origgi (2008), as well as my personal tendency to trust the word of experts more than non-experts. In addition, I hypothesized that (2) expert reports will be rated higher than non-expert reports (objective facts), regardless of whether they are experts on the topic or not. This is based on findings from Study 1 and the work of Sperber (2010) and Origgi (2008). Finally, based on the work of Fricker (2007) and findings from Study 1, I hypothesized that (3) male speakers will be trusted more than female speakers.

Method

The methods of Study 2 are very similar to Study 1 (participants and stimuli are different); however, a complete methods section is included here to make this section self-contained.

Participants

Participants (N = 333) were recruited from the Carleton University Cognitive Science SONA system. Before analysis, fifty-nine participants were excluded because they were not native speakers of English (N = 47) or because they completed less than half of the survey (N = 12).[10] The remaining participants (N = 274) included male (N = 68), female (N = 202) and other (N = 4) students, 16 years of age or older. All participants are Carleton University students who are currently enrolled in the first year cognitive science course CGSC 1001 - Mysteries of the Mind and have access to the Carleton University Cognitive Science SONA system. As previously noted, this course is not restricted to first year students. Participants were compensated a 0.5 grade percent in CGSC 1001 for their participation.[11]

Materials

Participants were asked to complete one of two versions of an online survey hosted on Qualtrics (Qualtrics, Provo, UT). The surveys consisted of forty-five scale-based questions (twelve target stimuli and thirty-three filler stimuli), took approximately fifteen minutes to complete and were completely anonymous. Two versions of the survey were created to explore differences in report ratings associated with the gender of the speaker. Differences were noted in Study 1 but further exploration was required. The two surveys were identical in all respects except the names in the four objective report stimuli were switched (e.g. Travis was switched to Sarah and Sarah was switched to Travis). Participants randomly chose one of the surveys to complete. Stimuli were randomized to prevent possible ordering effects. The data were collected between January 2018 and April 2018.

[10] Because this is a language based study, non-native speakers of English were excluded from analysis to eliminate potential language-related issues.

[11] Participants received 0.5% extra credit in Study 2, compared to 0.25% in the Pilot and Study 1 because the Carleton University Cognitive Science SONA system changed its compensation amounts.

At the beginning of the survey, participants answered a background questionnaire about their gender, education level, and current year of university study, university major, native language and age. Next, similar to the Pilot Study and Study 1, participants were presented with pairs of sentences and asked to judge how likely it was that the second sentence was true, given that the first sentence was true. The likelihood of the second sentence was reported along a seven-point scale, with 1 being neither likely nor unlikely to be true, and 7 being extremely likely. All target sentences were of the form "*Someone told someone*…" (e.g., Anna told Peter…) or "*Someone said to someone*..." (e.g., Anna said to Peter…).

Twelve target stimuli were divided into three categories: objective fact, expert report and non-related expert report. Each category comprised four of the twelve target stimuli questions.

1. The *objective fact* examples involved statements that can be verified as either true or false. For example: Assume that sentence (A) is true. Please rate the likelihood that sentence (B) is true on a scale from 1 (neither likely nor unlikely) to 7 (extremely likely).

    (A) Megan told Tom that the office was closed.

    (B) The office was closed.

2. The *expert report* examples involved objective fact statements made by a person with expertise on the statement's topic. For example:

    (A) The scientist said that the river is polluted.

    (B) The river is polluted.

3. The *non-related expert report* examples involved objective fact statements that were made by an individual with expert status but no expertise on the statement's topic. For example:

    (A) The judge said that there is a meteor shower tonight.

    (B) There is a meteor shower tonight.

Study 2 contained thirty-three filler stimuli which had a similar structure to both the Pilot Study and Study 1; these questions were focused around the five human senses (sight, smell, hearing, taste, and touch) and included sentence pairs where the responses would be clearly 1 or 7 to allow us to check whether participants were reading each question and not just clicking through the survey. It is important to note that I did not use the same filler stimuli for Study 2 as I did for the Pilot Study and Study 1. I created a new list of filler stimuli that included authority figures in order to try to prevent participants from guessing the goal of the study. For example:

(A) The policeman heard three dogs barking.

(B) Three dogs were barking.

For the full list of Study 2 stimuli (target and filler pairs), see Appendix F.

Finally, before exiting the survey, two optional sections were provided to the participants. The first section asked participants to describe any factors they felt influenced their ratings of the sentence pairs. The second was a comment section where participants could express any questions, concerns or thoughts they may have had while completing the survey.

As in Study 1, participants were categorized by their reported program of study, using the Statistics Canada (2016) MFS classification structure. Based on the distribution of the participants' reported majors, six categories were excluded for having less than 20 participants (Educational, Recreational and Counselling Services (N = 0), Fine and Applied Arts (N = 1), Engineering and Applied Science (N = 5), and Applied Science Technologies and Trades (N = 0), Health Professions and Related Technologies (N = 0), Undeclared (N = 15)).[12] As in Study 1, I included a new level to the classification structure called Cognitive Science because it is a newer discipline and falls within several MFS categories.

[12] Recall that Study 1 excluded majors with less than 40 participants. This number is different from Study 1 because I had fewer participants in Study 2.

This research project was cleared by the Carleton University Research Ethics Board-B (clearance # 106091) before the study began (January 2018).

Procedure

Participants were provided with an online link to the survey through the Carleton University Cognitive Science SONA System. When the link was accessed participants were directed to the survey hosted on Qualtrics (Qualtrics, Provo, UT). Participants completed the survey on their own time using any device with internet access (computer, laptop, smart phone, etc.). Upon completion, participants were automatically directed back to the Carleton University Cognitive Science SONA system and awarded a 0.5 percent grade bonus for the course CGSC 1001.

Results

The first part of my analysis will be on the combined data of both versions of the survey. This should not affect the data as the questions were exactly the same between versions except the names were switched for four stimuli. This switch of names should negate any differences potentially associated with gender differences.

I conducted descriptive statistics on the data and found that overall expert reports (M =6.05, SD = 1.17, SE = 0.04) were rated higher than objective facts (M = 5.38, SD = 1.47, SE = 0.04). This result was found in my Pilot Study and Study 1 and, supports the idea that expert reports are considered more trustworthy than non-expert reports. Contrary to my hypothesis, descriptive statistics showed that non-related expert reports (M = 5.42, SD = 1.63, SE = 0.05) were rated lower than expert reports, and similar to objective facts. This was not a result that I hypothesized and will be further explored later.

As with the Pilot Study and Study 1, the results of Study 2 were negatively skewed and the analysis will be based on percentiles; objective fact had a skewness of -0.97, expert reports had a skewness of -1.58 and non-related expert reports had a skewness of -0.97 (see **Figure 6**).[13] Four stimuli measured the likelihood of each testimony type, a total score (with a possible range of 0 to 28) was calculated by summing the four scores; higher scores indicating a higher likelihood the statements were true. The distribution of the scores was then divided into quartiles. Quartiles were generated by taking the maximum score (28) and dividing it by 0.25, 0.5 and 0.75. Resulting quartiles were as follows: Q1: 0 to 7, Q2: 8 to 14, Q3: 15 to 21, and Q4: 22 to 28. The quartiles can be more easily understood as: Q1 – low likelihood, Q2 moderate, Q3 – high, and Q4 – very high. This is representative of the 7-point scales used throughout the study. To maximize the number of participants for whom scores were calculated, one response left empty per testimony type was accepted. A score was calculated based on the items with responses and then adjusted by using the mean substitution technique to compensate for the item without a response. The mean substitution method was used for six participants.

