THE ROLE OF STATISTICAL NUMERACY IN COMPUTATIONAL MODELS OF RISKY CHOICE

Merle Werbeloff

A thesis submitted to the Faculty of Commerce, Law and Management,
University of the Witwatersrand, in fulfilment of the requirements for
the degree of Doctor of Philosophy

April 2020

WSG Plagiarism Declaration

I, Merle Werbeloff

(Student number: 7359680)

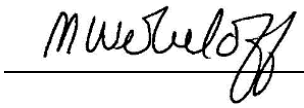
am a student registered for PADM9001: BDA00 – Doctor of Philosophy

in the year 2019/20.

I hereby declare the following:

I confirm that the work I submit for all assessment for the above course is my own unaided work. I have followed the required conventions in referencing the thoughts and ideas of others. I am aware that the correct method for referencing material and a discussion on what plagiarism is are explained in the WSG Study Guide, that these issues have been discussed in class during Orientation sessions and that there is extensive information available on the internet to assist students.

I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong. I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to correctly acknowledge the source of the ideas or words in my writing.

Signature:  Date: 20 April 2020

(assumed to be signed if submitted electronically)

Abstract

Numeracy is a strong predictor of general decision-making skill, and linked to differences in risk attitudes, such as risk aversion. However, the commonly used normative expected utility model assumes complete cognitive competence of the decision maker, and statistical numeracy is not considered directly in descriptive models of risky choice. These models are nevertheless used in policy-focused economics to assess individuals' economic welfare, regardless of the effect of statistical numeracy. Thus, if model validity is dependent on the statistical numeracy of individual decision makers, resultant policy decisions may be biased. In an online quantitative empirical study, student respondents were categorised into numeracy groups based on latent mixture analysis of responses to statistical numeracy tests. Using the students' risky choice responses to monetary lotteries, decision models were estimated using maximum likelihood parameter estimates on a subset of the data, followed by Markov Chain Monte Carlo Gibbs sampling methods for hierarchical Bayesian analysis. The results indicate significant differences between the numeracy groups on the utility parameter estimates, with risk aversion highest for low numeracy respondents. More complex models present identifiability problems. However, simpler models indicate successful outcomes in approximately two-thirds of in-sample estimates and out-of-sample predictions in the gain frame, based on parameter estimates specific to each numeracy group. The researcher proposes a numeracy-based modification to the models, citing the nudging and boosting policy initiatives of the behavioural economics literature as potential solutions to the presence of low numeracy and its effects on risky choice behaviour.

Keywords: risk, decision models, choice, numeracy, expected utility, rank-dependent, cumulative prospect theory, Bayesian modelling

Dedication

To dear Mirah

I share this with you.

Rest in Peace my friend.

Acknowledgements

I wish to thank my supervisors, Dr Michael Muller and Professor Pundy Pillay, for their encouragement and wisdom.

I thank the Dean of the Wits Faculty of Commerce, Law and Management, Professor Imraan Valodia, the Head of Wits School of Governance, Professor David Everatt, and my colleagues, for the opportunity to read for this thesis. Thank you to Dr Lynn Hewlett for your encouragement at all times.

I thank Dr Thorsten Pachur, Professor Paul Fatti, Professor George Wu, Professor Peter Wakker, Professor Nat Wilcox, Dr Sean Van Der Merwe, Dr Divan Burger and Jürgen Moller, for your insightful comments. My special gratitude goes to the late Professor John Verster who was selfless in his guidance.

To my husband, Arnold, my sons, Jason, Lance and Byron, my sister, Cheryl, family and friends, I thank you for your support. And, again, to Jason, for your generous and brilliant mind – I could not have done this without you.

Preface

In 2015, I was diagnosed with an eye disease.

My ophthalmologist had run a seemingly infinite number of tests on my eyes. She explained the condition, which she described as the “bad news”. It involved a 5% chance of progression and gradual loss of sight. The “good news,” she said, was that it was treatable through surgery, although the surgery involved a small chance of complications and risky side effects. It was my decision, she insisted. To me, the choice was clear.

What bothered me for a long time, though, was how someone in my situation who didn’t have a grasp of assessing probabilities might decide. How might such a person make an informed risky decision? Furthermore, risky decision making is ever-present. Aside from those pursuing risky medical treatments, many important decisions involve risk – decisions to purchase insurance, to buy lottery tickets and participate in other gambles, and for some people, to vaccinate your child. All these are examples of decision making under risk, as the decision maker is presented with risk in the form of objectively defined probabilities.

There is a large body of research in the psychology literature on numerical reasoning in risky choice in health and other domains. However, less attention has been paid to individual differences in knowledge of probabilities in the research on risky choice in the economic modelling literature. To address this gap, this thesis models the choices that individuals of different statistical numeracy levels make when faced with objectively defined probabilities of risk.

Table of Contents

List of Tables

List of Tables in Appendices

List of Figures

List of Figures in Appendices

List of Acronyms and Abbreviations

Acronym/abbreviation	Term
AIC	Akaike information criterion
BIC	Bayesian information criterion
BEU	Bernoulli's expected utility
BLRT	bootstrapped likelihood ratio test
BNT	Berlin Numeracy Test
C	contingency coefficient
c.i.	confidence interval
CFA	confirmatory factor analysis
CFI	comparative fit index
conf.	confidence
CPT	cumulative prospect theory
CRT	Cognitive Reflection Test
d	difference
df	degrees of freedom
DIC	deviance information criterion
ESEM	exploratory structural equation modelling
Est.	estimated/ estimate
EU	expected utility
FMM	factor mixture model
F	Fisher ratio
GAINS	Questionnaire with prospects in the gain frame
GLM	Questionnaire with prospects in the gain, loss and mixed frames
GAINS1, GAINS2	First and second administrations, respectively, of questionnaires with prospects in the gain frame
GLM1, GLM2	First and second administrations, respectively, of questionnaires with prospects in the gain, loss and mixed frames
ICs	information criteria
IP	Internet Protocol
IQ	Intelligence quotient
IRT	item response theory
IRT-2	item response theory 2-parameter model
k	R1,000
K-R 21	Kuder-Richardson Formula 21
LMR	Lo–Mendell–Rubin likelihood ratio
LS	least squares
MCMC	Markov Chain Monte Carlo
med	medium
mid	middle
MS	mean of squares
neg	negative
neg	negligible

Acronym/abbreviation	Term
OECD	Organisation for Economic Co-operation and Development
p	probability
PISA	Programme for the International Assessment of Adult Competencies
pos	positive
r	correlation
R	R programming language
RDEU	rank-dependent expected utility
RDU	rank-dependent utility
resp.	respectively
RMSEA	root mean square error of approximation
RMSR	standardized root mean square residual
SARS	severe acute respiratory syndrome
SD (*sd*)	standard deviation
SE	standard error
SL	Schwartz/Lipkus
SRMR or SRMSR	standardized root mean squared residual
SS	sum of squares
t	t value
TLI	non-normed fit index
ULS	unweighted least squares
val.	value
val 2.5% val 97.5%	95% credible intervals
WLS	weighted least squares
WLSM	robust weighted least squares
WLSMV	variance adjusted robust weighted least squares
WSG	Wits School of Governance
Z: Est./SE	Z statistic: Estimate/standard error
η_{p2}	partial Eta squared
χ^2	Chi-squared

Chapter 1: Introduction

On 30th January 2020, the World Health Organisation (WHO) declared the outbreak of a novel strain of Coronavirus *a Public Health Emergency of International Concern.* By doing so, the WHO communicated an extraordinary situation that met the criteria of a serious public health risk (COVID-19 Factsheet, 2020). On 11th March, the WHO upgraded the situation to the status of a pandemic. At the time of writing this thesis (16 April 2020), the outbreak of the potentially deadly virus, named SARS-CoV-2, continues its rapid spread internationally, with confirmation of over 2 million confirmed cases and 138,000 deaths in 185 countries or regions (Johns Hopkins University, 2020). Authorities have produced estimates of the risks associated with the virus based on objectively defined probabilities. For example, the WHO publishes probability estimates on the mortality rate of the virus (COVID-19, 2020), and epidemiologists calculate probabilities of its reproduction by estimating, for example, susceptibility to contracting the virus, and the transmission rates between infected and healthy individuals (Gorman, 2020).

The SARS-CoV-2 virus presents multiple examples of decision making under objectively defined risk estimates. At the highest level, regulatory authorities make decisions to fast-track registration of vaccines to counter a pandemic; states decide whether to restrict entry across national and international borders; those infected with mild symptoms decide whether to self-isolate from others; and people in public places decide whether to wear facemasks. Furthermore, epidemiologists model the outbreak of the virus and predict its spread and impact under various scenarios of risk. Their models are designed to aid risky decision making, given scenarios with different probabilities of events.

Models of decision making under risk have been studied over centuries and across the social and behavioural sciences. Among others, economists use neoclassical theories of risky decision making to build predictive models based on probabilities and outcomes of gambles; psychologists and behavioural economists incorporate cognitive biases into their understanding of judgment and predictions of risky choice; biologists use risk-sensitivity theories to predict risky choices of organisms (Mishra, 2014); and neuropsychologists and neuroeconomists use neurobiological and neuroimaging approaches to study the processes of decision making in the brain (Glimcher, Camerer, Fehr, & Poldrack, 2009).

Research on risky decision making has been especially acknowledged in the economic sciences. Since the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred

Nobel was awarded to Paul Samuelson in 1970, at least 15 other Nobel laureates have been recognised for their direct or indirect contributions to the study of decision making under risk in economics (Lindbeck, 1985; Wu, Zhang, & Gonzalez, 2004). The two most recent laureates recognised for their contributions related to decision making in behavioural economics are Daniel Kahneman in 2002, and Richard Thaler in 2017 (Nobelprize.org, 2002).

There is substantial literature on the relation between people's risk preferences, measured by the decisions they make under risk, and their *cognitive ability*. Cognitive ability, or general intelligence, is defined in terms of the general mental capability to manipulate, retrieve and process mental information (Lilleholt, 2019). In a recent review of the literature on risk preferences, Dohmen, Falk, Huffman, and Sunde (2018) emphasise the necessity for including the concept of cognitive ability in both theoretical and empirical research on risky decision making. Studies have found a link between higher cognitive ability and the avoidance of harmful risky behaviours in real life, such as alcoholism; cognitive ability is linked to self-reported attitudes to risk taking; and laboratory studies provide additional correlational evidence. On the other hand, individuals with higher cognitive ability tend to prefer riskier options to safer ones in financial settings.

While some researchers view cognitive ability as essential to the study of risky decision making, reviewers of the literature note that the relationship of risky choice with quantitative skills or *numeracy* is clearer and stronger than between risky choice and general cognitive ability, such as general memory tests or school achievement (Cokely, Feltz, Ghazal, & Allan, 2016; Peters et al., 2006). In particular, the relationship of risky choice with *statistical numeracy* is stronger than with general numeracy (Brand, Schiebener, Pertl, & Delazer, 2014; Dohmen et al., 2018).

Numerical ability, or numeracy, is commonly defined as one of the components (or facets) of cognitive ability. It is identified in terms of understanding and using numbers, performing basic arithmetic operations, calculating a percentage and converting probabilities to frequencies (Peters et al., 2006; Rammstedt, Lechner, & Danner, 2018).

Statistical numeracy is a component of numeracy that is specifically concerned with the practical understanding of probabilities, percentages and statistics for use in quantitative problem solving (Garcia-Retamero, Sobkow, Petrova, Garrido, & Traczyk, 2019). Cokely and colleagues (2016) conducted a series of studies on various types of judgment and

decision making, including general decisions and risky choice gambles. They also measured different components of numeracy including statistical numeracy. Their results showed that, while the various types of numeracy considered jointly strongly predicted superior decision making, statistical numeracy skill alone accounted for almost all the explanation. Furthermore, using structural equation modelling of decision-making skill, they found that numeracy, including statistical numeracy, largely explained (i.e. partially mediated) the relationship between intelligence and decision making (Cokely et al., 2016).

Other evidence of the importance of statistical numeracy for superior decision making emanates from the field of health. Decision making in medical contexts is an established subfield of the decision sciences, comprising a large body of research on numeracy and decision making. In the health field, statistics, and thus statistical numeracy, is essential for *health literacy* – the capacity to process and understand the information required to make appropriate medical decisions. Risk is expressed quantitatively in terms of percentages, proportions and survival rates, and physicians and policy makers are called upon to make risky choices between treatments and screenings (Anderson & Schulkin, 2014; Elstein, 2004; Galesic & Garcia-Retamero, 2010; Hibbard, Peters, Dixon, & Tusler, 2007; Reyna & Brust-Renck, 2014; Reyna, Nelson, Han, & Dieckmann, 2009).

Studies of decision making in financial contexts provide additional evidence of the importance of statistical numeracy in superior decision making. In real-life financial situations, people regularly face risky insurance and investment choices that involve probabilities.

However, statistical numeracy is not included in risky choice models in economics and behavioural economics. This thesis examines the potential role of statistical numeracy in these models.

In the remainder of this chapter, the key concepts of risky decision making are discussed and the challenges in the field are outlined to provide a context for the problem addressed. The researcher argues for investigating the role of statistical numeracy in the neoclassical model of risky choice and in two of the main behavioural models. The challenges in the state of research on modelling in the field of risky choice are described, in an attempt to set the study on a sound econometric and methodological base. An application for the potential results of the study are provided. Finally, the philosophical paradigm of the study is presented.

The next section of this introductory chapter defines the concept of decision making as used in this thesis.

1.1. Decision theory: decision making defined, as used in this thesis

The field of *decision making* is closely related to the field of *judgment.* While research in both fields examines responses to uncertain or complex information, they have different theoretical foundations and different foci. Decision-making research emphasises action, choice, personal preference and *overt* decision making. It is motivated by economists and statisticians who consider expected utility theory and other accounts. By contrast, judgment research emphasises appraising information, the state of mind, beliefs and feelings of the person (Goldstein & Hogarth, 1997). Judgment research is motivated primarily by psychologists who focus on the analogy between perception and judgment and the probabilistic nature of the natural environment (Hammond, 2001).

In summary, decision making concerns the overt action that a person makes, while judgement concerns the reasons for this action. This thesis is concerned with decision making only.

Decision making can occur under conditions of either *risk* or *uncertainty*. The distinction between the concepts was made by Frank Knight in 1921 (Knight, 1921/2006), and still forms the definition used in the literature on decision making today. *Risk* is defined in terms of objective probability, and is distinct from *uncertainty* which is defined in terms of subjective probability. For example, risk is involved in deciding between known side effects of a medical treatment, investments with different options, and in gambling decisions. In these examples, the outcomes are unknown but can be estimated as they are associated with objectively informed probabilities and empirical probability distributions. For example, the probabilities of the medical treatment's side effects are supplied in the accompanying literature, investment options are supplied with historical information, and winning in the casino is associated with known odds. Taking another health-related situation, the WHO estimated the probability of death from contracting the SARS (severe acute respiratory syndrome) virus objectively by the number of reported deaths relative to reported cases of infections (Roos, 2003). The fatality rate is therefore classified as a risk. However, the chances of government agencies containing the spread of the SARS-CoV-2 virus, and how quickly this may happen, are uncertain as there are too many variables to calculate the odds

of these outcomes. This situation would therefore be defined as uncertain under Knight's definition (Wakker, 2010). This thesis is concerned with decision making under risk only.

Decision making under risk in this thesis is considered from the perspective of the computational *modelling* of the structure of risky choice. Its concern is modelling for the purposes of estimating and predicting individuals' risky decisions, using formal structural mathematical models. These models are referred to as structural models, and also as computational models, in the risky choice literature. The term *model* is a special case of a theory, as it applies the theory to a particular situation or paradigm (Birnbaum, 2011).

1.2. Decision making models considered

This thesis considers three models of risky decision making: the von Neumann-Morgenstern expected utility model of neoclassical economic theory (1944, 1953), and two behavioural models of risky choice, namely, the rank-dependent utility model (Quiggin, 1981) and the model of cumulative prospect theory (Tversky & Kahneman, 1992). The latter two models are considered "major alternatives" to the expected utility model (Harrison, 2019, p. 109).

Neoclassical economic theory is the dominant paradigm of economics. It uses expected utility theory to explain the relation between phenomena, such as saving and spending. It also uses expected utility as a *normative* theory to make optimal or ideal decisions and policies (Anderson & Schulkin, 2014; Tversky, 1975). Expected utility theory is a special case of rank-dependent theory, and rank-dependent theory is a special case of cumulative prospect theory. This means that cumulative prospect theory can explain everything that rank-dependent theory can, and rank-dependent theory can explain everything that expected utility theory can (Dhami, 2016). This hierarchy provides a useful organisational structure for the thesis.

All three theories are premised on the individual acting on her own preferences. The thesis does not consider theories and decision models such as experimental game theory, which are premised on individuals having other-regarding preferences (or preferences that help other agents) (e.g. Crawford, 2013). Furthermore, this thesis does not research the cognitive processes and heuristics involved in qualitative models of risky choice, nor does it examine decisions with a time dimension, such as discounted utility models. It also does not consider the bounded-rationality models of experimental economics, for example, those of Harstad and Selten (2013) and Rabin (2013b).

1.3. The neoclassical assumption of the decision maker's complete cognitive capacity

Neoclassical economics is based on axiomatic expected utility theory first proposed by von Neumann and Morgenstern in 1944 (Von Neumann & Morgenstern, 1953). This theory assumes that individuals, or economic agents, are rational with unlimited cognitive ability. Furthermore, all of an individual's risky decisions depend on the principle of the individual rationally maximising her profit, utility or value, with no reference to cognitive or emotional processes (Dhami, 2016; Harstad & Selten, 2013). These key assumptions enable a decision maker to apply the axioms of normative expected utility theory and choose the option with the maximum utility, i.e. the option that makes her happiest, in a risky choice situation. The traditional theory is thus also referred to as *rational choice theory*. Using the terminology of Herbert Simon (1976), this theory is concerned only with *substantive rationality* – the fit of an individual's choices with her objectives. It does not consider *procedural rationality* – the process of reasoning in choice (Sugden, 2011). (In this thesis, neoclassical economics is also referred to as the 'normative theory' and the 'traditional theory'.)

By contrast, models of the behavioural economics paradigm are designed with parameters that reflect cognitive processes and approximate real-life phenomena using statistical science (Ntzoufras, 2009). Nonetheless, the models do not include statistical numeracy. (Because all the models discussed in the thesis involve decisions made under risk, the term 'decision making' will be used interchangeably with 'risky decision making'. Moreover, the researcher uses the terms 'choice' and 'decision' interchangeably.)

However, researchers, including the Nobel laureate, Herbert Simon (1955, 1956, 1977/1992), have rejected the idealised neoclassical assumption of the unlimited cognitive capacity and rationality of the individual. Simon proposed a process theory on *bounded rationality* in which the decision maker is limited by her cognition, information and time available (Simon, 1955, 1976). Following Simon, Tversky and Kahneman (1979, 1992), the latter also a laureate, produced an extensive body of evidence indicating that people's choices do not necessarily adhere to the axiomatic predictions of normative economic theory. They posited a dual-process theory of two systems: the quick acting *System 1* that is beset with cognitive biases from limited understanding of logic and statistics, versus the slower *System 2* associated with higher cognitive thinking (Dhami, 2016; Kahneman, 2011). Simon's influence also led to the research tradition of Gigerenzer and colleagues. Their focus is on how the boundedly-rational individual makes judgments and solves problems using smart and simple adaptations to uncertainty. They attribute risk illiteracy to poor risk

communication from poor representation of statistical information (Gigerenzer, 2001; Gigerenzer & Edwards, 2003).

Economists have approached the assumption of the decision maker's complete cognitive competence in different ways. Experimental economists (e.g. Rabin, 2013b) have proposed bounded rationality models of choice in an attempt to extend the conditions of the normative model, and introduce more psychological realism into economics. The instrumentalist position taken by the economist Milton Friedman (1953/2007), has been highly influential in mainstream economics (Dhami, 2016). As instrumentalists view theories as useful fictions, i.e. "…as *nothing more* than instruments" (Caldwell, 1980, p. 367) [italics in the original], Friedman claims that the purpose of assumptions is to specify the conditions under which a theory holds:

> …[T]he relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic," for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions." (Friedman, 1953/2007, p. 153)

In line with the assumption of neoclassical theory that individuals have unlimited cognitive capacity, expected utility theory should hold for respondents with high cognitive capacity as the theory assumes, but not necessarily for those with low. Friedman's view invites a test of the accuracy of the predictions of expected utility theory for individuals of different numeracy levels.

However, before the predictions of the model are tested, the challenges of past research on risky choice modelling need to be considered.

1.4. The problematic state of empirical research on risky choice modelling

The state of empirical research on risky choice modelling is not unlike the situation in other fields of science that has been labelled the *reproducibility crisis*, or the *replication crisis* in science (Baker, 2016; Shrout & Rodgers, 2018). In a meta-analytic study of the relation between cognitive ability and risk aversion, only 29% of the corresponding authors of 633 studies who were contacted were prepared to share their data or provide more details on their studies (Lilleholt, 2019). Additional evidence of replication challenges in science derives from an online survey of over 1,500 researchers by *Nature* in 2016. According to the survey, 70% of researchers in diverse fields of science tried to reproduce the experiments of others but failed, and most failed to reproduce their own results (Baker, 2016). Elsewhere, in

corroborated reanalyses of 100 published psychology studies, the rate of studies with significant results dropped from 97% to 36% on reanalysis, and original effect sizes were either overestimated or absent (Open Science Collaboration, 2015; van Aert & van Assen, 2018). The replication problem stems from multiple underlying challenges, including selective reporting, publication bias, low statistical power, poor statistical analysis and poor experimental design (Baker, 2016; Shrout & Rodgers, 2018).

In the field of risky choice modelling, prominent scholars identify two main problem areas that contribute towards the replication challenge (Bhatia & Loomes, 2017; Rieskamp, 2008; Wilcox, 2008, 2010; Zhou & Hey, 2018).

The first problematic area involves the emphasis of many decision modelers on the *structural* or theoretical aspect of risky choice models, at the expense of largely ignoring the stochastic or random component (Hey, 2001; Wilcox, 2008). As described in the following chapter, all theoretical models of risky choice share the common failing of being deterministic models. This means that they do not consider the random error and variation in the choices and preferences of decision makers. Stated in terms of decision theory, the models and their predictions are deterministic, not probabilistic. This problem is exacerbated if differences *between* individuals are also ignored. There are many such studies that fail to consider probabilistic risk attitudes, and their findings are likely to be wrong (Wilcox, 2008). Some prominent researchers in the risky choice modelling field have even suggested acknowledging the inherently stochastic nature of human decision making, and partially or fully abandoning the decision-based modelling of risky choice, opting instead for a judgment-type process orientation to try to understand the processes that people use in risky choice (Loomes & Pogrebna, 2014; Zhou & Hey, 2018).