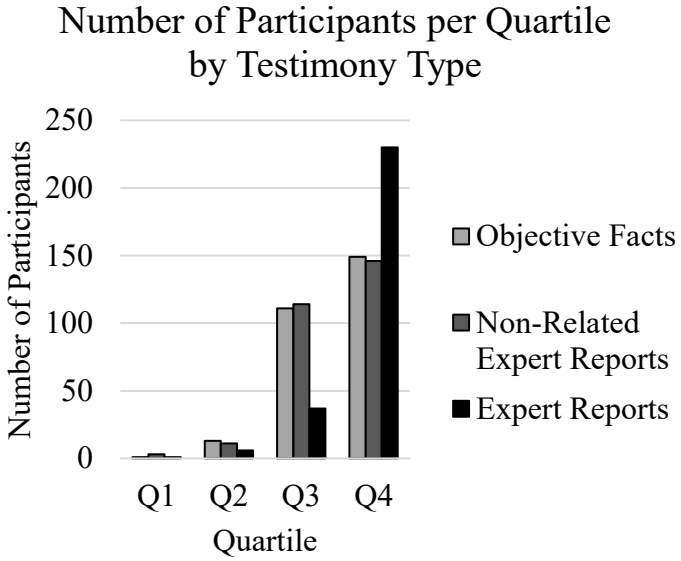

Figure 6. Number of participants in each quartile by testimony type.

Approximately 84% of participants scored in Q4 for expert reports, while only 53% of participants scored in Q4 for non-related expert reports and 54% for objective facts. I conducted

[13] The choice to run my statistics this way was made in consultation with Margot Shields. Mrs. Shields is an expert statistician with 30+ years of experience and 60+ publications with Statistics Canada. For examples of similar methods, see Shields & Wilkins (2009).

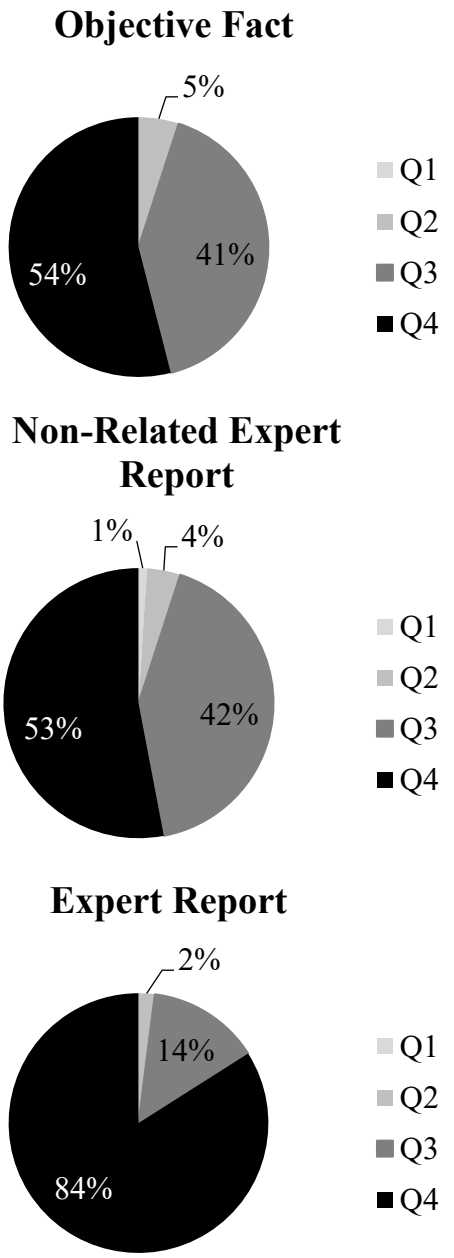

Figure 7. Percentage of participants who fell within each quartile for reports of objective fact, non-expert reports and expert reports.

t-tests between the different types of reports and found a significant difference between expert reports and objective facts ($t(273) = -9.38$, $p < 2.2^{-16}$). This result supports my hypothesis and is the same result that I found in the Pilot Study and Study 1; participants judge expert reports to be more trustworthy than the reports of others. In addition, and contrary to my hypothesis, expert reports were rated significantly higher than non-related expert reports ($t(273) = -7.4$, $p < 1.7^{-12}$), and no significant difference was found between the ratings of objective fact and non-related expert reports ($t(273) = -0.36$, $p = 0.72$). This finding does not support the work of Sperber (2010) or Origgi (2008) and suggests that participants are not always more likely to trust the reports of experts more than the reports of non-experts.

Approximately 69% of first year students, 72% of second year students, 74% of third year students, 79% of fourth year students and 88% of other students scored within Q4. Preliminary results from Pearson's product-moment correlation tests did not show a relationship between university year and testimony rating ($r = 0.07$, $n = 272$, $p = 0.24$). No significant correlations were found between university year and acceptance of

testimony for any type of report: objective fact (r = 0.07, n = 272, p = 0.26), non-related expert reports (r = 0.07, n = 272, p = 0.24) and expert reports (r = 0.00, n = 272, p = 0.87). I conducted ANOVAs on each testimony type to see if participants rated the reports differently based on their university level; First year (N = 154), Second year (N = 36), Third year (N = 27), Fourth year (N = 28). Just like in Study 1, I removed 'other' students (N = 9) from this part of the analysis because 'other' can refer to many different types of university education (e.g. fifth year students, special students, etc.) and I wanted to try to reduce any effects created by this unknown. When examining the results, no significant differences were found for any type of report; objective facts ($F(3, 261) = 1.84$, $p = 0.14$, $n_p^2 = 0.02$), expert reports ($F(3, 261) = 0.9$, $p = 0.44$, $n_p^2 = 0.01$) and non-related expert reports ($F(3, 261) = 0.89$, $p = 0.45$, $n_p^2 = 0.01$).[14] This reinforces our finding from Study 1 that university levels do not affect one's trust in the testimony of others. For an additional break down of the percentage of participants for each quartile by university level is provided in **Table 4**.

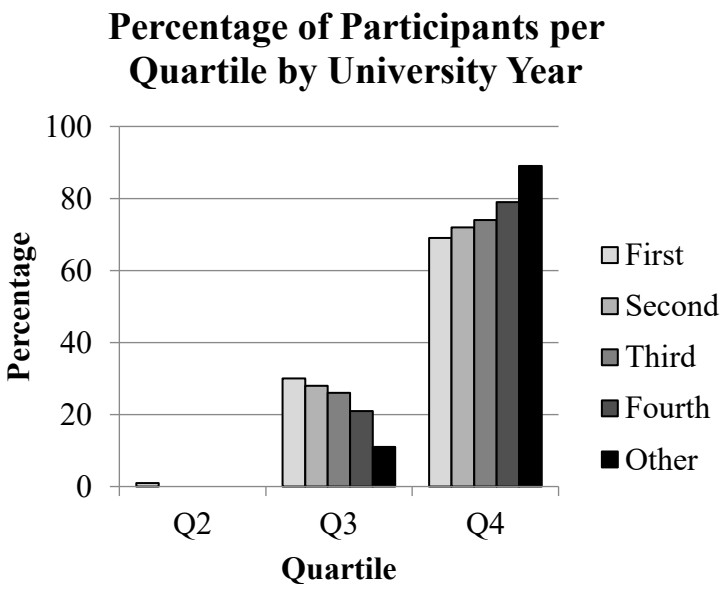

Figure 8. Percentage of participants by quartile and reported university year, for overall testimony ratings.

[14] I ran the same statistics with 'other' students included in the dataset and the results were only slightly different. No significant results were found for any type of report: objective fact ($F(4, 269) = 1.59$, $p = 0.18$, $n_p^2 = 0.02$), expert report ($F(4, 269) = 0.69$, $p = 0.6$, $n_p^2 = 0.01$), non sense expert report ($F(4, 269) = 1.75$, $p = 0.15$, $n_p^2 = 0.02$).

Table 4. Percentage of participants by quartile and reported university year for testimony type.

	Objective Fact				Expert Report				Non-related expert report			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
University Year												
First	1	7	41	51	1	3	16	81	1	3	48	48
Second		3	36	61			11	89	3	8	44	45
Third			56	44		4	7	89		7	26	67
Fourth		4	25	71		4	7	89		7	43	50
Other			38	62			25	75			12	88

Using the MFS, I divided participants by their reported university major; Humanities and Related Fields (N = 20), Social Sciences (N = 122), Commerce, Management and Business (N = 38), Agricultural, Biological, Nutritional and Food Sciences (N = 35), Mathematics, Computer and Physical Sciences (N = 21) and Cognitive Science (N = 17). I conducted ANOVAs on each testimony type to see if participants rated the reports differently based on their reported university major. Similar to Study 1, no significant differences were found for any type of report; objective fact ($F(5, 239) = 0.83$, $p = 0.53$, $n_p^2 = 0.02$), expert report ($F(5, 239) = 0.29$, $p = 0.92$, $n_p^2 = 0.01$) and non-related expert reports ($F(5, 239) = 1.04$, $p = 0.4$, $n_p^2 = 0.02$). This result confirms that regardless of their discipline, participants trust the reports of others similar to one another.

Next, I ran ANOVAs to test whether participants with the same major rated reports differently depending on their university level. When examining reports of objective fact, significant differences were found for participants within the Social Sciences ($F(3, 114) = 2.96$, $p = 0.04$, $n_p^2 = 0.07$), Commerce, Management and Business ($F(3, 34) = 3.47$, $p = 0.03$, $n_p^2 = 0.23$) and Mathematics, Computer and Physical Sciences ($F(3, 14) = 3.8$, $p = 0.03$, $n_p^2 = 0.45$) classifications. More specifically, within the Social Sciences, third year students rated reports lower than fourth year students ($p = 0.05$); within Commerce, Management and Business, second year students rated

reports lower than first year ($p = 0.02$) and fourth year ($p = 0.03$) students; and within Mathematics, Computer and Physical Sciences, third year students rated the reports higher than fourth year students ($p = 0.03$). No other significant results were found for the objective fact reports; Humanities and Related Fields ($F(3, 16) = 1.24$, $p = 0.33$, $n_p^2 = 0.19$), Agricultural, Biological, Nutritional and Food Sciences ($F(3, 30) = 2.16$, $p = 0.11$, $n_p^2 = 0.18$) and Cognitive Science ($F(2, 14) = 0.36$, $p = 0.7$, $n_p^2 = 0.05$).

For expert reports, ANOVAs found significant differences in ratings for participants who fell within the Mathematics, Computer and Physical Sciences classification ($F(3, 14) = 10.67$, $p < 0.00$, $n_p^2 = 0.7$). More specifically, fourth year students rated reports significantly lower than first year ($p < 0.00$), second year ($p < 0.00$) and third year students ($p < 0.00$). No other significant differences were found; Humanities and Related Fields ($F(3, 16) = 0.34$, $p = 0.8$, $n_p^2 = 0.06$), Social Sciences ($F(3, 114) = 0.83$, $p = 0.47$, $n_p^2 = 0.02$), Commerce, Management and Business ($F(3, 34) = 0.87$, $p = 0.47$, $n_p^2 = 0.07$), Agricultural, Biological, Nutritional and Food Sciences ($F(3, 30) = 0.86$, $p = 0.47$, $n_p^2 = 0.08$), Cognitive Science ($F(2, 14) = 0.1$, $p = 0.91$, $n_p^2 = 0.01$). Study 2 did not find any significant differences for the non-related expert reports; Humanities and Related Fields ($F(3, 16) = 0.95$, $p = 0.44$, $n_p^2 = 0.15$), Social Sciences ($F(3, 114) = 0.61$, $p = 0.61$, $n_p^2 = 0.02$), Commerce, Management and Business ($F(3, 34) = 1.2$, $p = 0.33$, $n_p^2 = 0.1$), Agricultural, Biological, Nutritional and Food Sciences ($F(3, 30) = 0.25$, $p = 0.86$, $n_p^2 = 0.02$), Mathematics, Computer and Physical Sciences ($F(3, 14) = 0.55$, $p = 0.65$, $n_p^2 = 0.11$), Cognitive Science ($F(2, 14) = 0.53$, $p = 0.56$, $n_p^2 = 0.07$). [15] These results suggest that there is some kind of interaction between major and university level, however, when looking at the dataset, I feel it can be explained as differences in the number

[15] I also conducted ANOVAs on the dataset as a whole to test whether there were interactions between any major and any university year. No significant differences were found for expert reports ($F(14, 222) = 1.61$, $p = 0.08$, $n_p^2 = 0.09$) or non-related expert reports ($F(14, 222) = 0.57$, $p = 0.89$, $n_p^2 = 0.03$). A significant difference was found for objective fact ($F(14, 222) = 2.1$, $p = 0.01$, $n_p^2 = 0.12$).

Table 5. Mean ratings for each testimony type by participant major and year of study.

	Objective Fact	**Expert Report**	**Non-Related Expert Report**
Humanities			
First Year	M = 5.23	M = 5.88	M = 5.42
Second Year	M = 5.5	M = 6.33	M = 4.92
Third Year	M = 6.58	M = 6.75	M = 5.75
Fourth Year	M = 6.25	M = 7	M = 5
Social Sciences			
First Year	M = 5.25	M = 6	M = 5.44
Second Year	M = 5.7	M = 6.36	M = 5.22
Third Year	M = 4.72	M = 6	M = 5.56
Fourth Year	M = 5.6	M = 6.21	M = 5.23
Commerce/Business			
First Year	M = 5.58	M = 6.28	M = 5.25
Second Year	M = 3	M = 4.75	M = 5.25
Third Year	M = 5.5	M = 6	M = 6.07
Fourth Year	M = 5.8	M = 6.18	M = 5.45
Biological Sciences			
First Year	M = 5.04	M = 5.68	M = 5.14
Second Year	M = 5.75	M = 5.93	M = 5.46
Third Year	M = 5.31	M = 5.81	M = 4.75
Fourth Year	M = 5.81	M = 6.5	M = 5.44
Math/Computer & Physical Science			
First Year	M = 4.72	M = 5.89	M = 5.75
Second Year	M = 5.31	M = 6.06	M = 5.5
Third Year	M = 6.44	M = 6.63	M = 5.63
Fourth Year	M = 3.25	M = 2.5	M = 5
Cognitive Science			
First Year	M = 4.96	M = 6.09	M = 5.39
Second Year	M = 5.25	M = 6.75	M = 4.25
Third Year	M = 5	M = 6.5	M = 6.25
Fourth Year	N/A	N/A	N/A

of participants for each university year within the major classification (for example, of the eighteen participants with math majors only one was a fourth year student). See **Table 5** for a more detailed comparison of rating averages for each type of testimony by participant major and university year.