The second problematic area involves major *methodological* deficiencies in the design of many of the empirical studies of risky choice (Harrison, 2019; Zhou & Hey, 2018). In recent meta-analyses of risky choice studies that employ different methods of eliciting risky choices with other factors controlled as far as possible, estimates of risk attitudes for the same individuals differed across methods (Loomes & Pogrebna, 2014; Pedroni et al., 2017a; Zhou & Hey, 2018). Other design factors in risky choice experiments that may influence the inconsistencies in results, include the realism or salience of the type of reward offered to participants, and the manner in which incentives are administered (Harrison, 2019; Harrison & Ross, 2017).

Thus, there are two main problem types in empirical research on risky choice, namely, a lack of consideration of stochastic error, and methodological design flaws. The combined effect of these structural and methodological challenges places doubt on the results of many studies that have compared risk preferences, for example, across sex, age and other demographics, across incentive methods, stress conditions and framing, and across choices involving losses versus gains (Bhatia & Loomes, 2017). Moreover, incorrect model predictions due to unaccounted random error will not be reproducible, and so contribute to the replication crisis in science. The flaws of these studies need to be considered before embarking on new research. As described in Chapters 4 and 5, the thesis has attempted to address these challenges.

1.5. Statistical numeracy as a source of subject heterogeneity in modelling risky choice

As previously stated, the practice of focusing only on the structural component of risky choice models and ignoring the stochastic component may result in incorrect estimates of an individual's modelled choices. This problem is even more serious if there is *subject heterogeneity* – or differences between individuals – on the structural and/or stochastic components of the model. As statistical numeracy has been found to be important for general decision making, this type of hazard may arise if there are meaningful individual differences in statistical numeracy that are not considered in the modelling of risky choice.

However, the normative model assumes the complete cognitive competence of the decision maker. Although descriptive models of behavioural economics relax this assumption, they too do not address differences in cognitive capacity or statistical numeracy of the decision maker, as heterogeneity among decision makers is not considered. Thus, the deviations from the normative model, or risk distortions, that these models posit are the same for all decision makers. For example, irrespective of the statistical numeracy level of the decision maker, the cumulative prospect theory model holds that low probabilities tend to be overweighted, and moderate to large probabilities tend to be underweighted (Tversky & Kahneman, 1992).

Importantly for this thesis, there is very little research on the role of statistical numeracy in the *modelling* of risky choice using structural models. As far as the researcher is aware, there are no empirical studies that have successfully incorporated statistical numeracy within structural models of risky choice. Some have regarded statistical numeracy as a covariate or extraneous variable to be controlled in the modelling, or examined for a different purpose (Di

Girolamo, Harrison, Lau, & Swarthout, 2015; Patalano, Saltiel, Machlin, & Barth, 2015; Xing et al., 2019). These studies are examined in depth in Section 3.6.

A dependency for an empirical modelling study of risky choice that incorporates statistical numeracy is the sound measurement of the statistical numeracy construct.

1.6. The measurement of the statistical numeracy of the decision maker

The scientific approach dominates the research in the decision sciences (Anderson & Schulkin, 2014). Influential figures in the field, such as Herbert Simon and Ward Edwards in the 1950s, stimulated empirical research on risky decision making. Particularly in the field of health, research on risky decision making has emphasised quantifying statistical numeracy, giving rise to a few frequently used measurement scales of statistical numeracy. Scales include those devised by Schwartz and colleagues in 1997, with extensions and elaborations to this scale by Lipkus and colleagues in 2001, Peters and colleagues in 2006/7, and Weller and colleagues in 2013 (Anderson & Schulkin, 2014; Garcia-Retamero & Cokely, 2014; Lipkus, Samsa, & Rimer, 2001; Schwartz, Woloshin, Black, & Welch, 1997; Peters et al., 2006; Weller et al., 2013).

Two other outstanding measurement endeavours in decision-making research involving cognitive ability and numeracy derive from other contexts. These are the Cognitive Reflection Test of Frederick (2005), which is often claimed to measure numeracy, and the more recent Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012).

Researchers who have made use of the original and newer measurement scales in risky decision-making studies have often found that the items of the health-related scales are too easy and thus do not discriminate among respondents at the higher levels of statistical numeracy, while the Cognitive Reflection and Berlin Numeracy tests are too difficult and thus do not discriminate at the lower levels of the construct. In response, some researchers have used more than one of these tests as indicators of statistical numeracy, treating these tests individually. This practice highlights the need for a consolidation of the individual statistical numeracy measures into a single measure with sound psychometric properties that is sensitive to variability across a wide range of statistical numeracy ability. However, to the knowledge of the researcher, no study has examined the factor structure *across* the items of these numeracy tests, including the Cognitive Reflection Test, and arrived at a composite

measurement scale with a known factor structure. This observation presents a gap in the statistical numeracy literature, and constitutes a necessary objective for the thesis.

The importance of an empirical modelling study that examines statistical modelling for economics and behavioural economics is discussed next.

1.7. The importance of economic modelling research: The case for welfare analysis

The modelling exercise of this thesis applies to the context of *welfare economics*. Welfare economics has been one of the common applications of neoclassical economics over the past 80 years (Sugden, 2011). In line with its commitment to the principle of individualism, normative economics tries to assess which, and how, economic policies and principles are good for each individual, all things considered. It traditionally assumes that preference satisfaction is what is good for the individual (Dhami, 2016; Hausman, 2007; Sugden, 2011). As the individual decision maker is assumed to have uncompromised cognitive capacity and access to relevant information, her preferences are assumed to be well-informed and stable, and the *considered* judgments of her own welfare. Furthermore, as preferences are revealed by choices, choices are the *primitives*, or basic concepts, of the neoclassical model of welfare economics, and can be measured or modelled by the maximisation principle of expected utility theory (Camerer, 2000; Hausman, 2007; Sugden, 2011).

Although neoclassical welfare economics assumes fully rational and informed individuals and is based on the liberal economic tradition, it nevertheless applies to policies and regulations, such as paying taxes, lockdown regulations during the COVID-19 crisis, laws regarding smoking, and the wearing of seatbelts while driving. These measures are considered as *externalities* applied on individuals. These externalities are also described as neoclassical paternalism or *heavy-handed* paternalism (Loewenstein, & Haisley, 2008 as cited in Dhami, 2016, p. 1579).

However, behavioural economists have a different view. Far from seeing the decision maker's choices as revealing her well-informed, considered preferences as best for her welfare, they see her preferences as *constructed*, influenced by heuristics, cognitive biases, and distortions, and thus not good for her welfare. Accordingly, most behavioural economists believe that her judgments need to be *re*constructed through policy adjustments so they are best for her welfare. To do so requires the principle of *paternalism* (Dhami, 2016).

Behaviour welfare economics attempts to reconstruct the underlying judgments of such individuals by addressing their *internalities* (Bhargava & Loewenstein, 2015). *Soft*

paternalism – or *light paternalism* – is the view that the state may intervene in an individual's choices if her rationality is compromised. The term is used in the context of policy interventions that behavioural economists advocate (Dhami, 2016). More specifically, *libertarian paternalism* is the term used by Thaler and Sunstein (2008) to describe attempts to influence or 'nudge' an individual to choose to change her behaviour in a way that will improve her economic welfare. According to Thaler and Sunstein, an individual – for example a boundedly rational decision maker – would be nudged in a direction of improved welfare as determined by a *benevolent planner* or *choice architect*. A parent, employer, and government agency are examples of such. Furthermore, the authors conceptualise the notion of nudging as associated with choice and sovereignty, with a minimum of intrusion and coercion (Dhami, 2016; Thaler & Sunstein, 2008). One of the examples that they provide is designing a form for new employees to complete in such a way that the option to enrol in a company health plan is positioned so they are enrolled by default.

The relevance of the thesis to the discussion on policy intervention is considered in terms of quantifying and addressing the gap between normative and behavioural models (Harrison & Ross, 2017). The behavioural economist who advocates paternalism may quantify the potential for enhancing the welfare of an individual by comparing model predictions of the normative expected utility model against those of a behavioural model, such as the rank-dependent or cumulative prospect theory models. Under the paternalism view, if this thesis were to find that the models apply differently to individuals of different numeracy levels, it could imply that certain behavioural policy interventions may be more appropriate than others for individuals of different numeracy levels. Improvement in the understanding of an individual's behavioural risky choices through considering her statistical numeracy would contribute to more reliable and accurate estimates of her risk preferences (Barseghyan, Molinari, O'Donoghue, & Teitelbaum, 2018; Brañas-Garza & Smith, 2016). In turn, the more reliable estimates of these models could improve the evaluations of the welfare implications of policy changes (Taylor, 2013). Alternatively, the thesis may find the models to be misspecified for individuals of a certain numeracy level, thereby leading to misleading predictions for these individuals. This finding, though negative, would be useful knowledge.

Policy interventions other than paternalism are discussed in Chapter 6. These interventions include, for example, decision aids for interpreting probabilities in medical literature, such as graphical portrayal of probabilistic risk (Garcia-Retamero & Cokely, 2014; Petrova, Garcia-Retamero, & Cokely, 2015). The appropriate risky choice models could provide more

accurate quantification of the potential effect of such policy interventions if applied to individuals in specific numeracy groups, compared to if applied without considering their numeracy levels.

The following sections on the problem statement, purpose statement, main research question and proposition, provide a more concise summary of the argument of the thesis thus far.

1.8. Problem statement

Numeracy, and statistical numeracy in particular, is a strong predictor of decision-making skill. Furthermore, statistical numeracy competence has been linked to differences in risk attitudes, such as risk aversion. As objectively stated probabilities are inherent in risky choice, one may expect that competence in statistical numeracy would be important in economic models that estimate and predict risk preferences. However, although consideration is given to numeracy in process models of risky choice, and as a correlate in some studies that involve modelling, there is a lack of adequate formal studies of the role of statistical numeracy in normative and descriptive modelling of risky choice. The normative expected utility model assumes complete cognitive competence on the part of the decision maker and does not consider the less-idealised individual. Descriptive models, such as the rank-dependent model and the cumulative prospect theory model, relax this assumption but do not address the statistical competence of the decision maker directly. Accordingly, distortions in outcome values and probabilities, for example, loss aversion for lotteries framed as gains, continue to be described irrespective of the characteristics of the decision maker (see Section 2.2).

Models of risky choice are used in policy-focused economics to assess the economic welfare of individuals, regardless of the effect of statistical numeracy. If the validity of these models is dependent on the statistical numeracy of individual decision makers, policy evaluations and resultant policy decisions may be biased, or inaccurate, for some individuals. Consideration of differences between statistical numeracy groups on these models' parameter estimates and predictions may offer policy makers a better understanding of individuals' behavioural choices under risk, and improved model predictions. More accurate and valid model predictions may identify those individuals for whom policy interventions may be required to address cognitive distortions in probabilities, dependent on their statistical numeracy. In turn, the potential costs and benefits of such interventions could be quantified more accurately. Alternatively, the modelling exercise may highlight misspecification of

structural models for individuals of certain numeracy levels. This finding would also be useful for identifying when model predictions are invalid. In summary, identifying the moderating role of statistical numeracy in models of risky choice may enhance the predictive validity of risky choice modelling in policy interventions.

1.9. Purpose statement

The ultimate purpose of this study is to estimate the latent structure of risky choice models for individuals of different statistical numeracy levels. To achieve this purpose requires an adequate measure of the construct of statistical numeracy across a wide range, with known factor structure and sound psychometric properties. As a measurement instrument of this description was not available in the decision science literature, the construction of such, based on a composition of existing measures, was identified as a dependency of the study. Thus, construction of the numeracy instrument is considered as a second purpose of the study.

For different levels of respondents' statistical numeracy, the thesis considers the structure of the models under expected utility theory, rank-dependent utility theory and cumulative prospect theory. The estimation is focused on the risk preference parameters of the models and their predictions in order to evaluate the role of statistical numeracy in the modelling of risky choice.

Finally, the implications of the findings on the role of statistical numeracy in risky choice modelling are discussed in the context of policy intervention.

1.10. Main research question of the thesis

The thesis has one main research question, referred to as the *thesis question*. It is framed as follows:

What is the role of statistical numeracy in computational models of risky choice?

The thesis question comprises two supporting research questions, each with sub-questions of their own. As they require concepts not yet described in this chapter, they are presented at the conclusion of Chapter 3 (see Section 3.10), following the literature reviews of decision making and numeracy research.

1.11. Main proposition

The thesis proposes that the statistical numeracy of decision makers under risk is a meaningful source of heterogeneity in the modelling of risky choices. That is, the parameters of utility models of risky choice are expected to differ substantively when applied to individuals of different numeracy levels.

The following section presents the philosophical questions that give rise to the paradigm used in this study.

1.12. Philosophical paradigm

This section discusses four elements of the philosophical paradigm used in this thesis: the ontology, metaphysics, epistemology and politics of the research.

Ontology is the philosophical study of *existence* – it investigates whether particular phenomena exist (Hofweber, 2018); metaphysics is the study of the *nature of reality*, including questions around the nature of causation, and the relationships among objects and among concepts (van Inwagen & Sullivan, 2018); epistemology concerns the nature of *knowledge* (Steup, 2018); and political philosophy concerns the *legitimacy of authority* (Peter, 2017).

Within the context of the thesis question, these four philosophical areas yield four questions:

Ontology: First, does statistical numeracy exist as a physiological structure within the brain; and, second, do the cognitive processes assumed by the models exist? For example, all three models considered in this thesis assume risk aversion, and cumulative prospect theory assumes probability weighting or distortion, and loss aversion. Do these exist as cognitive processes within the decision maker?

The researcher is agnostic about the ontological question, in both senses.

First, the research in this thesis is consistent with either the existence of a physiological structure within the brain responsible for statistical numeracy, or there being no such structure within the brain. Second, the researcher is agnostic on the question of whether cognitive processes like risk aversion, probability weighting and loss aversion actually occur in the decision maker.

However, the researcher *is* interested in whether the decision maker *behaves as if* she has a particular level of statistical numeracy and behaves as if she is following these cognitive

processes. That is, the researcher assumes *either* an objectivist *or* an instrumentalist approach toward statistical numeracy and cognitive processes. That is, the researcher assumes either that these cognitive capacities exist in areas of the brain, or that agents act as if they do. What matters for the researcher in this thesis is whether the constructs of statistical numeracy, risk aversion, probability weighting, loss aversion, and other cognitive processes are useful in predicting and explaining risky choice. A notable example of the instrumentalist position is provided by Friedman (1953/2007) who claimed that the goal of science is predictive rather than explanatory. For Friedman an economic model is an instrument meant to predict economic phenomena accurately, rather than to describe the processes underlying the activity. Thus, the assumptions and the processes underlying the theory are less important than whether the theory provides predictive accuracy. He called such models, *as-if* models.

Metaphysics: What are the conditions under which the estimates and predictions of the risky choice models hold? For example, do the risky choice models estimate and predict the risky choices of individuals with high numeracy levels, but not those of low numeracy levels? In statistical terminology, this question asks whether statistical numeracy *moderates* the predictions of the models.

The researcher does not attempt to specify or investigate all the possible conditions under which the risky choice models hold. There are many competitor variables, other than statistical numeracy, which may specify conditions for risky choices to occur as predicted by the models. However, this thesis does investigate statistical numeracy, specifically as a moderating variable in risky choice, since probabilities underlie statistical numeracy and are also inherent, by definition, in every risky choice situation.

The researcher is agnostic about whether statistical numeracy, the potential moderator, is the underlying *causal* variable that explains why the models hold in some conditions but not in others. Only its relationships to the model predictions are relevant here.

Epistemology: What can we know about the relation of statistical numeracy to risk attitudes? For example, what is the relation between statistical numeracy and pessimism, or risk aversion, in a risky choice situation, and how does statistical numeracy relate to optimism, or risk seeking? What is its relation to distortions in objectively-presented probabilities? What research methodology should we use to achieve this knowledge?

The researcher is committed to a thorough statistical analysis of the model estimates and predictions under different levels of statistical numeracy. This analysis requires careful

specification of the research design, considering the pitfalls identified in the existing literature on risky choice. It also requires a large set of quantitative data. The analysis of this data set relies on a combination of Bayesian and frequentist statistical methods and is purely quantitative in nature.

Politics: First, should the state implement policies to aid the statistical numeracy of its citizens? Second, if so, which policies should it implement? For example, should the state adopt a libertarian position as espoused by neoclassical economic theorists? Or should it rather adopt the soft paternalism of behavioural economists who seek to reconstruct the underlying judgments of individuals through policy interventions (Bhargava & Loewenstein, 2015)? Third, which persons should the policy interventions be directed at? Should they be directed at, for example, individuals with low numeracy, middle numeracy or both?

The researcher is agnostic on many of the political questions raised by research on decision making and statistical numeracy. Her concern is with the difference in the predictive accuracy of the models under conditions where the decision maker has different levels of statistical numeracy. However, the results yielded by this research could be used by policy makers when deciding whether to implement interventions to aid the statistical numeracy of its citizens (first political question). In Section 6.3.3, the researcher discusses the ways that educational policy could be altered to improve the long-term numeracy of the South African population as a whole.

Paradigm: Different research paradigms take different stances on each of these four philosophical debates (ontology, metaphysics, epistemology and politics). Furthermore, different texts split the field into distinct sets of paradigms. For example, Van de Ven (2007) refers to positivism, relativism, pragmatism, and realism. Chilisa & Kawulich (2012) argue that the field of paradigms comprises positivism, post-positivism, constructivism/interpretivism, transformative/emancipatory, and postcolonial/indigenous. It is beyond the scope of the research to define all the available paradigms. The epistemology of this research endeavour is entirely quantitative in nature, and the purpose of the research is to test theory and predict the results of theoretical models under different conditions. Thus, the choice of research paradigms most relevant for the study is narrowed to either positivism or post-positivism, so the researcher will focus on defining these paradigms.

Positivism uses the scientific method of the natural sciences, such as physics and chemistry, to establish knowledge of social phenomena. It uses a strict empirical approach to establish

objective, precise observations and causal relationships. On ontology, positivists believe that social phenomena (for example, statistical numeracy) exist independently of the researcher and are measurable. On metaphysics, positivists hold that the relationships among social phenomena (for example, between statistical numeracy and cognitive processes) are objectively real and independent of the researcher. On epistemology, positivists seek to verify the truth of their research enquiry, without a margin of error. They use the quantitative methods of the natural sciences to collect empirical data and establish objective causal relationships that are stable and generalisable (for example, the neoclassical model predictions should reflect the actual choices of the decision maker). On politics, positivists strive to be agnostic, as they see science as value-free and independent of any political system (Chilisa & Kawulich, 2012).

Post-positivism relaxes the strict scientific principles of positivism. As in positivism, post-positivists subscribe to the scientific method of objectivity and generalisability. By contrast, they acknowledge that observations are unlikely to be independent of the observer's human limitations and biases, and measurements are unlikely to be free of error. Accordingly, theories can be changed, and models can be modified. On ontology, post-positivists believe that social phenomena (for example, statistical numeracy) exist independently of the observer. They are thus objective, but observations are constrained by the limitations of the observer, and measured with some error. On metaphysics, post-positivists hold that the relationships between social phenomena (for example, between statistical numeracy and cognitive processes) are objectively real and independent of the researcher but cannot be established with certainty. On epistemology, post-positivists seek to falsify theories that do not hold, instead of seeking to verify theories as the truth. They, too, use the quantitative methods of the natural sciences to collect empirical data and attempt to establish objective causal relationships that are stable and generalisable. On politics, post-positivists seek no political base, as science is value-free and independent of any political system. However, they acknowledge the researcher's biases (Chilisa & Kawulich, 2012).

The post-positivist paradigm is the more appropriate and realistic framework for the thesis. The research cannot conform to the strict principles of the natural sciences. The nature of the research on risky choice conforms more to the psychological and social sciences where strict control, error-free measurement, and the causal nature of relationships are relaxed. The types of analyses conducted in this research do not assume causality, and so are offered within the bounds of probability appropriate under a post-positivist view.

1.13. Structure of the thesis

This introductory chapter to the thesis has cited evidence of the relevance of statistical numeracy for risky choice situations. The main conclusion of the chapter is that the normative model and prominent descriptive models of risky choice do not consider statistical numeracy in their predictions. The chapter also describes an application of the research findings for welfare economics. It concluded with a description of the philosophical paradigm that grounded the study. Much of the content of this chapter is discussed in more depth in later chapters, as the empirical aspects of the research are developed, and the findings are presented.

The remaining chapters are briefly summarised as follows:

Chapter 2 presents a review of relevant literature on risky choice modelling. The chapter presents first the traditional (micro)economics and behavioural economics literature on decision making under risk, including the history of the normative model and the two behavioural models of risky choice, their features and the associated functional forms, sources of stochastic error, analysis methods and various empirical design specifications. It presents the theoretical argument within the academic literature on the need to incorporate statistical numeracy in models of risky choice. The development of models of risky choice and relevant literature on the numeracy literature are described.

Chapter 3 presents the literature on cognitive ability, numeracy and statistical numeracy, as well as other numeracy-related literature relevant to the modelling of decision making. It presents the positions of Simon, Tversky and Kahneman, and Gigerenzer and colleagues, who link the individual's cognitive ability to decision making. It presents empirical studies of statistical numeracy and risky choice, and methodological caveats in empirical literature on risky choice and cognitive ability. It critiques empirical studies that have modelled risky choice in the context of statistical numeracy. It outlines measures of statistical numeracy in decision-making studies. Thereafter, it presents the theoretical framework of the thesis, and a description of its conceptual framework. It concludes with an elaboration of the research questions and sub-questions.

Chapter 4 presents the literature on the research methodology of the thesis. The topics covered include the quantitative methodology approach, sampling, the approach used to construct the questionnaires and details of the content of the questionnaires, the procedure and ethics, and the methods of analysis used for the numeracy and the risky choice data.

Chapter 5 presents the results of the thesis, structured by the research questions and sub-questions. It includes key tables and figures.

Chapter 6 discusses the results of the thesis, once again structured by the research questions and sub-questions. Findings on the research objectives are discussed with reference to key theories on risky decision making and numeracy. Implications for public policy are discussed.

Chapter 7 concludes the thesis, highlighting the implications of the main findings of the research. The results are synthesised in the context of risky decision making and statistical numeracy, and recommendations for future research are made.

Chapter 2: Literature review: Decision making

This thesis spans topics within economics, behavioural economics, econometrics, cognitive psychology and psychometrics. Owing to this interdisciplinary range, the supporting literature is split over two chapters. Chapter 2 deals mostly with theoretical aspects of decision-making models from the economics and behavioural economics literatures. Chapter 3 deals with the theoretical aspects of numeracy from the psychological, judgment and decision-making literature, as well as empirical studies from cognitive psychology and other literatures that relate cognitive ability and numeracy to risky choice.

There are three main sections in Chapter 2, each of which addresses a fundamental aspect of the mathematical modelling of risky choice: structural forms (2.1); functional forms (2.2); and probabilistic forms (2.3).

2.1. Model specification (i): Structural forms of risky choice modelling

Theorising about individuals' rational behaviour and mathematical modelling of decision making began in the 17th century and continues today (Heukelom, 2007). The most prominent theories include Pascal's expected value theory (1654), Bernoulli's expected utility theory (1738), classical and neoclassical economic theories of consumer choice and behaviour, von-Neumann-Morgenstern's expected utility theory (1944, 1947, 1953), the rank-dependent utility theory of Quiggin (1981) and Tversky and Kahneman's cumulative prospect theory (1992). There are several other theories and models of risky choice, including Loomes and Sugden's regret theory (1982), Birnbaum's configural weighting model (1992), and decision field theory (Busemeyer & Townsend, 1993). These theories have merits in their own rights. For example, decision field theory is a powerful cognitive model of decision making involving sequential sampling. However, although such models aim to provide a cognitive account of decision making, their incorporation into the thesis would have involved an expanded research design, which was beyond the scope of the thesis. Restricting this thesis to three frequently cited decision models in the utility tradition of the risky choice literature (Harrison & Ross, 2017) allows a cross sectional approach using a single online measurement event per respondent, which provides a manageable scope for the thesis.

This section (2.1) deals with the history, theoretical structure and features of the models, and the reactions of researchers to each. As most of this content is focused on theory rather than methodology, this section is targeted at the latent or theoretical level of risky choice modelling.

2.1.1. Expected value theory (Pascal-Huygens-de Fermat 1654)

The earliest attempts to model decisions involving risk date to the mid-17th century when philosophers and mathematicians exchanged letters over monetary gambles, or "games of fortune", and gave advice to noblemen (Hacking, 2006; Schuessler, 2016). In so doing, Blaise Pascal, Christiaan Huygens and Pierre de Fermat debated the fair price of participation in a lottery. They posited this price to be its expected monetary payoff, calculated as the sum of the monetary values of the outcomes of the lottery weighted by their respective probabilities. Thus, a lottery that offered 100 ducats (the currency of the day) with a probability of .5, and 50 ducats with a probability of .5 had an expected value of 75 ducats (.5x100 + .5x50 = 50 + 25). As the decision maker was assumed to *maximise* her winnings, this lottery would be chosen over another of lower expected value (Moscati, 2018). The expected value payoff method, whereby the lottery with the higher expected value is selected, became known as *expected value theory* (Schoemaker, 1982). Pascal is considered to have made the first significant contribution to the theory of probability and is also credited with having contributed expected value theory to what is now referred to as decision theory (Hacking, 2006).

The expected value (EV) of a prospect, bet or gamble with *n* outcomes, is represented in equation 1, where p_i is the probability of outcome i; x_i is the value of outcome i:

$$\mathrm{EV} = \sum_{i=1}^{n} p_i \, x_i \qquad (1)$$

The model involves a weighted sum of probability and value.

2.1.1.1. Reactions to expected value theory

Expected value maximisation was first challenged in 1670, when Pascal considered *personal* values instead of *monetary* values in a situation of uncertainty (Keren & Wu, 2015). A second major challenge to expected value theory was in 1738 when Daniel Bernoulli addressed a paradox in a game of chance suggested by his cousin, Nicholas Bernoulli, in correspondence with the mathematician Montmort in 1713 (Schoemaker, 1982). The game, known as the *St Petersburg problem*, involved flipping a fair coin – or ducat – repeatedly (*n* times) until a head is produced. The payoff for flipping the coin with the outcome of a head is 2^n ducats. The paradox of this game is that its outcome contradicts the prediction of expected value theory, as people are willing to pay relatively little to play, even though the expected value of the payoff is infinite as the prize keeps growing. (It is assumed that the player has an

infinite amount to stake.) The proposed solution to the game became the framework for the utility theory of choice.

2.1.2. Expected utility theory (Bernoulli 1738)

Daniel Bernoulli's proposed solution to the St Petersburg problem provided the framework for his expected utility model. [The paradox was also solved independently by Cramer, and a later solution was the risk-return model proposed by Markowitz in 1959 (Machina, 1987).]

Daniel Bernoulli's proposal was based on *personal* values rather than *monetary* values. The solution to the St Petersburg paradox was to multiply probabilities by personal *utilities* – or values of pleasure – rather than by monetary values. He thus arrived at the concept of moral expectation rather than mathematical expectation (Zabell, 1990). Bernoulli's moral expectation, or expected *utility*, was a function of the logarithm of monetary value. The logarithmic function is a marginally increasing function, and implied that the utility of the game increased at a diminishing rate and was thus finite. This utility function resolved the St Petersburg paradox as the utility of the game increased only marginally. Bernoulli's early idea of maximising expected *utility*, rather than maximising expected objective *monetary* return, became known as Bernoulli's expected utility theory.

Bernoulli's expected utility (BEU) of a prospect with n outcomes is represented in equation 2, where p_i is the probability of outcome i, x_i is the value of outcome i, and $u(x_i)$ denotes a logarithmic function defined on the outcome x_i under risk.

$$\text{BEU} = \sum_{i=1}^{n} p_i \, u(x_i) \qquad (2)$$

As in the case of the expected value model, Bernoulli's expected utility model uses algebra in which the utilities of the choices are multiplied by their likelihood or probabilities, or some function thereof.

2.1.2.1. Reactions to Bernoulli's expected utility theory

In the context of the intellectual and scientific progress of 18th century, Bernoulli's expected utility theory was the accepted model of human reasoning (Bristow, 2017). His theory represented a radical shift in the evaluation approach to gambles as it conceptualised subjective utility rather than objective wealth outcome (Wakker, 2010). It effectively introduced the concept of risk aversion and provided the framework for modern-day algebraic

models of decision making (Brandstätter, Gigerenzer, & Hertwig, 2006; Heukelom, 2007; Kahneman, 2011).

However, although the theory was studied extensively by the mathematicians of the 18th century, it was relatively ignored by the economic community. Bernoulli had not addressed why expected utility should be considered as rational, nor had he addressed how to fulfil the difficult task of measuring utility (Keren & Wu, 2015). Furthermore, it is speculated that scholars of the time insisted on clearly defined theories, and so disapproved of the imprecision of the utility function inherent in his utility theory (Hens & Rieger, 2016). By the late 1830s, the idea of a probability model of rational human reasoning ended, marking the beginning of the end of Bernoulli's expected utility theory.

2.1.3. Classical economic perspectives of choice behaviour

Efforts to understand choice behaviour in the economic community relate to early economic theory and the classical period of economics. In Adam Smith's 1759 publication, *The Theory of Moral Sentiments*, he described a theory of human behaviour with psychological insight into the choices people make. However, unlike Bernoulli, he viewed choice as a conflict between passionate short-term gratification, and impartially assessed long-term costs (Ashraf, Camerer, & Loewenstein, 2005). In *The Wealth of Nations* (1776), he further described the psychological determinants of choice behaviour and how choices aggregate into market activity (Glimcher et al., 2009).

In the early 1870s, the economists Stanley Jevons, Carl Menger and Léon Walras independently posited a utility-based view for explaining consumer behaviour and the exchange values of commodity prices (Moscati, 2018). Their utility approach determined a commodity's exchange value in terms of the utility it held for an individual. This approach posited that a commodity's exchange value is determined by its *marginal utility* or "the additional utility associated with an individual's consumption of an additional unit of the commodity" (Moscati, 2018, p. 1). This marginal utility diminishes as its consumption increases. The concept of marginal utility was used in major theories of prices, exchange and markets and, until 1945, it was used to evaluate market equilibria in the context of social welfare.

2.1.4. Neoclassical economic models of consumer choice and behaviour

Beginning in the 1930s, economists including Samuelson, Arrow and Debreu, began to use mathematical models of consumer choice and behaviour. Their style of modelling used

simple, primitive assumptions, known as axioms. The models were strongly normative as they were focused on idealised choices between goods, or combinations of goods and efficient allocation of resources using choice-based data to reveal unobservable preferences (Glimcher et al., 2009). This approach became known as the theory of revealed preference (Wakker, 2010).

Revealed preference in economic theory subsumed the underlying psychological reasons or motivations, so that preferences and utility were described as "all things considered" (Moscati, 2018, p. 142). This meant the psychological reasons for the preferences were no longer relevant for understanding choice behaviour. Furthermore, until the late 1930s, economic theorists did not include risk in their discussions on utility nor in their debates on how it could be measured. For example, in Pareto's 1906/1909 optimality criterion of reallocating resources, his preferences involved comparisons between a particular quantity of wine and a particular quantity of bread (Moscati, 2018). However, in the 20th century, when economists started to analyse economic decisions involving risk and uncertainty, they needed a theory of risky choice. Although there were some marginal utility theorists who had adopted Daniel Bernoulli's expected utility theory until the early 1900s, it had been virtually discarded by the economic community by the late 1930s. In 1944, the expected utility approach was revitalised in modified form when John von Neumann and Oskar Morgenstern published the *Theory of Games and Economic Behavior* (Moscati, 2018).

2.1.5. Von-Neumann-Morgenstern expected utility theory (von Neumann-Morgenstern, 1944, 1947, 1953)

Von Neumann and Morgenstern used the axiomatic approach of mathematics to develop a solid foundation for rational decision making under risk. They showed that if the preferences of the decision maker between risky alternatives satisfied certain axioms (self-evident mathematical principles), then she would prefer the risky alternative with the highest expected utility (defined below). Their axioms thus specified constraints on behavioural choices that, if satisfied, implied choosing the alternative with the maximum expected utility (Goldstein & Hogarth, 1997). Additionally, by the mid-20th century, expected utility theory had been formally axiomatised through the works of Frank Ramsey on the philosophy of belief, and Leonard Savage on the foundations of statistical inference and decision making under conditions of uncertainty (Keren & Wu, 2015; Machina, 1987).

Expected utility is calculated using a utility function, the existence of which is warranted by the axioms. The von-Neumann-Morgenstern expected utility (EU) of a prospect is represented in equation 3, where p_i is the probability of outcome i, x_i is the value of outcome i, and $u(x_i)$ denotes a utility function for obtaining the outcome x_i under risk, for a lottery with n outcomes (Schoemaker, 1982):

$$\mathrm{EU} = \sum_{i=1}^{n} p_i \, u(x_i) \tag{3}$$

In von-Neumann-Morgenstern expected utility theory, rationality requires a number of conditions. First, the decision maker must be capable of weakly ordering preferences over lotteries: for two lotteries A and B, she can prefer Lottery A over Lottery B, or *vice versa*, or be indifferent between them. This weak ordering (as opposed to strict ordering when one lottery is preferred to the other) is known as the completeness condition. Rationality further implies the condition of transitivity: if the decision maker prefers Lottery A to B, and Lottery B to C, then she will prefer Lottery A to C. Similar transitivity holds for indifference relations.

The second requirement for rationality in expected utility theory is the principle of maximisation, i.e. that the decision maker chooses the lottery that maximises expected utility (Edwards, 1954). Further conditions of rationality refer to the independence of irrelevant alternatives, and the invariance assumption. The former condition implies that a decision maker's preference is independent of other options; the invariance condition implies that preferences of logically equivalent options should not be affected by the format in which they are presented (Heukelom, 2007).

From here on, the term expected utility theory refers to von-Neumann-Morgenstern expected utility theory rather than to the expected utility theory of Bernoulli.

2.1.5.1. Reactions of economists to expected utility theory

Von-Neumann and Morgenstern published their proofs of the expected utility theorem in 1947, a time when there were other theories of decision under risk (as discussed by Kenneth Arrow in 1951, cited in Moscati, 2018). For the next five years, their theory inspired intensive debate among eminent theorists, including Friedman, Savage, Marschak, Samuelson, Baumol, Strotz, Alchian and Ellsberg. These theorists deliberated on the axioms' plausibility, and the descriptive ability of the theory (Moscati, 2018). Finally, von Neumann

and Morgenstern's axiomatic expected utility theory was accepted as the normative model of rational decision making under both objective and uncertain probabilities of risk (Schoemaker, 1982).

2.1.5.2. Main features of expected utility theory

As described by Dhami (2016), expected utility theory has four main features:

First, expected utility is linear in its probabilities, i.e.

$$EU(x_1, p; x_2, 1\text{-}p) = pEU(x_1) + (1\text{-}p)EU(x_2)$$

assuming a risky prospect with two outcomes, where p is the chance of receiving outcome x.

Second, utility typically depends on the final wealth level of the decision maker. [The final wealth is the sum of the outcome of the lottery and any external side-payment, for example, an endowment or incentive (Wakker, 2010).]

Third, risk neutrality corresponds to a linear utility function, risk aversion to a concave utility function, and risk seeking to a convex utility function (see Figure 1 for illustrations of concave and convex functions). By contrast, expected value theory restricts decision makers to a utility function that is linear and thus to risk neutrality.

Fourth, utility does not differ for prospects framed as gains versus losses.

2.1.5.3. Risk attitudes under expected utility theory

According to axiomatic utility theory, utility functions directly reflect attitudes to money, and only indirectly reflect attitudes to risk (Tversky, 1975). In 1948, Friedman and Savage used the preferences derived from expected utility theory to infer the shape of utility functions, and thereby explain attitudes to money.

Friedman and Savage used an *S*-shaped utility curve of money to explain why most people, irrespective of income levels, tend to buy insurance (risk-averse behaviour), at the same time as buying lottery tickets or engaging in other forms of gambling (risk-seeking behaviour). They explained these seemingly contradictory coexisting behaviours using the concavity region of an *S*-shaped, or sigmoid, utility curve to reflect the diminishing marginal utility of money, and the convexity region of the curve to reflect the risk-seeking behaviour associated with buying a lottery ticket (Figure 1). They defined a person as risk averse if she preferred a sure, or guaranteed, amount of money over a lottery with more risk but the same expected

payoff, and a person as risk seeking if she preferred the lottery with the expected payoff over the sure amount of money.

Risk aversion and risk-seeking attitudes are depicted in the left- and right-hand panes of Figure 1, respectively, reproduced from Machina (1987). In the figure, the graphs in both panes refer to a gamble that offers a probability of ⅔ for the outcome x′ and a probability of ⅓ for the outcome x′′. Thus, for both panes A and B, the expected value EV(x) = ⅔x′ + ⅓x′′ = $\bar{x}$.

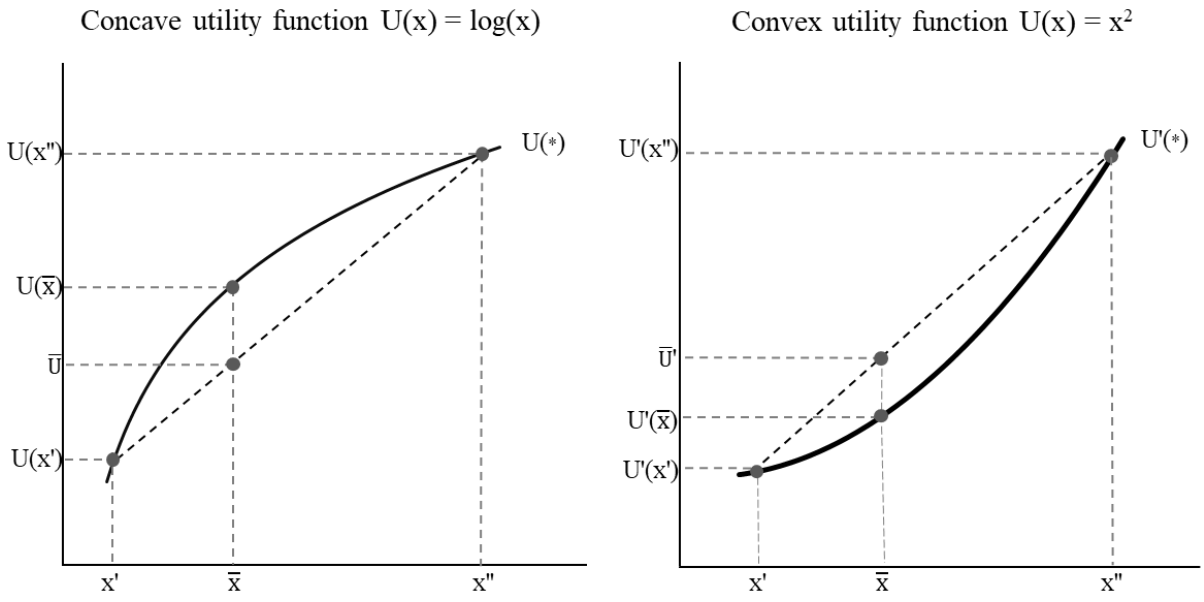


Figure 1: Concave utility function for risk aversion; convex utility function for risk seeking
Note: Reproduced from Machina *(1987, p. 538)*

In the left-hand pane of Figure 1, the utility of the ‘sure thing’ or expected value of the outcomes $\bar{x}$ is U($\bar{x}$), i.e. the point on the bowed down utility curve. This value is greater than the probability-weighted sum of the individual outcomes $\bar{u}$ i.e. ⅔U(x′) + ⅓U(x′′), using the utilities of the extreme points on the line. Thus, a person whose risk is reflected by this concave utility function would derive greater utility from selecting the bet with the expected value, than the bet offering the ⅔ : ⅓ odds with utility $\bar{u}$. Accordingly, she is said to be risk averse. The steeper – or more concave – the utility function, the more risk averse her preference compared to the sure thing.

By contrast, the utility of the expected value of the outcomes in right-hand pane of Figure 1, is U’($\bar{x}$). The utility of the expected value of the outcomes is less than the probability-weighted sum of the individual outcomes $\bar{u}$’. Thus, a person with this convex utility function

would derive less utility from selecting the bet with the sure thing or expected value, than the bet offering the ⅔ : ⅓ odds. She is thus said to be risk seeking. The steeper – or more convex – the utility function, the higher the person's risk seeking preference compared to the sure thing (Machina, 1987, p. 538).

2.1.5.4. Empirical tests of von-Neumann-Morgenstern expected utility theory

To test the descriptive and predictive validity of expected utility theory, researchers conducted controlled laboratory experiments to verify the shape of the utility curves derived from these risky choices (Moscati, 2018). In one such experiment, published in 1951, Frederick Mosteller and Philip Nogee observed conflicts with expected utility theory. Contrary to contemporary thinking, their participants were "… not so consistent about preference and indifference as postulated by von Neumann and Morgenstern" (Mosteller & Nogee, 1951, p. 404). This pronouncement was one of the first references to the stochastic nature of risky choice, and in direct contradiction to the deterministic nature of expected utility theory (Section 2.3). Another of their key objections was that the theory did not consider differences between objective probabilities and a person's psychological interpretation of these (Mosteller & Nogee, 1951).

However, experimental utility researchers of the 1950s continued to design their studies to ensure optimal conditions for supporting the theory (Davidson, Siegel, & Suppes, 1955; Moscati, 2018). At a major conference on expected utility in Paris 1952 and at an *Econometrica* symposium on the theory later in the year, most theorists expressed their public support for the validity of von-Neumann-Morgenstern expected utility theory (Moscati, 2018).

2.1.5.5. Early challenges to expected utility theory: The Allais paradox (1952)

There was, however, some opposition to the theory. One of the 1952 Paris conference organisers, a French economist named Maurice Allais, expressed his opposition informally to Savage by presenting him with a choice situation between pairs of lotteries that violated expected utility theory. His presentation involved two probability problems, one based on differences and the other on ratios. Collectively known as the *Allais paradox*, these problems revealed that the probabilistic linearity assumption of expected utility theory is violated if the probabilities are reduced by linear transformation. Under expected utility theory's principle of invariance to linear transformation, a person who prefers a chance p of winning an amount v, to a chance p' of winning an amount v', should have the same preference if the

probabilities *p* and *p'* are reduced by linear transformation. This would mean, for example, that a person who chooses the safer option of R1 million for sure, over a .9 chance of winning R2 million should also prefer the option of a .01 chance of winning R1 million over a .009 chance of winning R2 million (probability values having been divided by 100). This paradox is also known as "the common ratio effect (choices change when the probabilities in a pair of gambles are scaled by a common factor) and the common consequence effect (choices change when probability mass in a pair of gambles is shifted from one common consequence to another)" (Gonzalez & Wu, 1999, p. 134).

Allais's illustrations of shifts in people's preferences exposed normative expected utility theory as descriptively inadequate, as the expected utility formula could not account for the paradox (Lopes, 1995). The paradox was mostly dismissed by economists, with Savage later revising his response to the thought experiment of Allais to align with expected value theory (Moscati, 2018).

Consistent with Mosteller and Nogee's (1951) notion of different psychological and mathematical probabilities, Allais proposed subjective probability distortions dependent on the framing of prospects as gains or losses, and on their magnitude. He also argued for a more complex utility function to describe risky choice through incorporating the variance and other properties of the utility function with its expected value (Moscati, 2018). His common consequence and common ratio paradoxes are classic paradoxes in the decision-making literature. Other paradoxes include risk aversion and risk seeking in the same person, such as taking out insurance and gambling, the St Petersburg Paradox, the reflection hypothesis and preference reversals. Such paradoxes are also referred to as critical properties (Birnbaum, 2008, 2011).

In the main, deviations from expected utility theory were accommodated or dismissed, and von Neumann and Morgenstern's expected utility theory became the mainstream economic model of risky choice (Lopes, 1995; Moscati, 2018). Despite abundant evidence of its descriptive inadequacies, expected utility theory is still considered the dominant theory of decision making in neoclassical economics, and the basis of neoclassical economic applications (Cokely et al., 2016). It is applicable to any type of outcome, not money in particular. The theory is recognised for its simplicity and the elegance of using the utility function for representing risky choice (Schoemaker, 1982).

2.1.5.6. Later challenges to expected utility theory: Psychologists and behavioural economists

In contrast to the passive response of many economists to the inadequacies of expected utility theory, several psychologists actively challenged the theory. They developed two research programmes, both of which investigated people's responses to complex, unclear information. The one programme emphasised overt personal decisions or preferences, and the start of psychological research on *decision theory*. The other emphasised the psychological judgment of complex information, and the start of psychological research on *judgment*. Both programmes produced cognitive models of decision making, although based on different paradigms (Goldstein & Hogarth, 1997). Only the former programme is discussed here as the thesis is limited to decision making.

The psychologists concerned with decision theory had interests in modelling the relation between psychological phenomena and physical stimuli. They were inspired by Edwards (1954) who published a detailed review of economic and statistical research in the decision-making field. They also objected to modelling risk attitudes based on the utility of money (Wakker, 2010). Following demonstrations such as the Allais paradox, they began to test models of preferential choice derived from economics using choices between lotteries with different outcomes and probabilities. Their tests of expected utility theory showed systematic violations and procedural invariance (Kahneman & Tversky, 1979; Schoemaker, 1982; Tversky, 1975). Their work formed the beginnings of the field of behavioral economics in which psychological principles were incorporated into economic theory (Glimcher et al., 2009). In the same vein, although some years later, Rabin (2000) presented a calibration theorem which challenged the ability of expected utility theory to account for risk aversion over large stakes (or outcomes) unless risk attitudes are approximately neutral for stakes of moderate size. This theorem, known as Rabin’s paradox, represented a meaningful challenge to the theory (Wakker, 2010).

In general, these researchers sought non-expected utility models of risky choice based on alternative theories that represented psychological processes, modelling methods, and experimental testing (Glimcher et al., 2009; Lopes, 1995; Moscati, 2018). In contrast with the formula for the normative expected utility model, the formulae for the non-expected utility models involved nonlinear transformations of the probability functions with a psychological reference (Glöckner & Pachur, 2012).