In Study 2, I did not find a significant effect of participant gender on report rating for any type of testimony; objective fact ($F(269) = 1.07$, $p = 0.37$), expert report ($F(269) = 1.75$, $p = 0.14$), and non-sense export report ($F(269) = 0.32$, $p = 0.87$). This supports my finding in Study 1 that male and female participants rated all types of testimony similarly. Keeping with the idea of gender, there appears to be differences in ratings with respect to the gender of the speaker. The next part of my analysis compares the four objective fact questions of version 1 of the survey with the four objective fact questions of version 2. See **Table 6** for a list of Study 2 version 1 and version 2 objective fact stimuli.

Table 6. A comparison of the objective fact stimuli for version 1 and version 2 of Study 2.

Version 1 Stimuli	**Version 2 Stimuli**
(A) Travis said that the truck didn't stop at the stop sign. (B) The truck didn't stop at the stop sign.	(A) Sarah said that the truck didn't stop at the stop sign. (B) The truck didn't stop at the stop sign.
(A) Jeremy said that the sunflowers grew much taller this year. (B) The sunflowers grew much taller this year.	(A) Jennifer said that the sunflowers grew much taller this year. (B) The sunflowers grew much taller this year.
(A) Sarah said that she ate two hot dogs, three hamburgers and caramel popcorn at the fair. (B) Sarah ate two hot dogs, three hamburgers and caramel popcorn at the fair.	(A) Jeremy said that he ate two hot dogs, three hamburgers and caramel popcorn at the fair. (B) Jeremy ate two hot dogs, three hamburgers and caramel popcorn at the fair.
(A) Jennifer said that the car passed the safety test. (B) The car passed the safety test.	(A) Travis said that the car passed the safety test. (B) The car passed the safety test.

Preliminary descriptive statistics show a small difference in ratings of stimuli depending on the gender of the speaker; male reports (M = 5.39, SD = 1.43, SE = 0.06), female reports (M = 5.36, SD = 1.51, SE = 0.06). See **Table 7** for a more in-depth look at the descriptive statistics for the comparison of version 1 and version 2 objective fact stimuli. The ratings for stimuli with male speakers fell within in Q4 approximately 58% of the time, while ratings for female speakers fell within Q4 approximately 53% of the time. T-tests showed that this was not a significant difference, ($t(119) = -0.71$, $p = 0.48$). I conducted further t-tests to compare each sentence pair with its opposite gender sentence pair. There was no significant difference in ratings of male and female speakers for the "truck" stimuli, ($t(119) = -1.72$, $p = 0.09$) or the "food" stimuli, ($t(119) = 1.13$, $p = 0.26$). However, significant differences in ratings of male and female speakers were found for the "flower" stimuli, ($t(119) = -2.15$, $p = 0.03$) and the "car" stimuli, ($t(119) = 1.97$, $p = 0.05$). For the "car" stimuli, female speakers were rated significantly higher than male speakers. In contrast, for the "flower" stimuli, male speakers were rated significantly higher than female speakers. This result does not support the work of Fricker (2007) or the results I hypothesized. Recall that I was expecting male speakers to be rated significantly higher than female speakers, especially for topics such as trucks and cars. This result does not provide a better understanding of the effect the speaker's gender has on how likely participants judge the reports to be.

Table 7. Comparison of the mean, standard deviation and standard error of objective fact reports for male and female speakers.

	Truck	**Car**	**Food**	**Flower**
Male	M = 5.43 SD = 1.38 SE = 0.11	M = 5.42 SD = 1.48 SE = 0.14	M = 5.41 SD = 1.57 SE = 0.14	M = 5.3 SD = 1.33 SE = 0.11
Female	M = 5.08 SD = 1.57 SE = 0.14	M = 5.75 SD = 1.28 SE = 0.1	M = 5.57 SD = 1.48 SE = 0.12	M = 4.9 SD = 1.57 SE = 0.14

Discussion

Similar to the Pilot and Study 1, the overall results of Study 2 support both the anti-reductionist and reductionist views of testimony. Approximately 99% of participants scored within Q3 and Q4 for all types of testimony in Study 2, demonstrating not only our disposition to trust others, but our ability to assess sources in making decisions about whether to trust or not trust the information they are reporting. However, as will be demonstrated, a more detailed analysis of the results favour the reductionist point of view.

Study 2 findings align with the Pilot Study and Study 1 and found that participants are significantly more likely to trust expert reports than objective facts. This result supports the work of Origgi (2008) and Sperber (2010). However, I did not find a significant difference between non-related expert reports and objective facts. These findings align with the reductionist view of testimony because they show that hearers use information from internal and external experiences (Matheson, 2005; Sperber 2010) to judge the credibility of a report before believing the claims. It's possible that while participants originally take into account the authority of the expert (external factors), their own experiences with a person of the same expertise or their knowledge of the subject matter (internal factors) play a larger role in the decision to trust the report or not.

In contrast to what the results of Study 1 first indicated, this study did not show an overall effect of a speaker's gender on how believable hearers deem their report to be. Female speakers were rated higher than males for questions involving cars and food, while male speakers were rated higher for questions involving flowers and a truck. I wonder if participants assign more authority to males and females who report information on topics that are not stereotypically associated with their gender because they feel that if this person is talking about it then they MUST have some expertise on the topic. For example, if a female is discussing trucks she must

be an expert on them and likewise a male must be an expert on flowers if he is discussing them. This result also supports the reductionist view of testimony by demonstrating the effect of source of information on how people interpret the credibility of that information. However, further analysis should be done to try to get a clear understanding of whether the gender of the speaker has an effect on how trustworthy individuals perceive their reports to be.

In Study 2, I did not find a significant effect of university level on participants' trust in the provided reports; participants with higher levels of university study did not trust the testimony of others any more or less than participants with lower levels of university. In addition, this study did not find an effect of the participants' university major on the trust of the given report. Again, this result and the results of Study 1 may align with Piaget's Theory of Cognitive Development and suggest that participants reach their peak development before they reach university. These results could also suggest that higher levels of education do not emphasize critical thinking as much as they are believed to (McMillan, 1987). However, it is also quite possible that this limited study does not assess critical thinking at all and further examination is required.

3 Discussion

In support of the anti-reductionist view of testimony, all three of my studies (Pilot Study, Study 1 and Study 2) found that regardless of the source and/or type of a report, participants are more likely to judge reports as likely (i.e. the majority of participants scored within Q3 and Q4 for all types of testimony) than to not have an opinion. This echoes Hume's idea that God equipped humans with a disposition to believe in the honesty of others. It also supports Grice's idea that as hearers, we automatically assume speakers are following the Gricean Maxims and have no reason to lie (Grice, 1975; Levinson, 1983; Grice, 1989). However, this result, and further analysis of my data, also shows support for the reductionist view.

The results of my three studies showed that participants trust testimony differently depending on the type of information that is being reported. Participants were more likely to trust objective facts than subjective beliefs, demonstrating that they prefer information that can be verified as objectively true or false. Similar to the work of Sperber (2010) and Origgi (2008), my results also showed that participants were more likely to rate the testimony of expert figures higher than those of subjective opinion and objective facts. These results show that there are external, as well as internal factors that play a role in trusting and believing the testimony of others. Participants rated reports differently depending on their content (internal) and their source (external). Not only do these results demonstrate the importance of expert figures in our lives, it also demonstrates the ability of hearers to assign authority to 'speakers' based on the information they receive from their environment and the source (Adler, 2012).

Study 2 results showed that when experts reported information from outside of their field of expertise, participants rated them similarly to the objective fact reports of random male and female speakers and significantly lower than reports from experts within their field of expertise. When an expert in health (e.g. doctor) provides testimony on a sporting event, participants are most likely accessing previous knowledge and/or experiences from a similar event to make an informed decision on the report of this expert. This finding also supports the reductionist view that hearers require supporting information (e.g. perception and memory) before trusting a report because it demonstrates that participants are accessing information outside the given context of the report to judge its likelihood.

Interestingly, when examining the effect the source of the report has on its trustworthiness, I did not find any clear effects of gender. In Study 1 and Study 2, male and female participants rated the reports similarly. There was nothing to suggest that females are

more trusting of people than males or vice versa. Upon further analysis, Study 1 demonstrated, at least tentatively, that female speakers were trusted less than male speakers. A more detailed analysis of the effects of the speaker's gender was conducted in Study 2 but turned up inconclusive results. Overall, reports given by male speakers were rated higher than reports given by female speakers but the difference was not significant. More specifically, compared to female speakers, the male speakers were rated higher for two stimuli and only one was significantly higher. Similarly, female speakers were rated higher for two stimuli, with only one being significant. These results differ from the work of Fricker (2007) which makes me question whether the methods I used to examine this research question were designed appropriately. I do believe however that my results do warrant further investigation in supporting the reductionist view, pointing to the finding that depending on the type of report, male and female speakers are rated differently; males are rated higher on topics that are considered for all intents and purposes stereotypically female, while females are rated higher on topics that are stereotypically male. A study more focused on gender effects should be conducted to get a clearer understanding of this interaction.

In contrast to my original hypothesis, the findings of my three studies do not show an effect of education or university level on the acceptance of testimony. In other words, as individuals attain higher levels of education; my results did not indicate that they trust the testimony of others significantly more or less than participants with lower levels of education. This result could have three implications.

Firstly, the fact that I didn't find a significant difference between participants with fewer years of university study and those with more raises questions about the generalization that higher education makes individuals more inclined to critical thinking (McMillan, 1987). Jean

Piaget believed that learning is accomplished through active discovery, and therefore classroom learning should be student centered (McLeod, 2015). By using active methods requiring the student to rediscover and reconstruct truths, as well as collaborative and individual activities, teachers could enhance an individual's development (McLeod, 2015). It would be interesting to examine this idea across educational institutions and systems to see if some are more focused on student learning and development, i.e., individuals versus population as a whole and the associated costs and benefits of each with regards to teaching and student needs.

A different approach to studying the effect of education on our ability to think critically and accept the testimony of others would be to limit the study population to participants over the age of twenty-five who have completed their education. Using this technique, I believe it would also be interesting to study whether different environments favour different types of thinking. For example, do participants judge things the same when they are at work as they do when they are at home? I wonder whether it's possible that we wear separate hats in our social and professional lives when it comes to reasoning. Do individuals become critical thinkers in their workplace because that is what is required of them, not because they are critical thinkers in general? It would be interesting to analyze professionals at work and on their leisure time.

Secondly, Jean Piaget's Theory of Cognitive Development contends that humans progress through four stages of intellectual development (sensorimotor, preoperational, concrete operational and formal operational) between birth and adolescence resulting in the construction of a mental model of the world (McLeod, 2015). Throughout their development and as a result of biological maturation and environmental influences, children develop the ability to reason and think about things by way of progressive reorganization of mental processes. Piaget further postulates that the final stage of development, the formal operational stage, lasts into adulthood

(McLeod, 2015). In my research, it could be the case that by the time individuals reach this stage, it takes a much longer period of time to see notable changes in their reasoning (that is the four years of university is not long enough to study changes). Like the Pilot study, a future study could take a more in-depth look at differences between high school and university students' ratings of testimony.

Finally, it's possible that this is a subjective experience which requires a more personal approach to its study. A longitudinal study may be more appropriate and could be used to examine individual development throughout the course of one's education. For example, participants complete the same survey before entering high school (grade 8), before entering university, once they have finished their university program and 4 years after completing the program.

Looking forward, the results of this study highlight several potential issues. Firstly, in examining the study's methodology, it is not possible to evaluate or eliminate an individual's life experience from their decision-making process. While using random names in studies is a standard technique that has attained all kinds of significant results in the past, it would be interesting to study whether in this methodological approach the selected names had an effect on the rating of stimuli. It is highly unlikely for this study that all eight hundred plus participants had an issue with someone named John and therefore rated anything he said as lower than everyone else's, but there could be underlying personal reasons for people's ratings that are not captured by this study. In order to address this issue, a more qualitative approach could be used in the data collection process for the study, i.e., having the participant complete the survey in person, accompanied by a survey interviewer asking the participants why they chose the rating

they did. This new element would allow for some analysis of the thought process behind the decisions made by participants.

An alternative approach to studying this issue could be to run identical surveys, having the participants of one study read a paragraph describing the survey's speakers, say John and Jane, and using random names for the other participants. By providing participants with descriptions of the 'speakers' they are provided with enough information to make a decision about the character's trustworthiness, negating if not removing the need to access previous experiences regarding someone with the same name. This could possibly diminish any bias in rating caused by personal experience.

For future studies, I would also look into providing participants with a longer stimuli list in which each sentence pair topic was said by both male and female, experts and non-experts. For example:

(A1) Madison said that the ice cream was delicious.

(A2) Jared said that the ice cream was delicious.

(A3) The chef said that the ice cream was delicious.

(A4) The doctor said that the ice cream was delicious.

(B) The ice cream was delicious.

While this may be a tedious task for the participants, it would allow for more efficient testing and analysis on the part of the researcher. It would also allow for more filler stimuli to be added which would help reduce the chance that participants guess the goal of the study. I would also consider having participants watch videos of random people saying sentence A instead of reading it. This would simulate more of a conversation scenario which is how people acquire

testimonial reports in the real world. However, this may result in additional confounding variables such as the speaker's race, age and/or their accent.

Finally, I question whether the nature of my stimuli affected the participants ratings and my results overall. By this I mean that because all the stimuli had random topics with no direct implications or consequences in the lives of my participants, or the speakers, did the ratings not matter to them. It would be interesting to examine these factors on formal testimony. Perhaps differences in ratings are more apparent when there are obvious consequences of a hearer's belief (or disbelief) in the report as in the research examples cited earlier. For example, when they are a jury member and someone's life depends on how they interpret different testimonies in a trial setting versus being told about all the different flavours of cake from a stranger.

4 Conclusions

The anti-reductionist view was supported within the overall findings of this research; i.e. humans have a tendency to trust the word of others, regardless of source and/or information type. However, the more specific results supported the reductionist view suggesting that hearers use multiple factors to assess the credibility of speakers during conversation (e.g. type of information being reported, gender of the speaker, etc.).