The general formula $W(A)$ that accounts for nonlinear probability weightings is provided in equation 4, assuming an *n*-outcome risky prospect $A = (p_1x_1; \ldots; p_nx_n)$, where $w(p_i)$ denotes the outcome of a nonlinear weighting function of the probability of obtaining the outcome x_i.

$$W(A) = \sum_{i=1}^{n} w(p_i)x_i \quad (4)$$

An example of a nonlinear probability curve is shown in Figure 2, sourced from Wakker (2010, p. 152).

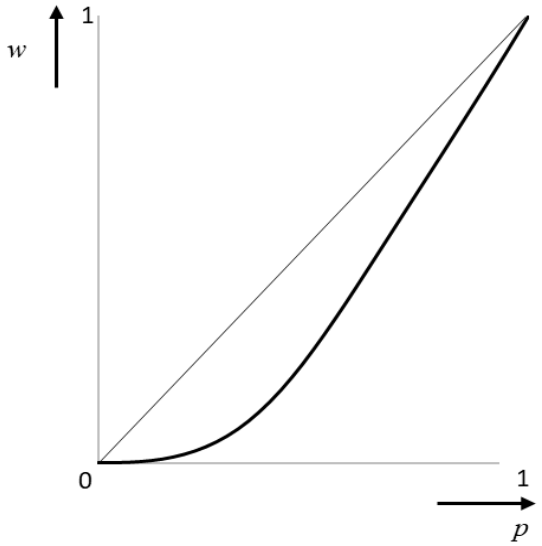


Figure 2: A nonlinear probability weighting function
Note: Reproduced from Wakker (2010, p. 152)

2.1.6. Early non-expected value models of risky choice: Preston and Baratta (1948)

The psychologists Preston and Baratta (1948) are thought to have been the first researchers to use nonlinear functions of probability in research on risky choice predictions (Gonzalez & Wu, 1999; Wakker, 2010). They based their nonlinear transformation of mathematical probabilities on a psychological scale in an experiment which required groups of participants to use faux money to bid competitively for bets involving different probabilities and payoffs. They found that players overvalued bets involving low probabilities of high winnings, and undervalued high probabilities of low winnings (Edwards, 1953; Preston & Baratta, 1948). As they found no evidence of a relation between knowledge and experience of the theory of probability and the psychological perceptions of probability, they attributed behavioural deviations from expected value theory to subjective probabilities, introspection and intuition.

In contrast to researchers of expected utility theory, who assumed a nonlinear utility function and a linear probability function, they explained their findings in terms of linear utility for the perception of outcomes, and nonlinearity for the perceptions of probabilities.

Other researchers who obtained similar findings on qualitative nonlinear probability weightings include Mosteller and Nogee (1951), Allais (1952), and Edwards (1954), all cited by Gonzalez and Wu (1999, p. 133).

2.1.7. Original prospect theory (Kahneman & Tversky, 1979)

In a somewhat similar fashion to Daniel Bernoulli's abandonment of objectivity in risky choice around 1738, Tversky and Kahneman abandoned the idea of rational choice (Wakker, 2010). They spent five years observing how people make risky choices, before arriving at a theory based on a utility (value) function and a probability weighting function that accommodated the Allais paradox and other behavioural choice anomalies in conflict with expected utility theory (Gonzalez & Wu, 1999).

Tversky and Kahneman based their thinking on Fechner's 19th century psychophysical law of a logarithmic relation between the magnitude of physical properties, such as brightness or pitch, and the psychological intensity experienced, conceptualised in Bernoulli's work on psychological utility some 100 years prior. As shown in Figure 3, the weighting function that they identified overweighted small probabilities, and underweighted large probabilities. It was not defined for extreme probabilities, a characteristic they termed *subcertainty* (Kahneman & Tversky, 1979, p. 283).

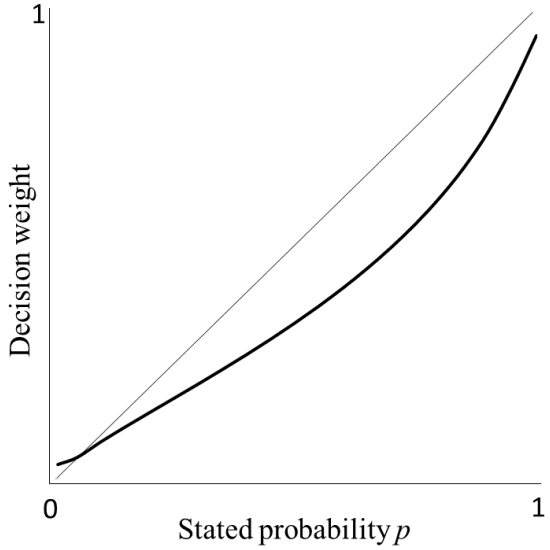


Figure 3: Probability weighting function identified by Kahneman and Tversky (1979)
Note: Reproduced from Kahneman and Tversky (1979, p. 283)

The result was their 1979 version of prospect theory, referred to in this thesis as *original prospect theory*. This psychophysical theory of choice provided the first descriptive model of choice behaviour that explained normative deviations using plausible behavioural foundations tested on overt choices (Wakker, 2010). However, in this early version of their theory, nonlinear decision weights were applied directly to multiple outcomes. This resulted in anomalous predictions of prospects, such as violations of stochastic dominance. [Stochastic dominance (of the first order) requires that shifting the probability from a lower outcome to a higher outcome should result in a preferred gamble.] Despite a 2-stage editing process proposed by the authors to overcome the problem, original prospect theory required a solution to the violation of stochastic dominance to be acceptable as a model of risky choice (Kahneman & Tversky, 1979). (As Kahneman and Tversky used the term 'value' in place of 'utility' in their models of risky choice, these terms are used interchangeably in this thesis from here on.)

2.1.8. Rank dependence (Quiggin, 1981 and others)

Several scholars worked independently on rank-dependent choices under risk and uncertainty and the stochastic dominance anomaly of original prospect theory. The economics honours student, John Quiggin, was the first such scholar to overcome the problem of violation of stochastic dominance under risk. His rank-dependent algebraic model of risky choice was based on a nonlinear probability function in which probabilities are ranked and transformed cumulatively (Quiggin, 1981).

The incorporation of rank dependence in algebraic models of choice using probabilistic sensitivity reflected the incorporation of psychological insight and behavioural economics within mainstream economics (Wakker, 2010). Diecidue and Wakker (2001) describe the intuitive arguments proposed in the literature on choice that provide psychological and economic support for rank dependence. Psychological applications include personality impressions formed on ordered personality traits (Birnbaum, 1974), and intuitive risks based on the relative weights of outcomes linked to inequality (Lopes, 1995). An economic application that focuses on extreme outcomes is Rawls' (1971) model of welfare evaluation based on the poorest individual.

2.1.8.1. Properties of rank-dependent utility theory (Quiggin, 1981)

Unlike under expected utility theory, risk attitudes under rank-dependent utility theory are determined jointly by a nonlinear probability weighting function as well as by the utility

function. Additionally, the weighting function applies to *cumulative* probabilities rather than to individual probabilities (Tversky & Kahneman, 1992). Thus, the magnitude of a transformed probability value or decision weight depends on its position in the distribution, which mostly affects the probabilities of the worst and best outcomes (Lopes, 1995).

Rank-dependent utility theory is a generalisation of expected utility theory. Stated differently, expected utility theory is a special case of rank-dependent utility when the weighting function is linear. As rank-dependent utility does not rely on the independence axiom of unrelated outcomes, it accounts for well-known violations of expected utility theory, including the Allais paradox. It also displays transitivity and stochastic dominance (Quiggin, 1991). The theory thus offered a solution to the stochastic dominance violation of original prospect theory.

An example of the ranking process of rank-dependent utility theory is provided in Appendix A, sourced from Wakker (2010, pp. 165-168).

2.1.8.2. Reactions to rank-dependent utility theory

Rank-dependent utility theory is said to be a satisfactory alternative to expected utility theory. In a recent review of behavioural utility models of risky choice, Harrison and colleagues identified Quiggin's model as superior to the expected utility model for explaining psychological biases in risky choice, and a serious contender to the cumulative prospect model that succeeded it (Harrison, 2019; Harrison & Swarthout, 2016). Wakker (2010) describes the principle of ranking as mathematically natural, and consistent with psychological decision-making processes. Accordingly, the theoretical parameters of the model have plausible psychological interpretations, and the model is viewed as *homeomorphic*. Homeomorphic models are preferred to *paramorphic* models in which the model parameters are not supported by plausible psychological interpretations (Harré, 1970, as cited in Wakker, 2010, p.3).

However, rank-dependent utility theory cannot explain certain cognitive biases, for example, loss aversion, the reflection effect and overweighting of low probabilities. An example of loss aversion is asymmetric price elasticities whereby purchases of consumer goods are more sensitive to price increases than to cuts; an example of the reflection effect is the disposition of holding losing stocks too long and selling winners too early; and an example of overweighting of low probabilities is consumers buying overpriced insurance (Camerer,

2000). The explanation of these phenomena is the preserve of the 1992 modification of original prospect theory, known as cumulative prospect theory (Dhami, 2016).

2.1.9. Cumulative prospect theory (Tversky & Kahneman, 1992)

The principle of rank-dependent probabilities informed Kahneman and Tversky's 1992 version of prospect theory, which they named cumulative prospect theory. Like rank-dependent utility theory, cumulative prospect theory incorporates the psychological principles of risk attitudes that are sensitive to both probabilities and outcomes.

There are two main differences between cumulative prospect theory and rank-dependent utility theory. The first is related to the distinction between losses and gains. Cumulative prospect theory evaluates prospects as losses or gains *relative to a reference point* rather than to the outcome, or to final wealth (see Section 2.1.5.2). Second, whereas rank-dependent utility theory comprises probabilistic sensitivity and utility curvature, cumulative prospect theory adds the phenomenon of *loss aversion* based on reference dependence (Wakker, 2010).

2.1.9.1. The reference point

In cumulative prospect theory, the outcome of a gamble is perceived as a loss or a gain in relation to a reference outcome or reference point. Consequently, preferences are sign dependent, i.e. positive or negative, relative to this reference point. Tversky and Kahneman provided various interpretations of the reference point, describing it as the point: "…that is judged neutral" (Tversky & Kahneman, 1981, p. 456); "… suggested or implied by the statement of the problem" (Kahneman & Tversky, 1984, p. 346); or "characterized by the abrupt changes in the valuation of gains and losses and of acceptable or reprehensible behavior [*sic*]" (Kahneman, 1992, p. 296). Alternatively, the reference point may be considered as the status quo.

The reference point implies the following: First, separate value (utility) functions, and separate weightings functions, are used for gains and losses. The term *value* function instead of *utility* function is used to emphasise outcomes as evaluated relative to the reference point. Second, wealth is conceptualised as a *change* relative to the reference point, and thus described as *local or net wealth*. By contrast, wealth under expected utility theory and rank-dependent utility theory is conceptualised as *final or integrated wealth*, which includes an initial, or house, endowment (Wakker, 2010).

The vagueness of the reference point is contentious for cumulative prospect theory, as it is the basis for categorising prospects as gains or losses, and for loss aversion. Harrison and Ross (2017, p. 162) criticise the reference point as "subtle and difficult, and … deliberately left outside the model", necessitating it having to be assumed or estimated in empirical research on cumulative prospect theory. Kőszegi and Rabin (2007) review evidence of the inconsistent results that emanate from cumulative prospect theory researchers' attempts to specify the reference point. However, others defend the concept, despite its vagueness. For example, Dhami (2016) cites the salience of the reference point as the *status quo* in many problems and contexts, and commends the way the concept has been developed in Kőszegi and Rabin's (2007) research framework. Wakker (2010) takes the pragmatic view of accepting the volatility of the reference point in different problem contexts.

2.1.9.2. Loss aversion

Loss aversion is the empirically-based phenomenon whereby people are more sensitive to losses than to gains (Wakker, 2010), i.e. "losses loom larger than gains" (Kahneman & Tversky, 1979, p. 279). Kahneman and Tversky further describe loss aversion as the aggravation experienced in the loss of a sum of money outweighing the pleasure of gaining that same amount, also implying that the avoidance of losses is greater than the attraction of gains. The concept is thus dependent on the reference point to determine losses and gains.

The loss aversion parameter λ is a function of the shape of the utility curve before and after the reference point. For values greater than 1, losses are overweighted relative to gains. Tversky and Kahneman (1992) found a λ value of 2.25. Loss aversion values less than 1 are found less often, but when found, suggest "gain seeking with little attention for losses" (Wakker, 2010, p. 239). Examples of value functions that illustrate loss aversion to a lesser and greater extent, are provided in the left and right panes of Figure 5, respectively, sourced from Wakker (2010, p. 240).

There are several caveats related to the phenomenon of loss aversion. First, although loss aversion is usually interpreted in terms of differences in *utility* for losses versus gains (utility loss aversion), an alternative interpretation is in terms of differences in *probability weightings* for losses versus gains, i.e. in decision weights (probabilistic loss aversion). These two interpretations imply different psychological pathways (Gonzalez & Wu, 1999; Harrison & Ross, 2017). For example, in terms of utility, $\lambda = 2$ is interpreted as the pain of a loss felt twice the extent of the pleasure of an equivalent gain; in terms of probability weighting, the

extent of the pain and pleasure would be the same, but the individual's distortion of the probability of a loss is twice her distortion of the probability of the gain. Wakker's (2010) interpretation is that the *importance* of the loss would be twice that of the gain.

Second, loss aversion is volatile and may be an experimental artefact. Wakker (2010) cites empirical evidence of the volatility of loss aversion, and its dependence on experimental manipulation. As loss aversion is based on the reference point, when subjects understand the difference between an overall loss (total wealth) and a loss relative to a reference point (local gain or local loss), loss aversion may disappear. Third, the theory assumes the same loss aversion value for small and large unit changes (in either value or probability) (Harrison, 2019). Finally, the value function proposed by Kahneman and Tversky (1979) is not defined near the extremes of 0 and 1.

2.1.9.3. The cumulative prospect theory model (Tversky and Kahneman, 1992)

As proposed by Tversky and Kahneman (1992), the cumulative prospect theory model comprises two functions: *the utility function* termed the *value* function U and the *probability weighting function* w_i .

As given in equation 5, the model assigns an overall value, *V(A)*, to an *n*-outcome risky prospect *A* with probability p_i (*i* = 1, . . . , *n*) where p_i is the chance of receiving outcome x_i. The prospect has outcomes ordered as follows: $x_1 \leq \ldots \leq x_k \leq 0 \leq x_{k+1} \leq \ldots \leq x_n$.

U is a continuous, strictly increasing value function. π_i^- and π_j^+ are decision weights for losses and gains, as in equations 6a and 6b, respectively (Booij, van Praag, & van de Kuilen, 2010).

$$V(A) = \sum_{i=1}^{k} \pi_i^- \mathrm{U}(x_i) + \sum_{j=\mathrm{k}+1}^{n} \pi_j^+ \mathrm{U}(x_j) \tag{5}$$

$$\pi_i^- = w^-(p_i) \tag{6a}$$

$$\pi_i^- = w^-(p_1 + \cdots + p_i) - w^-(p_1 + \cdots + p_{i-1}) \quad for\ 1 < i \leq k$$

$$\pi_j^+ = w^+(p_j) \tag{6b}$$

$$\pi_j^+ = w^+(p_j + \cdots + p_n) - w^+(p_{\mathrm{j}+1} + \cdots + p_n) \quad for\ k < j < n.$$

The value function

In cumulative prospect theory, the value function is *S*-shaped, concave for gains and convex for losses. As shown in Figure 4, the function has an inflection point separating gains from losses. Its gradient is steeper for losses than for gains indicating loss aversion with losses weighing more heavily than gains. The flattening of the curve towards high and low extreme values indicates preferences with diminishing sensitivity to values further away from the reference point (Kahneman & Tversky, 1979).

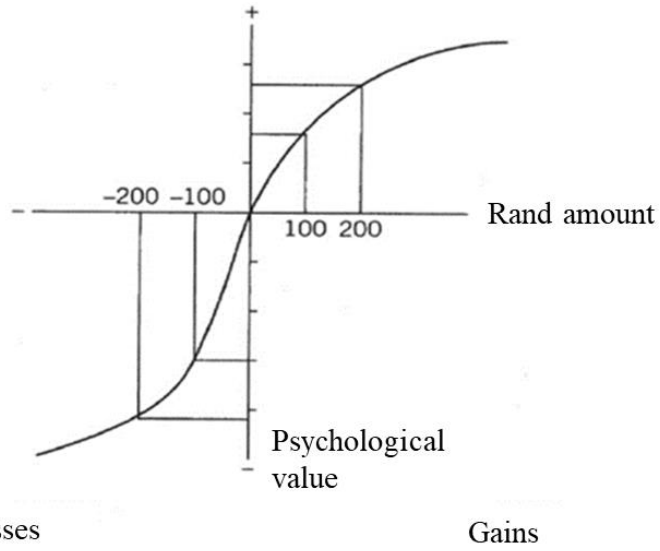


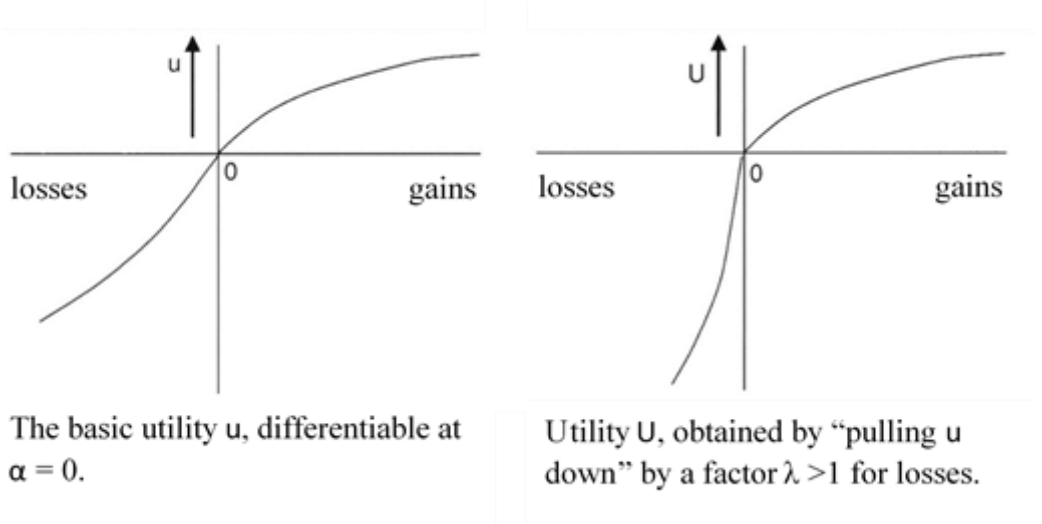
Figure 4: A hypothetical value function in cumulative prospect theory
Note: Reproduced from Kahneman (2011, p. 283)

Throughout this thesis, the power value function is used, as introduced by Tversky and Kahneman (1992) for cumulative probability theory. It is given by equation 7.

Loss aversion is modelled by the parameter λ

$$\begin{aligned} u(x) &= x^{\alpha}, & x \geq 0 \\ u(x) &= -\lambda|-x^{\alpha}| & x < 0 \end{aligned} \tag{7}$$

The phenomenon of loss aversion is illustrated more clearly in Figure 5, reproduced from Wakker (2010, p. 240).



The basic utility u, differentiable at $\alpha = 0$.

Utility U, obtained by "pulling u down" by a factor $\lambda > 1$ for losses.

Figure 5: Hypothetical value functions illustrating utility functions with degrees of loss aversion

Note: Reproduced from Wakker (2010, p. 240)

The figure in the left-hand pane of Figure 5 illustrates minor risk aversion via a value function (u) with a slightly steeper gradient for losses than for gains; the figure in the right-hand pane illustrates substantial risk aversion via a value function (U) as the slope of the curve in the loss frame is considerably steeper than for gains.

Figure 6 illustrates the value function with parameter values based on the median values of Tversky and Kahneman (1992), i.e. $\theta = \theta' = 0.88$, $\lambda = 2.25$. The shape of the curve approximates the theoretical *S*-shape, and the steeper slope for losses relative to gains illustrates loss aversion.

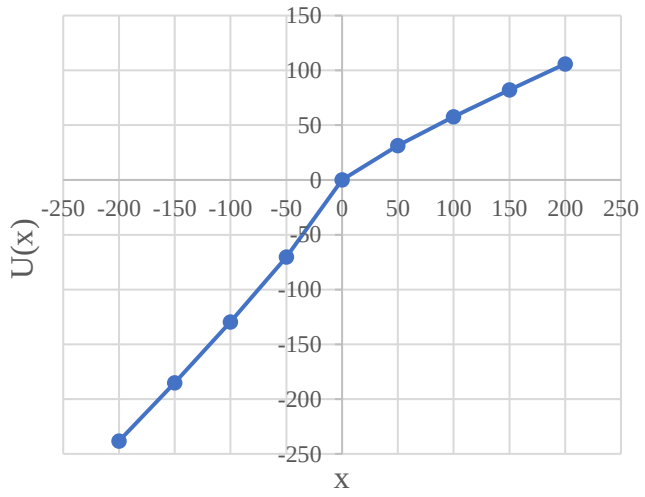


Figure 6: The power function applied to the value function, with median parameter values of Tversky & Kahneman (1992)

The probability weighting function (Tversky & Kahneman, 1992)

Tversky and Kahneman (1992) used the inverse *S*-shaped function $w(p) = p^{\gamma}/(p^{\gamma} + (1\text{-}p)^{\gamma})^{1/\gamma}$ to represent the risk attitudes of cumulative prospect theory for both negative prospects and positive prospects (Figure 7). The function demonstrates that low probabilities are overweighted, and moderate to large probabilities are underweighted. The slope of the weighting function in any particular range of probabilities represents the sensitivity of decision makers' preferences to changes in probabilities in that range. The curve thus indicates diminishing sensitivity to probability difference in the middle of the range (Tversky & Kahneman, 1992).

Tversky and Kahneman (1992) found a median parameter values for gamma = 0.61. The single parameter determines both the curvature and the point at which the curve crosses the neutral line, and thus the curvature and crossover are confounded (Wakker, 2010; Xing et al., 2019).

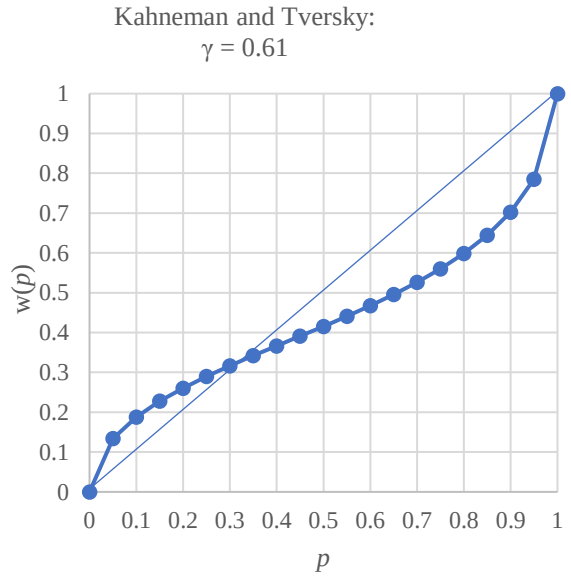


Figure 7: Illustration of weighting function (Tversky & Kahneman, 1992)

Thus far, the presentation of cumulative prospect theory model has been based on the value and probability parametric functional forms used by Tversky and Kahneman (1992). Other parametric functional forms are discussed in Section 2.2.

2.1.9.4. Reactions to cumulative prospect theory

Cumulative prospect theory has received substantial acclaim. It was recognised in the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel awarded to

Daniel Kahneman in 2002, by which time Amos Tversky had died (Nobelprize.org, 2002). Such has been the influence of Tversky and Kahneman's 1992 study, that their published parameter values for risk aversion, probability weightings and loss aversion have been used extensively in the risky choice literature. Furthermore, many researchers assume that these values are representative of the risk attitudes of the population of decision makers. As a result, loss aversion has been accepted by many as a generalised scientific principle with Tversky and Kahneman's 1992 value assumed largely impervious to change. By example, Thaler and Sunstein (2008) used the loss aversion parameter value to justify their *nudge* approach.

However, the practice of using Tversky and Kahneman's 1992 estimates as the normative standard has been severely challenged (Ert & Erev, 2013; Gal & Rucker, 2018; Simonson & Kivetz, 2018; Wakker, 2010; Yechiam & Hochman, 2013; Yechiam, 2019). These researchers query the reproducibility of studies that claim evidence of the phenomenon. As reviewed by Yechiam and Hochman (2013), some researchers have failed to find evidence of loss aversion, even when using the method of measuring loss aversion described by Kahneman Tversky in 1979. Another concern is that loss aversion may hold only for very large losses and not for small or moderate losses (Gal & Rucker, 2018; Yechiam, 2019).

Tversky and Kahneman claim that their theory captures additional patterns of risky behaviour. For example, they present empirical evidence that it yields a "fourfold pattern" of risk attitudes or biases (Kahneman, 2011, p. 317; Wakker, 2010). The four biases comprise two risk averse behaviours, and two risk-seeking behaviours. The two risk averse behaviours occur, first, with high probability gains, as reflected by the risk attitudes of perceiving diminishing returns of money; and second, with low probability losses, reflected by the taking out of insurance. The two risk-seeking behaviours occur, first, with low probability gains, reflected by buying lottery tickets; and, second, with high probability losses, reflected by taking desperate gambles (Kahneman, 2011). However, a contradictory claim to the success of the model predictions is that these biases may be due to systematic distortion in decision makers' risky choices caused by random errors (Bhatia & Loomes, 2017; Blavatskyy, 2007; Wilcox, 2008). When choices are analysed using a decision model that incorporates control of response error, the fourfold pattern disappears (Blavatskyy, 2007). (A more thorough review of the topic of stochastic error is presented in Section 2.3.)

Additionally, the method that Tversky and Kahneman used in their 1992 study has been described as both methodologically and analytically weak (Harrison & Ross, 2017).

Criticisms include the reporting of only the median value of point estimates for each the model parameters ($\alpha = \beta = 0.88$, $\lambda = 2.25$, $\gamma+ = 0.61$, $\gamma- = 0.69$). Tversky and Kahneman calculated these values at the individual level for their 25 subjects without reporting standard errors or any other measures of variability between- or within- person parameter variability. Thus, these median measures may not have reflected the range of values across their subjects. They also did not explain whether they assumed the values of the risk aversion parameters for gains and losses to be equal, or whether they calculated them as such. Despite caveats in the cumulative prospect modelling literature to hold these values equal (Nilsson, Rieskamp, & Wagenmakers, 2011a; Wakker, 2010), some researchers do not do so, for example, Rieger, Wang, and Hens (2017) and L'Haridon and Vieider (2019a).

Harrison and colleagues (Harrison & Ross, 2017; Harrison & Swarthout, 2016) have also reviewed the methodologies of other researchers who have claimed support for cumulative prospect theory based on controlled experiments and parametric modelling. They cite 29 such studies published between January 1992 and March 2017 (excluding their own). They challenge the methodologies of these studies, citing the lack of salience of the incentives offered as rewards, failure to include loss and mixed prospects (prospects involving both gains and losses) resulting in flawed estimates of loss aversion, and questionable use of probability weighting functions. Only 5 of the 29 studies reviewed included salient rewards in the gain, loss and mixed frames (Rieskamp, 2008; Nilsson et al., 2011; Glöckner & Pachur, 2012; Zeisberger, Vrecko, & Langer, 2012; Murphy & ten Brincke, 2017), all cited by Harrison (2019, pp. 122-124).

The practice of conducting experiments that claim to test cumulative prospect theory and estimate its parameters without incorporating mixed prospects implies testing rank-dependent utility theory instead, as the rank-dependent utility theory does not involve loss aversion (Harrison & Ross, 2017; Wakker, 2010). The study of Patalano et al. (2015) is such an example. Although the researchers claimed to be testing the role of numeracy in predicting distortions in value and probability using cumulative prospect theory, they used only prospects in the gain frame. This study is earmarked for later discussion as its involvement with numeracy is particularly relevant to the present thesis.

2.1.10. Conclusion

The review has outlined the underlying theoretical or *structural* component of risky choice models, focusing on the three models considered in this thesis. Cumulative prospect theory is

still considered by many researchers as prominent in meeting the requirements of a descriptive alternative to normative risky choice theory (Fox, Erner, & Walters, 2015; Keren & Wu, 2015; Pachur, Suter, & Hertwig, 2017c). However, laboratory studies that have attempted to test the theory, and claim support for it, are beset with methodological challenges as outlined in the previous section.

The review next provides a more detailed account of the *parametric functions* that these structural models require to satisfy their qualitative theoretical properties of risk aversion/seeking and probability distortions.

2.2. Model specification (ii): Functional forms for value and probability functions

As previously discussed, non-expected utility theories, such as rank-dependent utility and cumulative prospect theory are based on the qualitative theoretical principles of risk aversion and risk seeking and, in the case of cumulative prospect theory, loss aversion. These principles cannot be measured directly by asking individuals to assess them subjectively. Thus, the structural models that apply these theories use functional forms, or functions, that satisfy such principles. These functional forms need to be assumed and specified for weighting the value and probability functions when using a parametric approach for estimating the model parameters, as in this thesis. Value weighting functions are used to relate outcome values to their subjective values, and probability weighting functions are used to relate objective probabilities to their subjective weights. The parameters of these functions are estimated indirectly through observed risky choices (Cavagnaro, Pitt, Gonzalez, & Myung, 2013).

The most commonly used functional forms are presented for the value function in Section 2.2.1 and for the probability function in Section 2.2.2, respectively. (From here on, the terms ‘function’, ‘value curve’, and ‘curve’ are used synonymously.)

2.2.1. Parametric functional forms for the value function

The decision-making literature on value refers to three popular parametric functions: linear, logarithmic and power. Other forms include quadratic, exponential, Bell, and Hara (Stott, 2006), but these are not discussed here.

The linear function is plotted at 45° and represents risk neutrality as utility and expected values are equal.

2.2.1.1. The logarithmic value curve

The logarithmic curve is a negatively accelerated or concave curve. When used for a utility function, the curve reflects an attitude towards money consistent with risk aversion. As previously discussed, Bernoulli used the logarithmic function in the 18th century to illustrate the notion that utility would diminish incrementally in the same way as increments in wealth diminish with increases in existing wealth. Fechner also used the function in his 19th century psychophysical law to convey the principle of Weber's law. Cognitive psychologists have used the logarithmic function to model distortions in the perception of numbers of less educated individuals, as well as those of children. By contrast, more educated adults, and those with higher numeracy test scores, tend to have a more linear representation of numbers, and thus less perceptual distortion (Booth & Siegler, 2006, as cited in Reyna & Brust-Renck, 2014, p.217). Tversky and Kahneman (1974) used the logarithmic function in their psychophysical model of perceptual biases of numbers (Heukelom, 2007).

2.2.1.2. The power function

The power function, associated with Stanley Smith Stevens in 1957, replaced the logarithmic function in decision modelling. In economics, power curves are also referred to as the family of constant relative risk aversion (CRRA). The power family of utility curves (U) indicates strictly increasing functions and is frequently used to model risk aversion (Wakker, 2008).

Although the power function often provides a better fit to data than other functions, it presents problems for analysing loss aversion in the case of cumulative prospect theory (Wakker, 2008; 2010). The problem derives from the extreme behaviour of the power function near 0, where the differences in the slopes of the utility curves are necessary for estimating the loss aversion parameter (Wakker, 2010). Unless power parameters for gains and losses agree, loss aversion is dependent on the unit of money used. If a prospect is expressed in Rands (or Dollars) and then re-expressed in cents (all else held equal), the loss aversion parameter λ will have entirely different values in the two situations.

As previously illustrated (Figure 6), Tversky and Kahneman (1992) applied a power value function to derive predictions from their cumulative prospect theory model. They used different power functions for losses and gains, and arrived at identical values for the positive and negative parameters for losses, without explanation. Other researchers, including Harrison and colleagues, and Nilsson et al. (2011a) recommend holding the parameter values

equal to avoid the problem of dependence on the monetary unit. However, some researchers fail to do so.

2.2.2. Parametric functional forms for the probability weighting function

Under expected utility theory, the linear function represents probability neutrality. This section is thus relevant only to non-expected utility theories.

A probability weighting function for risky decisions reflects the decision maker's mental representation or *distortion* of objective probabilities (Dhami, 2016). The power function is popular for modelling the probability component under rank-dependent utility theory. The inverse-*S* shaped functions as used by Kahneman and Tversky, the Goldstein-Einhorn function, and the Prelec functions are commonly used under cumulative prospect theory (Glöckner & Pachur, 2012; Stott, 2006).

2.2.2.1. The power probability weighting function under rank dependence

As explained by Wakker (2010), rank dependence may be combined with the probability weighting power function (p^r with curvature r) to explain the risk attitudes of pessimism and optimism. Figure 8 displays two examples of the probability weighting power function plotted against cumulative probabilities of ranked outcomes. In the left pane, the probability weighting function is convex based on the power function $w(p) = p^2$; in the right pane, the function is concave based on the function $w(p) = p^{1/2}$. Recall that the decision weight "is the marginal contribution, measured in *w*-units, of the outcome probability to the total probability at receiving better outcomes" (Wakker, 2010, p. 160). The worst outcomes have the highest ranks corresponding to the highest cumulative probabilities. The convex curve on the left of Figure 8 implies the risk attitude of pessimism, as the *worst* outcome, i.e. the outcome with the *highest* rank towards the right of the graph, is the most *overweighted* where the curve is steepest, and thus has the *highest* decision weight. The curve thus describes the risk attitude of a risk taker who places greater weight on the lower (most highly ranked) outcomes and less weight on the higher outcomes, i.e. a pessimist. By contrast, the concave curve on the right implies optimism as the *worst* outcome, i.e. the outcome with the *highest* rank towards the right of the graph, is the most *underweighted* where the curve is shallowest, and thus has the *lowest* decision weight. The curve thus describes the risk attitude of a risk taker who places greater weight on the higher outcomes and less weight on the lower outcomes, i.e. an optimist (Wakker, 2010).

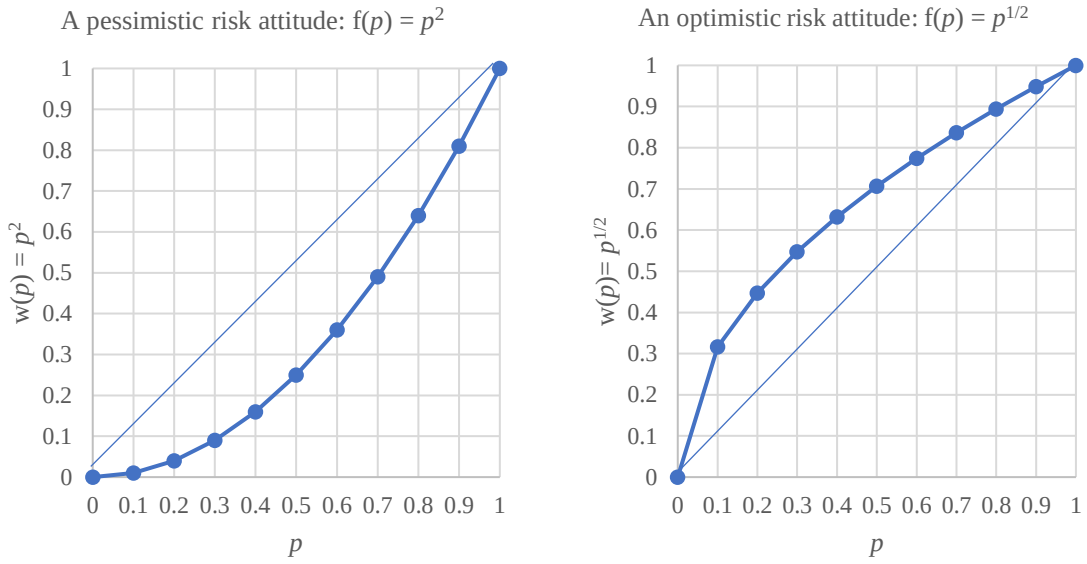


Figure 8: Convex curve showing pessimism (left pane), and concave curve showing optimism (right pane)

A worked example of the calculation of rank-dependent utility is provided in Appendix A.

As Wakker (2010) explains, under rank-dependent utility, these probabilistic risk attitudes impact utility: just as *concave value* curves indicate risk aversion, *convex probability* weighting curves represent pessimism and consequently risk aversion, as well. Likewise, just as *convex value* curves enhance risk seeking, *concave probability* weighting curves represent optimism and consequently risk seeking too. Thus, it is important to consider the parameters of both the value and probability weighting curves jointly, rather than individually, when evaluating the risk aversion or risk seeking of an individual. A risk averse person may have a probability weighting curve that reflects pessimism and that outweighs her seemingly contradictory convex value function reflecting risk seeking (and *vice versa*).

2.2.2.2. The 2-part power function under rank dependence: The inverse *S*-shaped probability weighting function

In empirical research on non-expected utility models under rank dependence, the most commonly observed probability weighting function is the inverse *S*-shaped curve (Wakker, 2010; Wu et al., 2004), as shown in Figure 9. Under rank dependence, the curve may be used to explain the classical paradox in the economics literature of the same individual taking out insurance and gambling. The *pessimism* of buying insurance for the certainty of avoiding a catastrophe with a small probability appears paradoxical to the *optimism* of buying lottery

tickets with an extremely small possibility of winning a large prize. Both events are associated with extremely low probabilities.

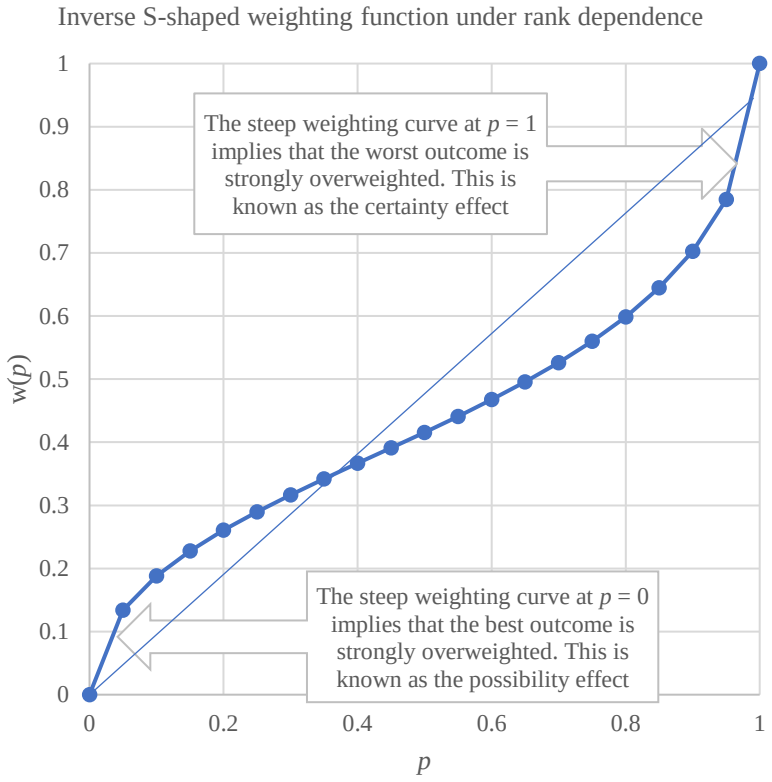


Figure 9: A hypothetical inverse *S*-shaped weighting function

The paradox of the overweighting of extreme probabilities is resolved by the inverse *S*-shaped weighting curve for outcomes ranked at the two extremes. At $p = 1$ the curve is steep, and the worst outcome is overweighted, implying *certainty* for the decision maker that she will avoid the *worst* outcome i.e. a catastrophe (thus suggesting pessimism and risk aversion, hence insurance). At $p = 0$ where the curve is similarly steep, the best *possible* outcome is overweighted, implying the overweighted *possibility* that the decision maker will win the lottery, in spite of the low odds of this occurring (thus suggesting optimism and risk seeking, hence buying a lottery ticket) (Wakker, 2010).

The region where the curve is flattest (in the middle range of the X axis), implies *likelihood insensitivity* for probabilities. Attitudes to outcomes that are ranked in the middle, i.e. those that are neither the best nor the worst, mostly reflect pessimism (Figure 9). However, attitudes to these outcomes contribute less to the overall evaluation of a prospect as the associated decision weights are relatively small.

The inverse *S*-shaped curve thus explains why non-expected utility theories are less dependent on the principal characteristic of expected utility theory that most rational people are generally risk averse. Under rank dependence, a person's behaviour may be generally risk averse, except when confronted with the possibility of a large outcome with extremely low odds of occurrence, in which case her behaviour changes to risk seeking (Quiggin, 1991).

There are various parametric forms of the inverse-*S* function. All involve some form of power in their formulae. Presented here is the function used by Tversky and Kahneman (1992), and the function proposed by Lattimore, Baker, and Witte (1992).

The single-parameter probability weighting function: Tversky and Kahneman (1992)

As previously illustrated (Figure 7), Tversky and Kahneman (1992) used the single parameter inverse *S*-shaped function $w(p) = p^{\gamma}/(p^{\gamma} + (1\text{-}p)^{\gamma})^{1/\gamma}$ to represent the risk attitudes of cumulative prospect theory. As previously described, the inverse *S*-function reflects optimism or overweighting for small probabilities, and pessimism or underweighting for moderate to large probabilities. The psychological interpretation of the weighting function parameter that Tversky and Kahneman used was based on the psychophysical principle of diminishing sensitivity. According to this principle, individuals are more sensitive to probabilities around the extreme values (in this case, probabilities of 0 and 1, an impossible event and a certain event, respectively), and become less sensitive as they move away from these points (Gonzalez & Wu, 1999; Kahneman & Tversky, 1984).

2.2.2.3. The inverse *S*-function: Lattimore, Baker, and Witte (1992)

Weighting functions with two parameters have superior properties to single-parameter functions when fitting individual choice data, as their parameters are independent of one another and more interpretable (Gonzalez & Wu, 1999; Stott, 2006). Such functions include those proposed by Goldstein and Einhorn (1987) and by Lattimore, Baker, and Witte (1992). The Prelec function is a third popular form, with single-parameter and 2-parameter versions. For brevity, only the function of Lattimore et al. (1992) is described in detail. It is henceforth referred to as the Lattimore function.

The Lattimore function *w* assumes that a linear transformation on the log odds scale can be used to vary the curve's elevation (intercept) and curvature (slope) separately (Gonzalez & Wu, 1999). The form of the Lattimore function is then:

$w(p) = \delta p^{\gamma}/(\delta p^{\gamma} + (1 - p)^{\gamma})$

where $\gamma > 0$ and $\delta > 0$.

The γ parameter measures *probability discrimination* via the degree of curvature of the function. Gonzalez and Wu (1999) provide the psychological interpretation of this parameter as *probability sensitivity.* The inverse *S*-shaped function has greater curvature around the probability boundaries of 0 and 1, which implies rapidly diminishing sensitivity (discrimination) to probabilities at these extremes, and less sensitivity to probabilities in the middle of the range where the function is relatively flat (Xing et al., 2019). Thus, the effect of changes to the probabilities of a gamble depend on where the probabilities are situated on the curve.

Gonzalez and Wu (1999) cite examples of two extreme situations of probability discrimination. The first extreme is when the γ parameter value is equal to 1, represented as a linear function of constant sensitivity to changes in probability throughout a probability interval, as when experts gamble in their specialised domain. The other extreme is when the γ parameter value approaches 0, represented as a step function, flat in the middle as there is no curvature. This function would imply a 'definitely-select/definitely-do-not-select' attitude, typical of a small child with no discrimination sensitivity between the probabilities of other outcomes.

The δ parameter measures the *elevation* of the weighting function. It measures the extent to which the curve is underweighted or overweighted relative to a 45° line. Gonzalez and Wu (1999) provide the psychological interpretation of elevation as the *attractiveness* of the gamble to the individual, with greater absolute elevation values indicating greater attractiveness of the gamble.

The two panes in Figure 10, reproduced from Fox and Poldrack (2009, p. 157), depict the Lattimore weighting function. The left-hand pane shows different elevation levels (**δ** values) with constant probability discrimination level $\gamma = .5$; the right-hand pane shows different probability discrimination levels (γ values) with constant elevation level **δ** =.5.

Greater elevation of the weighting function implies less overall risk aversion for gains, and more overall risk aversion for losses. Furthermore, these psychological properties are independent of one another which has the advantage that they are not confounded.