Given the findings, I believe testimony is a crucial source of belief and knowledge that is more complex than the proponents of anti-reductionism would argue. These studies provide support for the reductionist argument and a strong starting point for empirical research into natural testimony, cognitive processes associated with decision making (i.e. perception, memory and induction) and epistemology in general.

Appendices

Appendix A

Recruitment Notice

Study Name: Truth and Reliability

Description: We invite you to participate in a study entitled "Truth and Reliability". This study explores how we come to form beliefs and gain knowledge. Participants will be required to complete an online questionnaire.

The questionnaire will be anonymous and you will have the right to end your participation in the study at any time. Any incomplete questionnaires will not be analyzed and participants who withdraw from the study will have the data collected up to the point of the withdrawal destroyed.

Eligibility Requirements: Any speaker of English is eligible.

Risks: There is no risk associated with this study.

Duration and Locale: The entire session will take between 15 and 30 minutes and can be completed in a location and at a time of the participants convenience. Please follow the following link if you wish to participate: https://cuhealth.eu.qualtrics.com/jfe/form/SV_1MkolZaBvizyD7n

Researchers:

Emma McLarens (Principal Investigator); emma.mclarens@carleton.ca

Ida Toivonen (Faculty Sponsor); ida.toivonen@carleton.ca

This research has been cleared by Carleton University Research Ethics Board-B (clearance # 106091).

Should you have any ethical concerns with the study, please contact Dr. Andy Adler (Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or email: ethics@carleton.ca).

Appendix B

Participant Consent

Title: Truth and Reliability
Date of Ethics Clearance: December 16, 2016
Ethics Clearance for the Collection of Data Expires: November 31st, 2018
Ethics Clearance #106091

This form is to indicate that you are choosing to participate in a study on truth and reliability.

Purpose: The purpose of this study is to evaluate factors that influence how we form beliefs and gain knowledge.

Task Requirements: Your participation will consist of completing a short online questionnaire, approximately 15 - 30 minutes.

Compensation: Participants from the Carleton University SONA system will receive a 0.25% course credit for participating. All other participants will not be compensated.

Potential Risks: There are no risks associated with this study.

Right to Withdraw: Your participation in this study is entirely voluntary. At any point during the study, you have the right to withdraw without penalty. An option to withdraw from the survey will be presented at the bottom of each page of the survey. If you wish to withdraw at any point, select the withdraw option and the researchers will not use your data. However, it is not possible to withdraw your data after you have completed the survey because your responses are anonymous and the researchers will not be able to identify which responses were yours.

Confidentiality: The data collected in this study are anonymous. We do not ask for your name nor do we retain computer IP addresses. The anonymous data will be kept indefinitely and will be used for research and teaching purposes.

This survey is powered by Qualtrics. Qualtrics servers are located in the United States and are subject to the United States privacy and information laws.

Research Personnel: This project is carried out by Emma McLarens under the supervision of Dr. Ida Toivonen. Emma McLarens is a master's student in Cognitive Science at Carleton University. Ida Toivonen is Associate Professor in Cognitive Science and Linguistics at Carleton University.

REB Contact info for Concerns: This research has been cleared by Carleton University Research Ethics Board-B (clearance # 106091).
Should you have any ethical concerns with the study, please contact Dr. Andy Adler (Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or email: ethics@carleton.ca).

Researcher contact information: Emma McLarens
Institute of Cognitive Science Carleton University
Email: EmmaMcLarens@cmail.carleton.ca

Supervisor contact information: Ida Toivonen
Institute of Cognitive Science Carleton University
Email: Ida.Toivonen@carleton.ca

Selecting "I agree" indicates that you are 16 years of age or older, and indicates your consent to the publication of the results with the understanding that anonymity will be preserved.

Appendix C

Background Questionnaire

1. Are you a native speaker of English?
    Yes
    No

2. How old are you?
    18 – 25
    26 – 40
    41 – 60
    61+

3. Do you identify as male or female?
    Male
    Female
    Other
    Prefer Not to Say

4. What is the **HIGHEST** degree or level of school that you have completed?
    Master's degree or doctorate
    Univeresity bachelor's degree
    Some university (currently enrolled)
    Some college (currently enrolled)
    College
    High school
    Some high school, no diploma

5. If you are currently enrolled in university, what year are you in? *
    First year
    Second year
    Third year
    Fourth year
    Other (fifth year, special student, etc.)

6. If you are currently in University, what is your major? *

*Questions #5 and #6 appeared in Study 1 and 2, NOT in the Pilot Study

Appendix D

Bold sentence pairs indicate subjective opinion target stimuli, *italic* sentence pairs indicate objective fact target stimuli and ***bold-italic*** sentence pairs indicate expert report target stimuli.

Pilot Stimuli List

(A) Karen told Bobby that the concert was bad.
(B) The concert was bad.

(A) Kyle told Sue that his legs were sore.
(B) Kyle's legs were sore.

(A) Scotty said to John that he likes baseball.
(B) Scotty likes baseball.

(A) Kate said to Bob that she was cold.
(B) Kate was cold.

(A) Jesse said to Ray that the film won a Grammy award.
(B) The film won a Grammy award.

(A) Pat said to Erin that the cruise ship sank.
(B) The cruise ship sank.

(A) Sharon told Chris that the bus was cancelled.
(B) The bus was cancelled.

(A) Megan told Tom that the office was closed
(B) The office was closed.

(A) The doctor told Tim that his son has strep throat.
(B) Tim's son has strep throat.

(A) The mechanic said to Crystal that the car needs new brake pads.
(B) The car needs new brake pads.

(A) The dentist told Mac that he has a cavity.
(B) Mac has a cavity.

(A) The chef told Mindy that the soup contains garlic.
(B) The soup contains garlic.

(A) Daisy heard the door open.
(B) The door opened.

(A) It smells like Henry has been baking banana bread.
(B) Henry has been baking banana bread.

(A) It looks like a storm is starting.
(B) A storm is starting.

(A) It tastes like the soup has salt in it.
(B) The soup has salt in it.

(A) Joan likes tractors.
(B) It is raining.

(A) It looks like Greg has been crying.
(B) Greg has been crying.

(A) It sounds like Jessica's roommate is home.
(B) Jessica's roommate is home.

(A) The bread tastes moldy.
(B) The bread is moldy.

(A) It smells like Jennifer has been frying bacon.
(B) Jennifer has been frying bacon.

(A) Beth ordered five pizzas for dinner.
(B) Beth's family likes mozzarella cheese.

(A) Katie saw the punch being spiked.
(B) The punch was spiked.

(A) It smells like it is springtime.
(B) It is springtime.

(A) It looks like our neighbours are moving.
(B) Our neighbours are moving.

(A) It sounds like Emma is distracted.
(B) Emma is distracted.

(A) It sounds like the home team just scored.
(B) The home team just scored.

(A) It smells like the cake has vanilla in it.
(B) The cake has vanilla in it.

(A) The coffee tastes like it is burnt.
(B) The coffee is burnt.

(A) It sounds like there are dogs in Brian's backyard.
(B) There are dogs in Brian's backyard.

(A) Tyler looks like he has been working out.
(B) Tyler has been working out.

(A) Kristin goes to the library every day.
(B) Kristin likes books.

(A) It tastes like the cookies were made from scratch.
(B) The cookies were made from scratch.

(A) It sounds like the band has improved.
(B) The band has improved.

(A) Callie saw the chicken cross the road.
(B) The chicken crossed the road.

(A) The bus is always late.
(B) The bus was late this morning.

(A) Chad often has lunch with Judy.
(B) Chad had lunch with Judy yesterday.

(A) Melissa loves animals.
(B) Melissa owns 4 pigs.

(A) The fish tastes like it's undercooked.
(B) The fish is undercooked.

(A) It is snowing.
(B) It is snowing.

Appendix E

Bold sentence pairs indicate subjective opinion target stimuli, *italic* sentence pairs indicate objective fact target stimuli and ***bold-italic*** sentence pairs indicate expert report target stimuli.

Study 1 Stimuli List

(A) Karen told Bobby that the concert was bad.
(B) The concert was bad.

(A) Kyle told Sue the new restaurant was tasty.
(B) The new restaurant is tasty.

(A) Scotty said to John that the exams were confusing.
(B) The exams were confusing.

(A) Megan told Tom that the office was closed
(B) The office was closed.

(A) Jesse said to Ray that the film won a Grammy award.
(B) The film won a Grammy award.

(A) Pat said to Erin that the cruise ship sank.
(B) The cruise ship sank.

(A) The doctor told Tim that his son has strep throat.
(B) Tim's son has strep throat.

(A) The dentist told Mac that he has a cavity.
(B) Mac has a cavity.

(A) The chef told Mindy that the soup contains garlic.
(B) The soup contains garlic.

(A) It smells like Henry has been baking banana bread.
(B) Henry has been baking banana bread.

(A) Daisy heard the door open.
(B) The door opened.

(A) It looks like Greg has been crying.
(B) Greg has been crying.

(A) It sounds like Jessica's roommate is home.
(B) Jessica's roommate is home.

(A) It looks like a storm is starting.
(B) A storm is starting.

(A) It tastes like the soup has salt in it.
(B) The soup has salt in it.

(A) Joan likes tractors.
(B) It is raining.

(A) It smells like Jennifer has been frying bacon.
(B) Jennifer has been frying bacon.

(A) Beth ordered five pizzas for dinner.
(B) Beth's family likes mozzarella cheese.

(A) Katie saw the punch being spiked.
(B) The punch was spiked.

(A) It smells like it is springtime.
(B) It is springtime.

(A) It looks like our neighbours are moving.
(B) Our neighbours are moving.

(A) It sounds like Emma is distracted.
(B) Emma is distracted.

(A) It sounds like the home team just scored.
(B) The home team just scored.

(A) It smells like the cake has vanilla in it.
(B) The cake has vanilla in it.

(A) The coffee tastes like it is burnt.
(B) The coffee is burnt.

(A) It sounds like there are dogs in Brian's backyard.
B) There are dogs in Brian's backyard

(A) Kristin goes to the library every day.
(B) Kristin likes books.

(A) The bread tastes moldy.
(B) The bread is moldy.
(A) It tastes like the cookies were made from scratch.
(B) The cookies were made from scratch.

(A) It sounds like the band has improved.
(B) The band has improved.

(A) Callie saw the chicken cross the road.
(B) The chicken crossed the road.

(A) The bus is always late.
(B) The bus was late this morning.

(A) Chad often has lunch with Judy.
(B) Chad had lunch with Judy yesterday.

(A) Melissa loves animals.
(B) Melissa owns 4 pigs.

(A) The fish tastes like it's undercooked.
(B) The fish is undercooked.

(A) It is snowing.
(B) It is snowing.

Appendix F

Italic sentence pairs indicate objective fact target stimuli, ***bold-italic*** sentence pairs indicate expert report target stimuli and underlined sentence pairs indicate non-sense expert target stimuli.

Study 2 Stimuli List

(A) The doctor said that the house was for sale.
(B) The house was for sale.

(A) The scientist said that the popsicles melted.
(B) The popsicles melted.

(A) The professor said that the cookies contain raisins.
(B) The cookies contain raisins.

(A) The judge said that there is a meteor shower tonight.
(B) There is a meteor shower tonight.

(A) The doctor said that the infection is contagious.
(B) The infection is contagious.

(A) The security guard said that the lock was broken.
(B) The lock was broken.

(A) The fireman said that the dryer was the cause of the fire.
(B) The dryer is the cause of the fire.

(A) The scientist said that the river is polluted.
(B) The river is polluted.

(A) Travis said that the truck didn't stop at the stop sign.
(B) The truck didn't stop at the stop sign.

(A) Jennifer said that the car passed the safety test.
(B) The car passed the safety test.

(A) Sarah said that she ate two hot dogs, three hamburgers and caramel popcorn at the fair.
(B) Sarah ate two hot dogs, three hamburgers and caramel popcorn at the fair.

(A) Jeremy said that the sunflowers grew much taller this year.
(B) The sunflowers grew much taller this year.

(A) The judge spent a weekend fishing.
(B) The judge caught 8 fish.

(A) It sounds like the teacher is happy.
(B) The teacher is happy.

(A) Beth saw the fireman eat lucky charms.
(B) The fireman ate lucky charms.

(A) I heard the professor was in a car crash.
(B) The professor was in a car crash.

(A) The security guard heard noises coming from the basement.
(B) There is something in the basement.

(A) Alisha saw Chris hold his breath for 47 seconds.
(B) Chris held his breath for 47 seconds.

(A) Charlie saw the movie before it was officially released.
(B) Charlie saw the movie before it was officially released.

(A) Allen tasted salt in the soup.
(B) There was salt in the soup.

(A) Maya felt a spider crawling up her back.
(B) There was a spider crawling up Maya's back.

(A) Jacob believed he smelt burnt toast this afternoon.
(B) Someone burnt toast this afternoon.

(A) The train is always early.
(B) The train was early this morning.

(A) Brad will eat any kind of protein.
(B) Brad eats crickets.

(A) It sounds like Alana has been practicing the piano.
(B) Alana has improved her piano playing.

(A) It looks like Mario got his hair cut.
(B) Mario got his hair cut.

(A) It smells like your dad was eating beans.
(B) Your dad was eating beans.

(A) Eric heard the salesperson say that the ring is made of real diamonds.
(B) The ring is made of real diamonds.

(A) The scientist believes in the existence of aliens.
(B) Aliens exist.

(A) It sounds like Andrew is angry.
(B) Andrew is angry.

(A) It looks like the car was in an accident.
(B) The car was in an accident.

(A) The gymnast is very talented.
(B) The gymnast will make the Olympic team.

(A) The boys and girls were playing video games.
(B) The girls beat the boys at Mario Kart.

(A) Tessa turned 21 two days ago.
(B) Tessa is going to have a big birthday party.

(A) Danny couldn't get the lamp to turn on.
(B) The light bulb was burnt out.

(A) It sounds like the train is coming.
(B) Mark is going to miss the train.

(A) It tasted like the chicken at dinner was fake.
(B) The chef served tofu for dinner.

(A) Josh believed the museum was haunted.
(B) The museum was haunted.

(A) Chris believed he made the right decision.
(B) Chris made the right decision.

(A) Jenna believed she made the right decision.
(B) Jenna made the right decision.