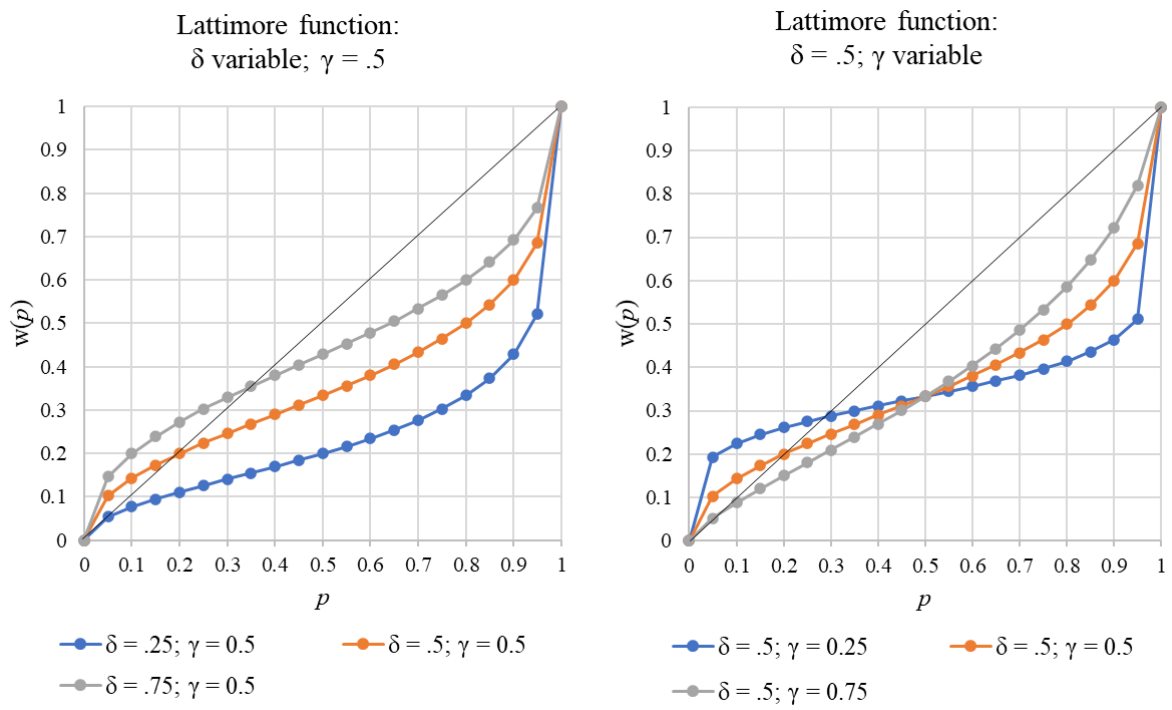


Figure 10: Lattimore et al. (1992) assuming γ = .5 (left pane); δ = .5, (right pane)
Note: Reproduced from Fox and Poldrack (2009, p. 157)

2.2.3. Conclusion on functional forms

Thus far, the literature review has focused on two aspects of risky choice modelling at the latent, or theoretical, level. These aspects are the structural forms of the models (expected utility, rank-dependent and cumulative prospect models), and the *functional* forms of the models for specifying value and probability weightings. Although models, such as rank-dependent utility and cumulative prospect theory address the descriptive inadequacies of expected utility theory, the latter continues to be accepted as the normative model of risky choice in economics. Furthermore, there is no clear indication of the best functional forms to use in modelling, although those characterised by separate parameters, such as that of Lattimore et al., (1992), have better explanatory power.

In the following section, the third latent level of the models is addressed. This concerns the random variability, or stochastic component, in the choices of decision makers. Consideration of this component overcomes the problem of the original deterministic nature of the model specification.

2.3. Model specification (iii): Probabilistic forms of error models of risky choice

Since the time of Bernoulli, the emphasis in the risky choice modelling literature has been on research involving structural models to predict the direction of a person's decision. The structural models considered in this thesis are represented by a mathematical equation, or model, based on the sum of the probability-weighted outcomes of the prospect, with their appropriate functional forms. In a binary discrete choice situation in which an individual chooses between competing risky prospects, the model is used to quantify the overall value of each of the prospects under consideration. According to the principle of maximisation, each time the decision maker is faced with a risky choice, she is expected to choose the prospect with the higher result as determined by the model. As the prediction of the direction of a person's risky choice is strictly determined, i.e. always the same for a choice situation, these binary discrete choice models are termed *deterministic*.

The assumption of randomness in choice behaviour is the "crucial missing element" (Wilcox, 2008, p. 199) in structural risky choice models. "Preoccupation with theories of the deterministic structure of choice under risk, to the exclusion of theories of error, is a mistake" (Wilcox, 2015, p. 1).

2.3.1. Empirical evidence of the probabilistic nature of preferential choice

The earliest empirical evidence of inconsistencies in risky choice experiments was noted by Mosteller and Nogee (1951) (see Section 2.1.5.4). Contrary to the deterministic thinking of the time, Mosteller and Nogee noted inconsistencies in their participants' choices of a risky bet, and moreover noted that the frequency of selecting the risky choice represented a "gradation" (p. 374) as its value increased. Participants' choices were thus *probabilistic*, as they were based on probabilities, rather than deterministic. Another frequently cited study of randomness in choice responses is that of Hey (2001), in which he presented participants with a particular set of 100 pairs of risky gambles over five experimental sessions. His participants displayed a high level of switching in their choices, even with the benefit of hindsight in the latter sessions. Such studies show risky choices to be probabilistic rather than deterministic.

In general, studies in which participants are faced with repetitions of identical binary choice problems show reversal rates of between 21% and 32% of initial choices, irrespective of whether the time interval between repetitions spans days or only minutes (Blavatskyy & Pogrebna, 2010; Wilcox, 2008).

2.3.2. Early stochastic modelling

In decision theory, observed choice inconsistencies are termed *stochastic* to indicate choices that are "less than perfectly predictable" (Wilcox, 2008, p. 198) and governed by probability theory (Singh, 2007). The term stochastic was used by Marschak in 1952 to suggest that a more 'stochastic' approach to choice would accommodate the probabilistic nature of the findings of Mosteller and Nogee (Moscati, 2018). Marschak went on to publish theory on stochastic choice in 1955, and in the early 1960s he collaborated with Becker and DeGroot to design stochastic models of choice behaviour that would account for inconsistencies in the experimental choices of their subjects' (Becker, DeGroot, & Marschak, 1963; Moscati, 2018). Other significant contributions to stochastic modelling include that of the mathematician Duncan Luce who devised a model of probabilistic choice in 1959, and McFadden who developed an economic version of Luce's choice model in 1976 (Anderson & Schulkin, 2014; Pleskac, 2015).

More recently, Wilcox proposed and defended the contextual stochastic error model (2008; 2011; 2015). Contextual utility essentially rescales the payoffs for each of the choice tasks to a unit 0 - 1 scale based on the minimum and maximum payoff in that choice task.

2.3.3. Sources of stochastic error

There are two main sources of error in risky choice modelling: response error and preference error (Bhatia & Loomes, 2017). Although some researchers have included stochastic error in the specification of their risky choice models, many have not included both types of stochastic error.

2.3.3.1. Response error

People do not respond perfectly accurately when they respond to risky choice situations due to random factors, such as lapses in concentration or attention, or change in mood. There is thus random error in their responses to risky choice situations, and their observed choices and underlying preferences may differ randomly. This type of stochastic error is known as *response noise* (Bhatia & Loomes, 2017).

As reviewed by Wilcox (2008) and Blavatskyy and Pogrebna (2010), there are several options for modelling stochastic response errors in binary discrete choices. These include the Tremble model, the Random Preference model, the strong utility or Fechnerian model, the strict utility model sometimes referred to as the Luce model, the moderate utility model,

Busemeyer and Townsend's (1993) decision field theory, and the contextual utility model (Wilcox, 2008, 2010, 2011, 2015).

In a comparison of models of response error, Wilcox (2008, 2010) found that only contextually-based models that consider the ratio of the utilities of the outcomes, satisfy the requirements of all the decision-making functionals. He concluded that predictions of contextually-dependent models are superior to the strong utility models that are conventionally used in behavioural economics and applied econometric modelling. Compared to using alternative stochastic specifications in econometric modelling of random parameters, contextual utility provided the best fit and forecasts (Wilcox, 2010). The stochastic expected utility theory of Blavatskyy (2007) is another such compatible model.

The probabilistic choice model developed by Luce in 1959 is often used to accommodate response noise in risky choice experiments (Bhatia & Loomes, 2017), for example, in the studies of Glöckner and Pachur (2012), and Nilsson et al. (2011a). This model is also referred to as the strict utility model, and can be rewritten as a Fechner model of homoscedastic random errors. Despite the popularity of this approach in the literature, the use of this formula is problematic when used with a decision model that has a utility function determined up to an affine transformation, as in the case of the decision models considered in this thesis. "Formally, there is a theoretical mismatch inherent in combining affine structures (such as EU or RDEU) with strict utility" (Wilcox, 2008, p. 218). Doing so would result in an arbitrary normalisation of the utility function which would affect the parameter estimates (Blavatskyy & Pogrebna, 2010). Thus, the Luce model is not considered in this thesis. Wilcox's contextual model is used instead.

2.3.3.2. Preference error

A second source of error in risky choice modelling resides in random variability from trial-to-trial of the parameters of the functions used by the structural model. Thus, these parameters are considered as random variables. Under expected utility theory, random error manifests as variability in the parameter that determines the shape of the utility function; under rank-dependent utility theory, it manifests as variability in the parameters of both the utility and the probability functions; under cumulative prospect theory, it manifests as variability in the parameters of both the value and probability functions and in the loss aversion parameter. Bhatia and Loomes (2017) refer to this type of error as *preference noise*.

Determinants of this type of error are fluctuations in the decision maker's computation errors and thoughts (Bhatia & Loomes, 2017).

2.3.4. Potential effects of stochastic errors in modelling

Both sources of error variability need to be considered in the analysis of choice data to inform and test theory, particularly in light of the nonlinear nature of models of risky choice (Rouder, Morey, & Pratte, 2016). In their research on the cumulative prospect theory model fit and stability of the parameters over the course of a week, Glöckner and Pachur (2012) found improved model predictions when both response and parameter variations were taken into account in the model. Prominent researchers in the field (Bhatia & Loomes, 2017; Wilcox, 2015) stress that whenever there is a nonlinear relationship between preference and utility parameters, the *coexistence* of response and preference noise can systematically distort choice probabilities. Distortions occur irrespective of the functional forms used in the structural models and the error models used to specify response noise. Bhatia and Loomes (2017) cite several empirical examples where the omission of one of the error sources – either response noise or preference noise – leads to errors in predictions. The omission of response noise results in unrealistically few violations of dominance, whereas the omission of preference noise results in excessive predictions in which the dominant bet is not selected, contrary to empirical findings.

Such is the importance of adjusting for randomness that well-established violations of expected utility theory may be negated if compatible stochastic models are considered (Wilcox, 2008). Blavatskyy's (2007) reanalysis of 10 prominent experimental studies on risky choice showed that 'stylized empirical facts', including the fourfold pattern of risk attitudes of cumulative prospect theory and the Allais paradox, can be accounted for by the distortion of random errors in computing the expected utility of risky lotteries. Similarly, Wilcox (2008, 2011) re-modelled the risky choice data of Hey & Orme (1994) and found that refutations of expected utility theory based on violations of the common ratio effect would not have resulted if stochastic error was incorporated in the analysis. Stochastic models are thus 'identifying restrictions' for models – they account for violations of models, and thereby effectively protect the underlying theory from wrong falsification (Wilcox, 2008).

> *"…[P]arts, perhaps substantial parts, of what we normally think of as violations of EU may be due to aggregation and stochastic models, rather than nonlinear probability weighting arising from rank-dependent structure"* (Wilcox, 2008, p. 231) [italics in original].

Further implications of misspecified error models and unsuitable analyses are that the correlates of risk preference found in the literature on risky choice could be anomalous. Bhatia & Loomes (2017) provide the example of Charness and Gneezy (2012) who maintain that men are more risk seeking than women based on modal choices in the raw data of the two genders. However, if there was more preference noise in the responses of males than of females, the gender difference could have disappeared or reversed under appropriate analysis. Nevertheless, these authors are somewhat consoling: "Our point is not to reject out of hand all of the differences reported in the studies cited earlier but rather to alert researchers to the possibility that the combination of parameter and response noise may in some cases lead to misestimates of the degree (and occasionally even the direction) of such differences" (Bhatia & Loomes, 2017, p. 684).

Two important conclusions emerge from the modelling and stochastic error research (Bhatia & Loomes, 2017; Blavatskyy, 2007; Blavatskyy & Pogrebna, 2010; Hey, 2001; Krefeld-Schwalb, Scheibehenne, & Pachur, 2019, February 1; Rieskamp, 2008). First, a contextual error model (or compatible equivalent) must be specified and incorporated when using models, such as expected utility theory and rank-dependent utility theory. Second, aggregated or averaged parameter estimates cannot be used for checking model fit or for making predictions or comparisons between the preferences of groups of individuals, as individual level properties of specifications may become lost or distorted.

2.3.5. Evaluation of functional forms for cumulative prospect theory

Several functional forms have been used in risky choice modelling for the value, probability and stochastic error components. Stott (2006) examined the explanatory power of the variety or "menagerie" (Stott, 2006, p. 101) of functional forms available for modelling cumulative prospect theory. As the functions operate jointly, Stott fitted combinations of value, probability weighting and error functions to experimental data and then tested their explanatory power (Stott, 2006). He found the best combination to be "…a power value function, a risky weighting function due to Prelec (1998), and a logit function" (Stott, 2006, p. 101). However, the results of Stott's (2006) analysis showed only marginal differences between the various functions in terms of best fit, with none clearly superior. What did emerge more clearly was that functions with two parameters tended to fit individual participant data better than did single parameter functions (Stott, 2006).

2.3.6. Conclusion on modelling stochastic errors

The literature on the effect of omitting the two sources of stochastic error (response and preference error) from models of risky choice, illustrates that modelling the stochastic error component is a critical aspect of risky choice modelling.

For accommodating response error, there are several stochastic error models, some more suitable than others. The Luce model remains popular but may be statistically incorrect for use with models, such as expected utility and rank-dependent utility. Wilcox's contextual error has been found to be superior and suitable for use with the expected utility, rank-dependent and cumulative prospect theory models. For accommodating preference error, modelling experts advise that a method be used that considers random variability in responses such as Bayesian statistical modelling (Section 4.8.4). They strongly advocate against the common practice of taking modal choice results or sample proportions as indicative of risky preferences. These practices may lead to wrong results, particularly in light of the nonlinear relationships between the model parameters. (See Section 4.8 for consideration of types of statistical techniques and analysis paradigms.)

The possibility exists that the previous work on correlates of choices analysed through modal responses without considering preference error and error in the parameters of the preference functionals themselves, may be at least partly spurious. In addition to the issues of analysis discussed, these issues of methodology highlight the importance of checking and replicating scientific study results.

2.3.7. Conclusion

Sections 2.1 - 2.3 have focused on three fundamental aspects of decision modelling under risk. The literature reviewed on the structural, functional and error model specification of risky choice models sets the scene for the methodology and empirical research of the thesis (Chapters 4 and 5).

The objective of the second part of the literature review (Chapter 3) is to link the literature on the modelling of risky choice to the literature on cognitive ability, numeracy and statistical numeracy as pertinent to the thesis.

Chapter 3: Literature review: Numeracy

This chapter presents numeracy-related literature relevant to the modelling of decision making. The first section (3.1) presents the positions of Simon, Tversky and Kahneman, and Gigerenzer and colleagues. As all of these researchers link the individual's cognitive ability to decision making, this first section bridges the risky choice, cognitive ability, and numeracy literatures. Subsequent sections deal with models of cognitive ability in psychology (3.2), reasoning with probabilities (3.3), empirical studies of statistical numeracy and risky choice (3.4); methodological caveats in empirical literature on risky choice and cognitive ability (3.5); a critique of empirical studies that have modelled risky choice in the context of statistical numeracy (3.6); and the measurement of statistical numeracy in decision-making studies (3.7). Thereafter, the theoretical framework of the thesis is presented in the form of a scientific argument (3.8), followed by a description of its conceptual framework (1.1). An elaboration of the research questions and sub-questions concludes (3.10).

3.1. Theories of the relation between risky decision making and cognitive ability

Herbert Simon (1955, 1956, 1976) was an influential critic of the neoclassical assumption of the perfectly rational decision maker with unlimited knowledge and cognitive ability. His work influenced the judgment and heuristics experiments of Kahneman and Tversky, and the heuristics research of Gigerenzer and colleagues (Campitelli & Gobet, 2010). The next three sections (3.1.1 - 3.1.3) present a brief review of the theoretical positions of Herbert, Tversky and Kahneman, and Gigerenzer and colleagues.

3.1.1. The position of Herbert Simon

Simon viewed individuals as mentally constrained with limited access to information (Navarro-Martinez, Loomes, Isoni, Butler, & Alaoui, 2017). In his Nobel lecture (1977/1992), he summarised his eminent stance on normative theory that decision makers with such limitations were cognitively bounded and could not be rational. If they were, normative theory and descriptive theory would yield closely related results. His information processing theory described *how* a decision maker with bounded cognitive capacity would go about making decisions with incomplete information. It posited selective information searches based on heuristics – or simple rules of thumb – that lead to satisfactory solutions, instead of the complex algorithms of neoclassical rational theory (Shah & Oppenheimer, 2008). He coined the term *satisficing* to describe such solutions as they entailed both

satisfying and *sufficing*, or being good enough, for individuals with bounded computational and information processing capacity (Simon, 1956).

Simon jointly considered the type of task, the environment and the cognitive system of the decision maker, and stressed that formal models of decision making had to be investigated empirically. He likened the task environment to one blade of a scissors and the other blade to the computational capabilities of the decision maker (Simon, 1990), proposing two approaches to research on decision making. The first approach was to use simple behavioural experiments to test neoclassical decision theory, which describes the judgment heuristics experiments of Kahneman and Tversky. The second approach was to use experiments to study the processes of decision making, which later stimulated the research of Gigerenzer and colleagues on heuristics (Campitelli & Gobet, 2010). These approaches are now discussed in turn.

3.1.2. The position of Tversky and Kahneman

Tversky and Kahneman's programme of heuristics and biases of *judgment* preceded their work on cognitive modelling of risky decision making. Their work on judgment was based on observations of people's systematic cognitive biases. Likewise, their work on decision making began with their observations of decision makers' systematic biases in risky choices that violated the rules of normative rational utility theory (Kahneman, 2011). A brief review of their judgment research is presented here, as it suggests an approach to addressing the effects of biases in risky decision making (see Section 6.3) (Toet, Brouwer, van den Bosch, & Korteling, 2016).

3.1.2.1. The heuristics and biases programme

Following Simon, Tversky and Kahneman observed some 20 biases, or errors in judgment, that arise when people use simplifying mental shortcuts, or heuristics, in their intuitive thinking (Kahneman, 2011). They posited that these biases in judgment happen because people answer a simpler question intuitively, instead of using a process of deliberation (Kahneman, 2011). By relying on simple decision rules or limited intuitive heuristics to simplify complex probabilities, the problem solving of cognitively challenged decision makers may be irrational (Toet et al., 2016; Tversky & Kahneman, 1974). An example of a System 1 *heuristic*, or simple rule of thinking that is inconsistent with predictions of normative theory, is the *representativeness heuristic*. This heuristic is the tendency to believe, wrongly, that small samples represent the statistical characteristics of large samples, and may

result in cognitive biases, such as the *hot hands fallacy*. Players in a casino who observe a frequently-occurring number may see themselves on a winning streak and be less likely to quit. By ignoring the known randomness of the underlying process, they are displaying the hot hands fallacy (Dhami, 2016).

Following their work on biases in judgment, Tversky and Kahneman (1971; 1974) spent five years engaged in a series of simple behavioural experiments, often exploring decision makers' probabilistic assessments and inferences in risky decision making. These experiments demonstrated that when people are faced with probabilities they do not understand, they make decisions that deviate from normative rational theory. Tversky and Kahneman thus challenged the assumption of neoclassical economics that an *agent*, or individual, is perfectly rational, and abandoned the idea of rational choice.

3.1.2.2. Dual system process theory of mind

Tversky and Kahneman's explanation of people's cognitive biases and systematic violations of the axioms of rational decision theory, was based on a dual system theory, or model, of the mind. According to dual system theory, when a person's thinking style is appropriate to the problem at hand, she will be less prone to cognitive biases. By contrast, if she employs an intuitive thinking style involving heuristic processing (System 1), when the problem requires an analytic-rational style involving deliberate processing (System 2), she will be more prone to biases (Toet et al., 2016). Furthermore, the intuitive heuristic thinking style is usually associated with individuals with low numeracy, whereas the deliberate, analytical processing style is usually associated with those with high numeracy (Reyna & Brust-Renck, 2014; Toet et al., 2016). While intuitions are processed in System 1, statistical thinking takes place in slower System 2, which *can learn* to avoid the heuristics and biases. There is thus a *prescriptive* principle in the heuristics and biases approach to improve the situation by overcoming the biases measured relative to the normative standard (Basel & Brühl, 2013; Dhami, 2016; Fiedler & von Sydow, 2015; Tversky & Kahneman, 1974).

There are many extensions of dual process theory. For example, the dual process approach is the basis of Lipkus and Peters' (2009) information processing model of numeracy in risky decision making in health. An essential aspect of their framework is how people with different numeracy skills process the numerical information associated with risk. It highlights interventions, such as formats of presentation, and education (Lipkus & Peters, 2009). Further

extensions of the dual process model include the tri-partite model of mind of Stanovich and colleagues (Stanovich & Stanovich, 2010; Stanovich, West, & Toplak, 2016).

3.1.3. The position of Gigerenzer: Fast and frugal heuristics, and simplifying statistics

Unlike supporters of Tversky and Kahneman's tradition of viewing heuristics as cognitive biases, Gigerenzer and colleagues view heuristics as the decision maker's simple adaptations in uncertain environments (Gigerenzer, 1991, 2001; Gigerenzer & Brighton, 2009; Grün-Yanoff & Hertwig, 2016). Under this view, a decision maker's heuristics represent an *adaptive toolkit* modelled on her bounded cognitive abilities. Adaptations involve the selective search for cues that enable smart, efficient and intuitive decision making (Gigerenzer, 2001). Particularly under conditions of uncertainty, but also under risk, the decision maker uses her toolkit for simplifying and adapting to a specific choice situation (Brandstätter et al., 2006). The rules of smart heuristics replace the principle of optimisation to arrive at a satisficing solution (Pachur, Hertwig, Gigerenzer, & Brandstätter, 2013).

Gigerenzer and his colleagues view risky choice as associated with inductive inference and the use of probability theory and statistics (Neth & Gigerenzer, 2015). He attributes the statistical innumeracy of patients and physicians facing risky choices to the way statistics are presented, rather than to internal cognitive errors (Gigerenzer & Edwards, 2003). In Gigerenzer's view, risk literacy and Bayesian thinking can be taught and learned (Gigerenzer, 2015). An example of his solutions to statistical innumeracy and statistical thinking, is changing the presentation format of problems involving probabilities from percentages and probabilities to natural frequencies (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2008; McDowell, Galesic, & Gigerenzer, 2018). The impact of this strategy is disputed, however (Pighin, Gonzalez, Savadori, & Girotto, 2016; Weber, Binder, & Krauss, 2018). Moreover, the entire approach of enabling people through education is different from the 'nudge' approach of Thaler and Sunstein (2008) (see Section 6.3.2).

3.2. Models of cognitive ability in psychology

The research literature on the relation of risky choice to numeracy and statistical numeracy, is often integrated, or confounded, within the psychology literature on general cognitive ability. For instance, in a systematic review and meta-analysis of 287 studies on the relationship between cognitive ability and risk aversion, Lilleholt (2019) considered studies collectively that involved tests of numeracy and those that involved tests of cognitive ability (for example, the Raven's Progressive Matrices) and more general cognitive ability test batteries.

However, as this thesis focuses on numeracy and statistical numeracy, rather than on general cognitive ability, the researcher needs to position numeracy and statistical numeracy within past and contemporary structures or models of general cognitive ability. These models derive from the literature of cognitive psychology.