(A) Jessica believes the world is going to end in 2025.
(B) The world is going to end in 2025.

(A) The conservation officer jumped out of his boots.
(B) The conservation officer jumped out of his boots.

(A) The President thought he felt jelly in his boots.
(B) There was jelly in the President's boots.

(A) The doctor looks like he had an allergic reaction.
(B) The doctor is allergic to peanut butter.

(A) The policeman heard three dogs barking.
(B) Three dogs were barking.

References

Adler, J. (2012). Epistemological problems of testimony. *The Stanford Encyclopedia of Philosophy*, Retrieved from <https://plato.stanford.edu/archives/win2017/entries/testimony-episprob/>

Audi, R. (1997). The place of testimony in the fabric of knowledge and justification. *American Philosophical Quarterly, 34*, 405-22.

Ballantyne, N. (2015). Debunking biased thinkers (including ourselves). *Journal of the American Philosophical Association, 1(1),* 141-162.

Benton, M.A. (n.d.). Knowledge norms. *The Internet Encyclopedia of Philosophy*. Retrieved from http://www.iep.utm.edu/kn-norms/

Blackburn, S. (2016). Epistemology. In *The Oxford Dictionary of Philosophy (3rd).* Retrieved from http://www.oxfordreference.com.proxy.library.carleton.ca/view/10.1093/acref/9780198735304.001.0001/acref-9780198735304-e-1113?rskey=JLROhw&result=1141

Boyle, J.A. (2014). Vocal pitch manipulation and credibility in the courtroom: Differences between male and female expert witnesses. *The University of Alabama, ProQuest Information & Learning*

Burge, T. (1993). Content preservation. *The Philosophical Review*, *102,* 457–488.

Crystal, D. (1985). Maxims of conversation. In *A Dictionary of Linguistics and Phonetics (6th).* Malden, MA: Basil Blackwell.

Davies, J. & Matheson, D. (2013). The cognitive importance of testimony. *Principia:an International Journal of Epistemology*, *16(2),* 297-318. doi:http://dx.doi.org/10.5007/1808-1711.2012v16n2p297

Faulkner, P. (2000). The social character of testimonial knowledge. *Journal of Philosophy*,

97(11), 581–601.

Faulkner, P. (2002). On the rationality of our response to testimony. *Synthese*, *131*, 353–370.

Finstad, K. (2010). Interpolation and scale sensitivity: Evidence against 5-Point scales. *Journal of Usability Studies*, *5(3),*104-110.

Gilmore, L. (2016). *Tainted Witness: Why we Doubt What Women Say About Their Lives.* Columbia University Press.

Gilmore, L. (2017). Testimony. *A|b: Auto|Biography Studies, 32(2),* 307-309.

Goldman, A. (1999). Testimony. In *Knowledge in a Social World*. Oxford: Oxford University Press. Retrieved from http://www.oxfordscholarship.com.proxy.library.carleton.ca/view/10.1093/0198238207.001.0001/acprof-9780198238201-chapter-4

Graham, P. (1997). What is testimony? *The Philosophical Quarterly, 47 (187)*, 227 – 232.

Grice, H. P. (1975). Logic and conversation. In Cole, Peter, and J. L. Morgan. (1975). *Syntax and Semantics: Speech Acts,* 3. New York..

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

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy, 82,* 335-49.

Hume, D. [1777] (1975). An enquiry concerning human understanding. In P. H. Nidditch (ed.) Enquiries concerning human understanding and concerning the principles of morals, *3*, 1–165.

Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching, 14(2),* 169-175.

Lackey, J. (2005). Testimony and the infant/child objection. *Philosophical Studies*, *126*, 163 - 190. https://doi.org/10.1007/s11098-004-7798-x

Lackey, J. & Sosa, E. (2006). *Epistemology of Testimony*. Retrieved from <http://www.myilibrary.com?ID=116466>

Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge.* Retrieved from http://www.oxfordscholarship.com.proxy.library.carleton.ca/view/10.1093/acprof:oso/97

Lackey, J. (2010). Testimonial knowledge. *Routledge Companion to Epistemology*. Retrieved from http://faculty.wcas.northwestern.edu/~jal788/documents/TestimonialKnowledge.pdf

Lehrer, K. (2006). Testimony and trustworthiness. In J. Lackey and E. Sosa (eds.) *The Epistemology of Testimony*, 145-59.

Lesage, C., Ramlakhan, N., Toivonen, I., & Wildman, C. (2015). The reliability of testimony and perception: connecting epistemology and linguistic evidentiality. In D. C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P.P. Maglio (eds.), {it Proceedings of the 37th Annual Meeting of the Cognitive Science Society}, 1302--1307.

Levinson, S. C. (1983). *Pragmatics.* Cambridge, ENG: Cambridge University.

Lewis, J. (1993). Multipoint scales: Mean and median differences and observe significance levels. *International Journal of Human-Computer Interaction*, 5, 383-392.

Matheson, D. (2016). Testimonial reasons. *Erkenntnis: An International Journal of Scientific Philosophy, 81,* 757-74.

McLeod, S. A. (2015). Jean Piaget. Retrieved from www.simplypsychology.org/piaget.html

McMillan, J.H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education, 26 (1),* 3 – 29. https://doi.org/10.1007/BF00991931

Nagle, J.E. (2016). Smiling behaviors of expert witnesses. *The University of Alabama, ProQuest Information & Learning*

Noormohamed, A. (2014). 10 Fascinating facts about elephants. Retrieved from https://africageographic.com/blog/10-fascinating-facts-about-elephants/

Origgi, G. (2008). Trust, authority and epistemic responsibility. *Theoria. An International Journal for Theory, History and Foundations of Science, 23(1)*, 35-44. Doi:http://dx.doi.org/10.1387/theoria.4

Pritchard, D. (2016). Epistemology of testimony. In *Oxford Bibliographies.* Retrieved from http://www.oxfordbibliographies.com/view/document/obo9780195 396577/obo- 9780195396577-0119.xml.

Reid, T. [1764] (1997). *An Inquiry into the Human Mind on the Principles of Common Sense*. In D. Brookes (ed.), University Park, PA: The Pennsylvania State University Press.

Shields, M. & Wilkins, K. (2009). Factors related to on-the-job abuse of nurses by patients. *Health Reports.* 20(2). Retrieved from: www.statcan.gc.ca/pub/82-003-x/2009002/article/10835-eng.pdf

Siegel H. (2005). Truth, thinking, testimony, and trust: Goldman on education. *Philosophy and Phenomenological Research*, *71,* 345–366.

Sperber, D. (2010). The guru effect. *Review of Philosophy and Psychology. 1(4)*, 583 – 592.

Steup, M. (2016) Epistemology. *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu/archives/fall2016/entries/epistemology/>.

Stroll, A., Martinich, A.P. (2013). Epistemology. *Encyclopædia Britannica.* Retrieved from https://www.britannica.com/topic/epistemology.

Wagemans, J. H. M. (2011). The assessment of argumentation from expert opinion. *Argumentation, 25(3),* 329 – 339. doi:10.1007/s10503-01109225-8

Walton, D.N. (1997). *Appeal to Expert Opinion: Arguments from Authority*. University Park, PA: Penn State University Press.

Walton, D.N. (2006). Examination dialogue: A framework for critically questioning an expert opinion. *Journal of Pragmatics, 38,* 745–777