3.2.1. General cognitive ability

General cognitive ability reflects broad classes of abilities that include reasoning, problem solving, abstract thinking, and the processing of mental information (Lilleholt, 2019). (The terms 'general cognitive ability' and 'general intelligence' are used interchangeably here.)

3.2.1.1. Fluid versus crystallised intelligence models

In the early 1940s, cognitive abilities were dichotomised and referred to as either fluid or crystallised intelligence. The distinction was based on the work of Cattell and Hebb, and known as the *Gf-Gc* theory (Rammstedt et al., 2018). Fluid intelligence is the ability to solve novel problems and reflects capacities, such as memory and processing speed. It develops in early life, peaking in young adulthood. Crystallised intelligence represents knowledge and acquired skills and reflects capacities, such as numeracy, verbal skills and financial literacy that develop over a lifetime (Dohmen et al., 2018).

In 1965, Horn added other categories to the *Gf-Gc* model, including short-term memory and visual processing, as separate abilities. Quantitative knowledge (*Gq*) was added in the early 1990s. By then, the original *Gf-Gc* dichotomy had been expanded into an 8-factor model known as the Cattell-Horn *Gf-Gc* theory. Under the expanded model, numeracy was a separate, additional ability, categorised as neither fluid nor crystallised (Flanagan & Dixon, 2014).

In 1997, Carroll expanded the Cattell-Horn model into a 3-strata hierarchy of cognitive abilities, known as the Cattell-Horn-Carroll model of cognitive abilities (Flanagan & Dixon, 2014; Newton & McGrew, 2010). The model's highest level (stratum 3) is overall general intelligence, or 'g'. The middle level (stratum 2) comprises broad abilities or classes, including fluid reasoning and crystallised intelligence. The narrowest level (stratum 1) comprises the narrow abilities of the broad classes of stratum 2. At the level of stratum 1, fluid reasoning includes inductive and deductive reasoning, and quantitative reasoning defined as the ability "… to inductively…and/or deductively reason with concepts involving mathematical relations and properties" (Newton & McGrew, 2010, p. 623). At this narrow

level, crystallised intelligence, or acquired knowledge, includes verbal knowledge and long-term memory.

A possible theoretical explanation for the unique contribution of numeracy to risky choice behaviour may lie in increasing evidence of a reciprocal relationship between domain-general and domain-specific skills in developing mathematics ability (Cragg & Gilmore, 2014). Numeracy, although categorised as a fluid skill and partially reliant on fluid intelligence capacities, is *domain specific* as it requires specific content and uses crystallised knowledge (for example, the rules of algebra). By contrast, fluid intelligence is *domain general* as it does not use knowledge from a specific content area (Dieckmann et al., 2015).

3.2.1.2. The relation between general cognitive ability and numeracy (empirical evidence)

Various studies note empirical evidence of a weak-to-moderately strong relation between numeracy and general cognitive ability. For example, in a study of 200 students, Brooks and Pui (2010) found a moderately high correlation (r = .41) between scores on the Lipkus numeracy scale and on a general cognitive ability test (the Wonderlic Personnel Test). Furthermore, moderately high correlations are found between numeracy scores and other measures of fluid intelligence, such as working memory (for example, recall of numeric digits in reverse), and reasoning ability (Dieckmann et al., 2015; Peters, Baker, Dieckmann, Leon, & Collins, 2010). Such findings provide evidence that numeracy, although related to general cognitive ability, is nevertheless somewhat distinct from it. Other empirical support for numeracy to be considered a separate construct derives from a national survey of respondents between 23 and 28 years old in the United States, wherein general cognitive ability failed to explain the variance in numeracy scores (Mitchell & Lusardi, 2015).

Dieckmann, Peters, Leon, Benavides, and Norris (2015) attempted to isolate the effect of the cognitive skill of numeracy from fluid and crystallised intelligence, to determine whether numeracy could predict risky choice behaviour uniquely. After controlling for measures of fluid and crystallised intelligence, and for the direct effect of education and other confounding sociodemographic factors, the effect of numeracy on risky behaviour was significant, albeit not strong. The authors interpreted their results in terms of the potentially unique role for numeracy as a cognitive ability (over and above other cognitive abilities).

The moderate-to-strong overlap between numeracy and other cognitive skills apparent in several studies may account for the confounding of the constructs in studies on numeracy and risk.

3.2.2. Cognitive ability, intelligence and rational thinking

The terms ‘cognitive ability’, ‘intelligence’ and ‘rationality’ are often used interchangeably, particularly in the psychological literature. This practice is exacerbated as some theoretical definitions of *intelligence* refer to *rationality* indirectly through references to its components of judgment, decision making and behavioural adaptiveness. However, at a practical level, tests of general intelligence do not measure these – or any – components of rational thinking (Stanovich et al., 2016). Stanovich and colleagues stress that rationality, and cognitive ability as measured by contemporary tests of intelligence, are largely unrelated. Furthermore, the strength of correlations among cognitive biases are generally weak (lower than .2), and at best moderate with measures of fluid intelligence, such as probabilistic and scientific reasoning (maximum correlations range between -.25 and -.35) (Teovanović, Knežević, & Stankov, 2015; Toet et al., 2016; Stanovich et al., 2016). Tests of general intelligence are thus ‘missing’ measures of rationality.

Accordingly, Stanovich and colleagues have developed a test specifically aimed to measure rational thinking. Known as the *Comprehensive Assessment of Rational Thinking*, the test is based on the empirical judgment and decision-making literature, rather than on the intelligence literature. Two of the 20 subtests of this instrument measure statistical numeracy (probabilistic and statistical reasoning, and probabilistic numeracy). Additionally, several other subtests that measure cognitive biases in the heuristics and biases tradition are numerically based, for example, tests that measure ratio bias, avoidance of framing, and temporal discounting (Stanovich et al., 2016). By contrast, the criterion of whether the subtests correlate with intelligence tests was incidental to its development.

3.2.3. Conclusion on the relation between cognitive ability, intelligence, numeracy, and rational thinking

Three conclusions emerge from theoretically-based research on cognitive ability, intelligence, numeracy and rationality. First, general cognitive ability and intelligence should be regarded as distinct from rational thinking, as the measurement of cognitive ability and intelligence in research has failed to measure the components of rational thinking. Second, as the relation among numeracy and general cognitive ability and fluid intelligence is at best moderately

strong, numeracy should be considered as a distinct component of general cognitive ability, and furthermore, distinct from measures of fluid and crystallised intelligence. Third, as statistical numeracy is either inherent in – or correlated with – several tasks involved in cognitive biases in the judgment and decision-making literature, it may mediate or explain the relation between general cognitive ability and decision-making biases.

3.3. Reasoning with probabilities

Reasoning with probabilities is thought to be one of the most difficult domains of mathematics. As risky decision making is defined as decision making under objectively known probabilities of risk (Knight, 1921/2006), the outcome of an individual gamble or prospect is not guaranteed initially. Rather, its outcome is known in the long run, after many trials. It thus entails inferences about the likelihood of uncertain events.

Reasoning with probabilities also incorporates the idea of the random outcome of an individual event, although outcomes are known over a large number of events (Morsanyi & Szucs, 2015). Similarly, inductive logic involves an inference based on premises that provide *a degree of support* for the conclusion of an event rather than *guaranteed* support (Pendlebury, 2013). Thus, in a risky choice situation, the decision maker uses inductive logic as inferences are less than certain (Hawthorne, 2017). Decisions in a complex and uncertain world are often made using inductive logic (Cokely et al., 2016).

Statistical numeracy is seen as necessary for an educated society, as statistically literate citizens who understand risks and ask critical questions will be better-informed (Gigerenzer et al., 2008). However, studies show pervasive low levels of numeracy involving probabilistic thinking. Many people struggle with basic numerical computations and probabilistic concepts used in comparing risk magnitudes, such as computing and comparing proportions and percentages (Carr, 2016; Lusardi, 2012; Peters et al., 2017; Rodríguez et al., 2013). In the 2015 survey of the *Programme for the International Assessment of Adult Competencies* (PISA), designed for cross-cultural and international validity, only a third (34%) of approximately 200,000 adult participants from over 40 countries displayed numeracy levels consistent with, or exceeding "…recognising and working with mathematical relationships, patterns and proportions expressed in verbal or numerical form; and interpretation and basic analysis of data and statistics" (OECD, 2019, p. 48).

Superior numeracy skill has also been linked to the sex of the individual, with males tending to outperform females in numerical tests, higher education levels and higher income levels (Skagerlund, Lind, Strömbäck, Tinghög, & Västfjäll, 2018).

3.3.1. Mechanisms of risky choice

Individuals with higher statistical numeracy tend to make normatively superior choices under risk (Stanovich, 2016). This means they tend to make choices *consistent with* the maximisation principle of expected value theory. Statistical numeracy may have broad predictive power as its measurement involves computations of probability and risk, and inductive thinking. These skills are also required in risky decision making (Reyna & Brust-Renck, 2014). Stated differently, statistical numeracy challenges inductive reasoning in conditions of risk and uncertainty (Ashby, 2017; Cokely et al., 2016). If normatively superior decision making is associated with inductive logic as Cokely and colleagues maintain, it would also be expected to correlate with superior scores on statistical numeracy, as both involve probabilistic thinking.

However, the actual cognitive *process* followed in making normatively superior choices is rarely based on calculations of expected value (i.e. by multiplying probabilities by values). Results from protocol analysis and cognitive process tracking studies of general decision making show that few individuals, even normatively superior decision makers, calculate expected values in situations of risky choice (Cokely et al., 2016). In an experiment of risky choice options involving cognitive process tracing and retrospective protocol analysis, fewer than 5% of 80 undergraduate students actually calculated expected values when making risky decisions in a naturalistic setting (Cokely & Kelley, 2009). However, expected value calculations are a strategy used more frequently in problems involving high payoff choices than in choices between trivial payoffs (Traczyk et al., 2018).

The general conclusion on expected utility is that the normative model of risky choice is an *as-if* model, or at most paramorphic, meaning that the model may describe risky choice correctly, but the process posited in the model – utility maximisation – does not match the underlying empirical processes (Wakker, 2010, p. 3).

Various mechanisms or processes of risky choice have been posited, some of which are described here. They all involve statistical numeracy.

3.3.1.1. Metacognition and smart heuristics

Cokely and colleagues have adopted the smart heuristics theory of Gigerenzer and colleagues. Their position is backed by empirical process tracing results which show that decision makers with higher scores on numeracy tests, compared to those with lower scores, tend to use adaptive cognition with more efficient, simple heuristic search strategies (Cokely & Kelley, 2009). The heuristic searches involve, for example, the relative size of the outcomes and transforming probabilities. In general, reflective cognitive thinking and selective heuristics searches are used rather than cognitive computation of expected values. Decision makers of higher numeracy also tend to spend more time deliberating on the problem prior to making normatively better decisions (Ghazal et al., 2014). They are less affected by emotional distractions and less sensitive to stimuli irrelevant to the decision, for example, the framing of a problem (Peters et al., 2006; Traczyk & Fulawka, 2016).

Corroborating evidence for these cognitive process mechanisms derives from studies of the processes used by experts in risky decision making. An expert is defined as an individual who has acquired special domain-specific skills (Chi, Glaser, & Farr, 1988, as cited in Toet et al., 2016, p.6). Examples of experts are grandmaster players in strategy games, such as chess and Go. Process tracing studies show that such experts use their specialised knowledge in a process of inductive logic, rather than engaging in memory-intensive calculations of expected values. In so doing, they use skilled deliberation (metacognitive processing), and select from the vast numbers of potential options through evaluation of relevant options using elaborate heuristics (Cokely et al., 2016; Garcia-Retamero et al., 2019; Toet et al., 2016).

Despite the emphasis in the cognitive risky choice literature on the superior metacognitive processes of higher numeracy individuals, the relationship between numeracy and superior decision making is not completely explained – or mediated – by metacognitive processes. In studies in which indicators of metacognitive processes are controlled statistically, a significant relation remains between numeracy and risky decision making (Ghazal et al., 2014).

3.3.1.2. Dual process theory

An alternative theoretical explanation for the mechanism of risky decision making involving statistical numeracy is based on dual process theory and the heuristics and biases programme of Kahneman and Tversky (Kahneman, 2011; Stanovich & Stanovich, 2010). As previously outlined, dual process theory posits the interaction of fast Type 1 cognitive processes

associated with intuitive decision making, and slow Type 2 cognitive processes associated with deliberate, rational decision making (Lilleholt, 2019). The executive function and working memory required for Type 2 processing are associated with higher cognitive functioning, specifically with fluid intelligence and numeracy (Lilleholt, 2019; Stanovich, 2016). (The terms 'Type' and 'System' are used interchangeably in the thesis.)

Furthermore, dual process theory accounts for the negative relation frequently found between risk aversion and cognitive ability (Section 3.1.2.2). Individuals with higher levels of cognitive ability who engage in Type 2 deliberation are more likely to see that it is irrational to be risk averse when choosing between risky choices of small or moderate outcomes (Lilleholt, 2019). Thus, they would tend not to show risk aversion over choices of the size typically offered in laboratory experiments.

3.3.1.3. Affect

A third mechanism involved in the evaluation of probabilities in decision making under risk is thought to be *affect* or emotional reactions to stimuli (Peters et al., 2006). Probabilities have been found to be obscured by emotional reactions to stimuli (Pachur, Hertwig, & Wolkewitz, 2014). At the moment of choice in an extreme situation involving highly aversive outcomes, for example, when presented with the serious side effects of a drug, people may focus on the consequences of risky choices only, and disregard the probabilities of their occurrence. Taking out insurance for low probability risk is another example. This phenomenon has been termed "risk as feelings" (Loewenstein, Weber, Hsee, & Welch, 2001, p. 267), probability neglect, and "the primacy of outcome over probability" (Brandstätter et al., 2006, p. 5). By contrast, in affect-rich or positive affect decisions, such as winning big lottery prizes, small probabilities tend to be overweighted.

Studies have found a stronger relation between risk aversion and affect, than between risk aversion and cognitive factors (Pachur, Mata, & Hertwig, 2017a). In a study of age-related differences, Pachur et al. (2017a) found older adults (63 years or older) to be less risk averse than younger adults (30 years or younger) in gain-framed choices, despite having lower cognitive ability than the younger participants. They attributed this finding to the older adults' positive affect and motivation to take risks in positively framed lotteries.

The affect mechanism appears to be associated with numeracy, as it applies less to higher numeracy decision makers than to those with lower numeracy (Petrova, van der Pligt, & Garcia-Retamero, 2013). Petrova et al. (2013) found more numerate respondents to be more

sensitive to probabilities than less numerate respondents on reappraising the loss of an affect-rich outcome. The affect-choice-numeracy relation is complex as it may also be explained by the type of heuristics employed by respondents of higher versus lower numeracy, or by cross-cultural differences (Pachur & Galesic, 2013). Then again, the same type of thinking and emotion may be present when taking a statistical numeracy predictor test as when making emotive decisions in real-world situations (Cokely et al., 2016; Lusardi, 2012).

3.3.1.4. Neuroscience

Brain research in psychology and economics explores a physiologically-based mechanism underlying the relation between cognition and risky choice (Dohmen et al., 2018). Although neuropsychological research was conducted initially on the risky decision making of patients in clinical settings, the focus has been on investigating the risky choices of healthy adults over the past decade (Brand, Labudda, & Markowitsch, 2006).

Research evidence on the areas of the brain that underlie numeracy and decision making is inconclusive (Reyna & Brust-Renck, 2014). As reviewed by Dohmen et al. (2018), studies of nonhuman primates and rats on the processing of risk suggest that processes of risk and probability may be generalised across the brain at the level of single neurons (Burke & Tobler, 2011). However, neuropsychological studies on humans suggest that numeracy and statistical numeracy are fundamental to a neuropsychological theory of risky decision making. In Brand et al. (2014), the processing of ratios and complex probability mediated, or explained, the relationship between basic cognitive functions, such as logical reasoning and basic mental calculations, and decision-making performance under risk (Brand et al., 2006). Furthermore, good probability processing may compensate for weak executive brain functioning in making risky choices (Schiebener, Zamarian, Delazer, & Brand, 2011).

Neuropsychological research finds support for the importance of the mechanism of affect (see Section 3.3.1.3) as findings suggest that the region of cognitive processing in risky decision making may depend on the affect level of the outcomes (Suter, Pachur, Hertwig, Endestad, & Biele, 2015). In a study on the neural basis of risky choice of 23 adults involving both neuroimaging analyses and computational modelling of risky choice data, Suter et al. (2015) identified different regions of brain activation dependent on whether the risky choice involved less emotional, i.e. affect-poor, outcomes (risky monetary choices), or highly emotional, i.e. affect-rich, outcomes (risky medical choices). Decision mechanisms underlying affect-poor choices involved the activation of neural regions that reflected the

processing of probability information, whereas mechanisms underlying affect-rich choices were associated with the emotional value of the outcomes. Furthermore, a complementary computational modelling exercise showed less probability distortion in affect-poor choices than in affect-rich choices. When faced with affect-poor choices, participants selected options with higher expected values more often than when facing affect-rich choices. They tended to ignore the objective probabilities of affect-rich choices, and avoid options with worse side effects. These findings suggest that the association of numeracy and risky choice holds when choices are not highly emotive.

3.3.1.5. The education-numeracy relation

A fifth mechanism of the association of statistical numeracy in risky decision making involves the education level of the decision maker. There are two opposing views of the role of education: *numeracy* may mediate (or explain) the relation between education and risky choice, or *vice versa*. On the former, Dieckmann et al. (2015) review studies that show that individuals who are educated acquire cognitive skills, such as numeracy, which in turn correlate with superior risky choice. According to the principles of mediation (Baron & Kenny, 1986), numeracy would partially or fully mediate the relationship between education and risky choice if, when the numeracy variable is controlled, the effect of education on risky choice is subsequently reduced. By contrast, *education* may mediate (or explain) the relation between numeracy and risky choice. Several studies have found a relation between cognitive skills – in domain-general fluid abilities, such as numeracy and crystallised intelligence – and superior decision making in health settings (Dieckmann et al., 2015). Individuals with high cognitive abilities, such as numeracy, may seek out further education which then relates to superior decision making. Education would partially or fully mediate the relationship if, when the education variable is controlled, the effect of numeracy on risky choice is subsequently reduced. [It is acknowledged that more modern mediation tests exist in the literature, for example, VanderWeele (2016).]

3.3.1.6. Number processing skills

Finally, an intuitive magnitude system associated with the encoding of numerical information is thought to relate to the processing of numerical values and probabilities, and so may explain distortions in risky choice (Xing et al., 2019). An example of such processing is contained in *Fuzzy-trace* theory. People who understand probability and risk extract and encode the intuitive, fuzzy meaning, or *gist*, of the probability information in risky choices,

rather than the exact or *verbatim* facts (Reyna & Brust-Renck, 2014). A second example of such processing is thought to occur in the *approximate number system*, associated with superior number approximation, as distinct from calculating exact proportions. A third example involves the extent of the accuracy of mappings of symbolic numbers onto mental magnitudes (Schley & Peters, 2014). Superior gist processing, approximation skills, and symbolic-number mappings are thought to be associated with superior numeracy, lack of distortions in value and probability judgment, and advantageous for decision making under risk (Mueller, Schiebener, Delazer, & Brand, 2018; Patalano et al., 2015; Reyna & Brust-Renck, 2014; Schley & Peters, 2014).

3.4. Empirical studies of numeracy and risky choice

This section presents empirical studies on the relation among risky choice and cognitive ability, numeracy and statistical numeracy. They are presented as part of the rationale of the thesis for considering statistical numeracy in models of risky choice.

There are a number of reviews of such studies. These include the reviews of Reyna and colleagues (Reyna & Brust-Renck, 2014; Reyna et al., 2009), Cokely and colleagues (Cokely et al., 2016; Garcia-Retamero et al., 2019), Dohmen and colleagues (2018), the review and meta-analysis of Lilleholt (2019), and recent working papers (Amador Hidalgo, Brañas-Garza, Espín, Garcia, & Hernández Román, 2019; Alipour, 2019). The studies considered derive mainly from the cognitive psychology and judgment literatures. They do not involve estimating the parameters of structural models based on parametric value and probability functions. Instead, they are correlational or descriptive in design, rather than causal. To qualify as causal studies, they would need to meet specific requirements of experimental design, control and measurement. This would be extraordinarily difficult in the context of risky decision making from ethical and design perspectives.

The studies in this section are grouped into the following four categories according to the dominant content of each study, as categories do overlap. These are: statistical numeracy and decision making in health (Section 3.4.1), numeracy and financial decision making (Section 3.4.2), numeracy and decision making using risky prospects (Section 3.4.3), and the relation between cognitive ability and risk aversion (Section 3.4.4).

3.4.1. Numeracy and decision making in health

Increased recognition of collaborative physician-patient decision making involving risk has produced a subfield of decision making in medical judgment and decision making (Anderson

& Schulkin, 2014). Studies of risky choice in health usually involve choices based on quantitative risks and benefits of treatments or screenings expressed in terms of percentages, proportions and survival rates. To make informed medical decisions, patients and practitioners must understand the associated statistics (Anderson & Schulkin, 2014; Elstein, 2004; Galesic & Garcia-Retamero, 2010; Hibbard et al., 2007; Reyna & Brust-Renck, 2014; Reyna et al., 2009).

As reviewed by Garcia and colleagues (2014, 2019), several studies document the positive relationship between statistical numeracy skills and the quality of patients' risky medical decisions and behaviours. Statistical numeracy has been positively related to the accuracy of perceptions of risks and benefits of medical interventions, the quality of health-related decisions involving risk, and of shared decision making. Patients with lower numeracy levels often fail to understand probabilistic information on risk. Compared to patients with higher numeracy, they tend to perceive risky outcomes less accurately, and to rely on physicians to make their risky choices on their behalf. They also tend to be less responsible in managing their diseases, and are at least 40% more likely to develop comorbid conditions (Garcia-Retamero & Cokely, 2014; Garcia-Retamero et al., 2019).

Research on decision making in health has emphasised quantitative measurement of numerical ability, giving rise to well-researched measurement scales of statistical numeracy (Garcia-Retamero & Cokely, 2014; Peters et al., 2006; Schwartz et al, 2005).

3.4.2. Numeracy and financial decision making

Complex numerical calculations involving percentages and interest rates are often required for financial decision making. In Lusardi's (2012) review of studies on the relation between numeracy and consumer behaviour, individuals with lower numeracy had inferior retirement planning, inferior pension fund participation, less ownership of stocks, and less accumulated wealth. Furthermore, numerical ability – as opposed to financial literacy, verbal and other aspects of cognitive ability – is strongly negatively related to defaults in mortgage payment (Gerardi, Goette, & Meier, 2013). Benjamin et al. (2013) cite examples of studies, including their own 2006 research, that indicate a relation among cognitive ability, the accumulation of assets, and successful financial market-related risk preference. Likewise, Dohmen et al. (2018) review several studies that find that individuals with superior numeracy and other cognitive skills tend to make more financially successful decisions under risk. Compared

with others with lower cognitive skills, these individuals take considered risks in their investment portfolios associated with successful economic outcomes.

3.4.3. Numeracy and decision making using risky prospects

In laboratory experiments of risky prospects, individuals with higher statistical numeracy levels tend to make normatively superior choices under risk, i.e. choices advocated by expected value theory (Anderson & Schulkin, 2014; Cokely et al., 2016; Garcia-Retamero & Cokely, 2014; Garcia-Retamero et al., 2019). Laboratory studies have also shown that more numerate individuals are less affected by incidental, irrelevant and emotional attributes of the decision situation (Garcia-Retamero et al., 2019; Jasper, Bhattacharya, & Corser, 2017; Peters et al., 2006). University students higher in statistical numeracy are less susceptible to differences in the wording used to frame a risky choice. More numerate students tend to display more consistent responses to risky choices, and are less affected by the presentation format of numerical information (for example, an option expressed as a percentage versus a frequency) (Peters et al., 2006). [They also exhibit a more linear probability weighting function with fewer deviations from the linear expected value model (Traczyk & Fulawka, 2016).] People with low numeracy are less likely to use numeric information to make a decision and instead recruit nonnumerical information, such as the strength of the affective reaction to an option (for example, Peters, 2008; Peters et al., 2006, 2009).

Participants' susceptibility to decision paradoxes under expected utility theory have been found to be dependent on their statistical numeracy level (Liberali, Reyna, Furlan, Stein, & Pardo, 2012). In the first major replication study of Kahneman and Tversky's (1979) study, Millroth, Nilsson, and Juslin (2019) measured the risky choices of over 1,800 respondents using the original set of 16 risky prospects from the 1979 study. These prospects were originally constructed to represent observed psychological principles that were paradoxes under expected utility theory, and were the motivation for original prospect theory. Further to the 1979 study, Millroth et al. (2019) measured their participants' statistical numeracy using the Berlin Numeracy Test. Their analyses compared the extent of replication of the original study results dependent on statistical numeracy, and did not involve the modelling of the parameters of the structural model. Millroth et al. found that the original paradoxes were more typical of *highly* numerate individuals, than of low. Individuals with lower numeracy levels tended to minimise the risk of receiving the worst outcome of each choice pair. Their systematic choice strategy did not reflect the paradoxes posited by Kahneman and Tversky in 1979. This finding concurs with studies that have shown the tendency for people with lower

cognitive ability to be more risk averse (Dohmen, Falk, Huffman, & Sunde, 2010). However, this finding is couched more in gain-framed prospects, and even then, may be weak. Furthermore, Pachur and Galesic (2013) found that less numerate respondents tended to follow the affect (nonnumerical) heuristic, and more numerate respondents tended to follow the minimax heuristic. Additionally, as reviewed by Lilleholt (2019), several studies, including those of Stanovich and colleagues, have found lower cognitive ability associated with more cognitive biases.

3.4.4. Cognitive ability and risk aversion

Several experimental studies have found a negative relation between cognitive ability and risk aversion. These studies usually use the multiple price list method, or a variant of it (see Section 4.4.2.1), to estimate risk aversion, and various indicators of cognitive ability (Andersson, Holm, Tyran, & Wengström, 2016). For example, Dohmen et al. (2010) found a negative relation between risk aversion and non-numeric measures of fluid and crystallised dimensions of cognitive ability (a symbol correspondence test and a test of word fluency). Their study involved approximately 1,000 German adults. Burks, Carpenter, Goette, and Rustichini (2009) measured the cognitive ability of 1,000 trainee truckers using a test of basic quantitative literacy and a standard intelligence test designed to measure fluid intelligence (Raven, 2000, as cited in Burks et al., 2009, p. 774). Their results showed a significant negative correlation, but of weak effect size (r = -.16). Other examples are two laboratory studies on the risk attitudes of Chilean high school students (Benjamin, Brown, & Shapiro, 2013). In these studies, Benjamin et al. measured cognitive ability using mathematics and verbal achievement scores from a standardised national examination. Their results showed that students' small-stakes risk aversion was negatively related to standardised cognitive test scores, and the effect size was moderately large. L'Haridon and Vieider (2019a) related grade point average (GPA) as a proxy for cognitive ability in a study of the risky choices of approximately 3,000 individuals across 30 countries. They too found a negative relation between cognitive ability and risk aversion, although weak.

The theoretical justification for a negative relation between intelligence and risk aversion, and between numeracy and risk aversion is that highly intelligent or numerate people are likely to make risky choices that correspond to the maximisation principle of expected utility theory, thereby avoiding risk aversion. By contrast, those of lower intelligence or numeracy are not. This reasoning is also consistent with the calibration theorem of Rabin (2000). The theorem holds that risk attitudes must be approximately neutral for small and moderate

stakes, or outcomes, in order for risk aversion to hold for high stakes. In general, responses of higher numeracy individuals that correspond to the maximisation principle of expected utility theory will show little-to-weak risk aversion, especially for the small stakes or outcomes typically found in laboratory experiments. Thus, in laboratory experiments, the attitudes of highly numerate participants are likely to be risk neutral, irrespective of the associated probabilities. By contrast, individuals with lower levels of fluid intelligence, whose attitudes are inconsistent with the maximisation principle and with Rabin's theorem, may display risk aversion for stakes of any size. Hence the negative (and weak) relation.

However, findings on the relation between cognitive ability and risk aversion are not all in agreement. The next section presents methodological caveats in this category of research.

3.5. Caveats related to methodology in empirical literature on risky choice and cognitive ability

There are two main concerns of methodology in studies of the relation among risk attitudes, cognitive ability, numeracy and statistical numeracy. These concerns relate to the framing of the choices, and the confounding of error and biased elicitation methods employed to measure risk preference.

3.5.1. The framing of risky decisions

Risk preferences seem to depend on whether risky choices are presented in a gain, loss or mixed frame. Lilleholt (2019) conducted a meta-analysis of experimental studies that involved measures of cognitive ability, including numeracy (but not financial numeracy) and risk taking, considering all three frames. For choices framed as gains (97 studies), Lilleholt's results showed negative correlations between cognitive ability and risk. On meta-analysis, individuals with higher scores on indicators of cognitive ability tended to be less risk averse than those scoring lower. However, the correlations were weaker than originally reported. Further, no consistent significant relationships between cognitive ability and risk aversion were found in studies in which choices were framed in the loss and mixed frames (12 and 41 studies, respectively). The negative relation between risk aversion and cognitive ability appears to hold only in the frame of gains, where it is weak.

3.5.2. Confounding errors in risky decisions with cognitive ability and choice architecture

Section 2.3 presented theory on the importance of considering both stochastic and response error in specifying models of risky choice. This section presents another potential challenge related to random response errors. This is the challenge of systematic measurement error in elicitation methods. Systematic measurement error occurs, for example, when a multiple price list, or a list of choices, is constructed in such a way that it offers more opportunities for choices reflecting risk aversion than risk seeking, or *vice versa*. In the decision-making literature, this effect is known as the "choice architecture" of the elicitation method (Dohmen et al., 2018, p. 125).

Various researchers have conducted elaborate experiments and analyses to test whether the choice architecture or structural design of risky choice tasks can induce decision biases (Andersson, Holm, Tyran, & Wengström, 2018; Holden & Tilahun, 2019). Based on the results of their studies on the multiple price list method, Andersson and colleagues (2016, 2018) dispute claims of a negative relation between cognitive ability and risk aversion. Their argument is based on two frequent findings: First, random errors are more prevalent in the risky choice responses of individuals with lower cognitive ability (Benjamin et al., 2013; Burks et al., 2009). Second, methods used to elicit risky choices are often *systematically* biased towards choices reflecting a particular risk attitude, such as risk aversion. The noise in the responses of individuals with lower cognitive ability, combined with measurement bias, results in the choices appearing more risk averse than they are. Furthermore, as the responses of individuals with higher cognitive ability are less susceptible to both errors, a spurious negative relation between cognitive ability and risk aversion results. A counterclaim is that the bias in choice architecture is equally likely to occur in the direction of risk seeking.

In this thesis, the researcher adopts the neutral stance of Dohmen et al. (2018), who maintain that the mechanisms through which cognitive ability affects risky choice have not been identified as error, nor through some aspect of cognitive ability. The nature of the relationship between the latent traits of cognition and risk preference has not been established, and remains open to investigation (Dave, Eckel, Johnson, & Rojas, 2010; L'Haridon & Vieider, 2019a). Furthermore, the research questions and sub-questions of this thesis are not directly concerned with the underlying processes or pathways involved in modelling the relation between statistical numeracy and risk attitudes.

The following section reviews studies that have aimed to relate cognitive ability or numeracy to estimates of risky choice model parameters.

3.6. Studies on modelled risk attitudes and cognitive ability/numeracy

Studies that have estimated parameters of structural models and also related these estimates to indicators of cognitive ability or numeracy have used either comparisons or correlations.

The comparison method involves estimating and comparing the parameters of structural models of distinct numeracy groups. The quality of the inferences derived is dependent on the adequacy of the numeracy groupings (aside from the modelling imperatives of structure and stochastic specification). Poorly constructed numeracy groups that overlap, or are unreliable or invalid, detract from the power and validity of the analysis and subsequent inferences.

The correlation method examines the linear relations between indicators of cognitive ability or statistical numeracy, and the individual model parameter estimates. The relations are computed via pairwise correlations or via multiple regression analysis of each estimated parameter on the numeracy indicators.

3.6.1. Comparison of model parameters of numeracy groups

To the knowledge of the researcher, few studies have used the comparison method to categorise respondents into statistical numeracy groups and then estimated structural models of risky choice using parametric, function-dependent estimations of utilities/values and probabilities. This method of comparison was used in the study of Patalano et al., (2015). Among other questions, the authors investigated whether numeracy plays a role in value and probability distortion in modelling risky choices. They formed two numeracy groups, and estimated parameter values of the cumulative prospect theory model for each. As their objective was the same as one of the central objectives of this thesis, their study methodology and results are presented in detail here.

Patalano et al. (2015, p. 1820) claim to have found “moderately strong negative correlations between numeracy and value and probability distortion”. Their sample comprised 57 undergraduate university students, 51 of whom had usable scores. They measured risk attitudes using certainty equivalents obtained from a series of monetary gambles. They measured statistical numeracy using the 11-item numeracy scale of Lipkus et al. (2001) and allocated the students to high versus low numeracy groups based on a median split of the numeracy scores. They published the following numeracy summary statistics, relative to the

possible maximum score of 11: mean = 9.7, standard deviation = 1.5, range = 6-11, median = 10.5, and skewness = 0.80 (Patalano et al., 2015, p. 1826).

One of the dependencies for their findings to be credible was the reliability and validity of their allocation of participants to numeracy groupings. However, reconstruction of their data, based on their published statistics, does not support this dependency. Crucially, their numeracy-based median split is highly questionable. A median split of the sample, based on a value of 10.5 out of 11, would imply that all participants (except perhaps one) in the higher numeracy group must have scored 11/11 as scores on the test are always integer values. By simulation, given the remaining summary statistics supplied, there must have been several participants who scored 10/11 in the lower numeracy group. Thus, the numeracy scores of these individuals would have been negligibly different from those of the high numeracy group. Importantly, these differences could well have been due to measurement error. Measurement error is indicated, given that the internal consistency of the Lipkus scale (alpha coefficient = .54) supplied by Patalano et al., was below the barely acceptable cutoff of .6 (Hair, Black, Babin, & Anderson, 2010). Moreover, this alpha value is gravely low considering the number of items in the Lipkus scale: at 11 items, this internal consistency value is already inflated as the alpha coefficient is correlated with the number of items in the scale. [It is noted that the alpha coefficient is more correctly referred to as Kuder-Richardson Formula 21 for tests with dichotomous items (Ghiselli, Campbell, & Zedeck, 1981). Coefficient alpha implies Kuder-Richardson Formula 21 throughout the thesis.] Another observation in the results is that the positive skewness value supplied for the numeracy scale is not feasible given the direction of the difference between the mean and median values. Other (non-measurement) concerns of the study include the authors' assessment of the fit of the cumulative prospect theory model without considering individual heterogeneity (Bhatia & Loomes, 2017; Wilcox, 2008; Wilcox, 2015), and without including negative and mixed prospects. The methodology of this study is problematic, and thus the findings are unsupported.

Other studies have used numeracy as a fixed effect in mixed factorial designs for categorising individuals. Examples include those of Traczyk and Fulawka (2016), and Millroth and Juslin (2015). Some of the results of these designs are complex. For instance, Millroth and Juslin categorised their respondents into high versus low numeracy groupings based on the Berlin Numeracy Test (Cokely et al., 2012). Inadmissibly, they used gain-framed prospects only to estimate cumulative prospect theory parameters. They further examined the effect of the

presentation format of probability information on the risky choices of 340 participants in an elaborate 3x2x3x11 mixed factorial design. The results claimed a significant 4-way interaction involving cumulative prospect theory parameter etimates and value, probability, numeracy and format effects. The conclusion drawn was that the presentation format of probability information changed people's evaluations of risk, depending on their numeracy levels. Similarly, Traczyk and Fulawka found an interaction effect between the curvature parameter of the probability weighting function and affect irrelevant to the decision task, with low numeracy individuals more sensitive to the irrelevant stimuli. Although these research results are important for motivating the research, they do not directly fill the knowledge gap addressed in this thesis.

Pachur et al. (2017a, 2017b) provide an example of a study that estimated structural model parameters for younger versus older individuals, incorporating stochastic error and prospects in all frames. Although their categorising variable was age rather than numeracy, the study provides useful analytic methods for the thesis.

3.6.2. Correlation of model parameters with statistical numeracy indicators

In a study that included some of the authors of the Patalano et al. (2015) study, Xing et al. (2019) used the correlational approach to relate risky choice model parameters to statistical numeracy. As in their 2015 study, the procedure of calculating certainty equivalents using gain-framed prospects was used, this time to test a version of cumulative prospect theory based on multiple reference points. The authors once again used the 11-item numeracy scale of Lipkus et al. (2001), but included the 8-item abbreviated numeracy scale of Weller et al. (2013) to measure statistical numeracy. The two sets of numeracy scores correlated highly (r = .65), but were considered separately. Both correlated negatively with distortions in value, and correlated moderately positively with the reliability of responses across repeated trials, with more numerate individuals providing more consistent responses (Xing et al., 2019).

Another example of the use of the correlational approach is the recent study of L'Haridon and Vieider (2019a). These authors related cognitive ability to estimates of the parameters of the Prospect theory model, using the 2-parameter weighting function of Prelec (1998). Their study was based on the risky choices of approximately 3,000 individuals across 30 countries. The data included the grade point average of each individual as a cognitive ability proxy, and other demographic and country-level covariates. They incorporated sources of stochastic error. Their estimations were based on a log-likelihood function linearly dependent on the

covariates. Regression of the individual parameters of the structural model on the cognitive ability proxy showed that grade point average scores correlated positively with sensitivity to risky choices in both gain and loss frames, and correlated negatively with loss aversion. Compared to individuals with higher GPA scores, those with lower GPA scores discriminated less well between probabilities that lie away from the extreme values of 0 and 1. Their responses were also noisier. Although these effects were weak (L'Haridon & Vieider, 2019b), they were nevertheless significant in the context of the extreme power associated with a large sample size (n = 2939).

3.6.3. Conclusion on model parameter estimation and statistical numeracy

To the knowledge of the researcher, there is no study that adequately assesses the role of statistical numeracy in modelling risky choice. Existing empirical studies that estimate the parameters of risky choice models under expected utility, rank-dependent or cumulative prospect theory models are inadequate for the main purpose of the thesis. Aside from the imperatives of considering stochastic error and other modelling and design factors, the results of studies that compare structural model parameters estimated for separate numeracy groups are questionable, largely in terms of the numeracy categorisation employed. These studies generally use statistical numeracy scales that are overly easy or difficult. They then use a convenient split of the numeracy score distribution to form the numeracy groups, without considering the implications for the construct validity of numeracy.

Studies that correlate estimated model parameters with statistical numeracy, or proxies thereof, are also problematic. These studies typically consider statistical numeracy as one of several covariates, and either regress each of the parameters on the set of covariates, or examine the pairwise correlations between each of the covariates and the separate parameter estimates. One potential problem associated with this approach involves the assumption of a *linear* relation between the numeracy covariate and each of the estimated structural parameters. As model parameters are transformed via nonlinear functions, relationships may not be linear. Second, risk attitudes may depend on a combination of parameters. For example, the degree of risk aversion under rank-dependent utility depends on both the utility and the probability weighting parameters (Wakker, 2010). Further, the curvature and elevation parameters of the weighting function together describe risk attitudes better than separately. Finally, the parameter estimates may be highly intercorrelated, and may compensate for each other (Spektor & Kellen, 2018), rendering pairwise correlations suspect.

Thus, neither approach fully, nor adequately, describes whether the model(s) fit the risky choices of individuals in different numeracy groups.

The review now turns to the state of measurement of statistical numeracy in the psychology and decision-making literature.

3.7. Numeracy measurement

There are distinct differences in the approaches to the measurement of numeracy in psychological settings compared to in decision research.

3.7.1. The measurement of numeracy in psychological research

Many contemporary psychometric tests of intelligence are structured within the Cattell-Horn-Carroll model of cognitive abilities (Newton & McGrew, 2010). In general, intelligence testing is undertaken in a school, personnel environment, or other situation related to the need for psychometric measurement. In these instances, validated and reliable psychometric tests of numeracy and other cognitive skills are administered by a suitably qualified psychometrist or psychologist under controlled, timed conditions. As reviewed by Lilleholt (2019), measurement scales of general cognitive activity include the Wechsler Adult Intelligence Scale (WAIS), and the *Ravens Progressive Matrices* test, an example of a measure of fluid intelligence. The place of numeracy within the theoretical structure of general cognitive ability was discussed in Section 3.2.

3.7.2. The measurement of numeracy in risky decision making

The testing of numeracy in the risky decision-making literature differs from psychometric testing in several ways. First, as previous stated, the concept of risky choice is defined in terms of objectively stated probabilities. A risky choice, by definition, involves *statistical numeracy*. Thus, the testing of statistical numeracy, rather than general numeracy, is more relevant in empirical decision-making research. Second, while general numeracy skills are defined in terms of the ability to use and understand numbers based on fundamental mathematical operations, statistical numeracy is more specifically defined in terms of being able to calculate and compare percentages and proportions (Cokely, Ghazal, & Garcia-Retamero, 2014; Peters et al., 2006). There are various alternative terms for statistical numeracy used in the numeracy literature, such as ‘quantitative literacy’ or ‘risk literacy’ (Anderson & Schulkin, 2014). Third, statistical numeracy is intended to predict risky choice. Thus, the intention of the testing is different from the psychometric setting (Brooks & Pui,

2010). Finally, statistical numeracy tests used in the decision-making literature are typically administered by the researcher, not necessarily timed, nor administered under controlled conditions.

Throughout this thesis, the term ‘statistical numeracy’ is used to refer to the ability to process information presented in terms of proportions, percentages or probabilities. Furthermore, only performance-based numeracy measures are considered. These measures require mathematical computation, rather than subjective reporting of ability. Additionally, the terms ‘tests’ and ‘scales’ are used interchangeably, as scales refer to measurement scales or instruments. Finally, the terms ‘numeracy’ and ‘statistical numeracy’ are used interchangeably.

3.7.3. Measurement scales of numeracy in risky decision making

Measurement scales of statistical numeracy emphasise risk, probability and computations involving ratios. These scales require the respondent to make probability judgments (Reyna & Brust-Renck, 2014). The most popular tests found in the risky choice literature are the numeracy scale of Schwartz, Woloshin, Black, and Welch (1997), the extension of this scale by Lipkus et al. (2001), the Berlin Numeracy Test (Cokely et al., 2012), and the Cognitive Reflection Test (Frederick, 2005). All of these tests other than the Cognitive Reflection Test were developed specifically to predict the ability of respondents to understand risk.

3.7.3.1. The 3-item numeracy scale of Schwartz, Woloshin, Black, and Welch (1997)

The measurement of numeracy in decision making in health was popularised in 1997, when Schwartz et al. published a 3-item numeracy scale. The items were all framed in a general, i.e. non-medical, context and involved probabilites and proportions. To illustrate, one of the items asked:

> *In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize is 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy by a single ticket to BIG BUCKS?*

Schwartz et al. (1997) used this scale to assess whether the female patients’ perceptions of the risks and benefits of mammography screening were related to their numeracy levels. They found that only patients who understood basic probability and numerical concepts could judge the meaningfulness of quantitative risk reduction, regardless of the mode of presentation of the data on risk. The three items are framed in a general, non-medical, context.

3.7.3.2. The 11-item numeracy scale of Lipkus et al. (2001)

Lipkus et al. (2001) extended the Schwartz 3-item scale by adding 8 items. The additional items all involve the risk of a disease or infection. For instance:

> *Which of the following numbers represents the biggest risk of getting a disease?*
>
> *___ 1 in 100, ___ 1 in 1,000, ___ 1 in 10?*

The scale was intended to discriminate between the numeracy levels of highly educated respondents, and Lipkus et al. (2001) maintained that even college-educated adults found his relatively simple numeracy questions *difficult*. However, the mean score of these respondents on the 8 additional items was 6.2 out of 8, i.e. 78%, with 80% of respondents answering at least half of these items correctly. Moreover, numerous other studies have found the Lipkus 11-item scale overly easy for educated respondents (Cokely et al., 2014; Cokely & Kelley, 2009; Patalano et al., 2015; Peters et al., 2006). In the 2014 study of Cokely et al. based on an educated adult sample, the score distribution on the Lipkus scale was strongly negatively skewed with a mean value of 82% and median value of 91% (or 10/11), i.e. close to the maximum score, indicating a measurement ceiling effect. The ceiling effect was even clearer in the study of Patalano and colleagues (2015) with a mean of 88% and median of 95%.

3.7.3.3. The Berlin Numeracy Test (Cokely et al., 2012)

The Berlin Numeracy Test was developed initially at the Max Planck Institute for Human Development. It is part of RiskLiteracy.org, a nonprofit university-based virtual collaboration of international researchers of informed decision making, directed by Cokely, Feltz, and Garcia-Retamero (RiskLiteracy.org, 2020).

Cokely et al. (2012) designed the 4-item test of numeracy to predict risk literacy, or the ability to use information about risk to make good decisions. Their initial testing was based on a community sample from Berlin, hence the name of the test.

Whereas the Schwartz, Lipkus and Cognitive Reflection tests were all constructed in the tradition of classical test theory, Cokely et al. used a second-generation method of test construction known as an *adaptive decision tree* to identify items that discriminated optimally across a range of numeracy difficulty levels (Cokely et al., 2014). This method departs from the assumption of classical theory, that a person's measured test score is a composite of her true score on the latent construct, and measurement error. To minimise measurement error, the classical test constructor eliminates items that are correlated more weakly with the total

test score, thereby arriving at a shortened test with higher internal consistency (Hair et al., 2010; Weller et al., 2013). However, this practice may narrow the construct, and thus restrict its predictive scope (Smith, McCarthy, & Anderson, 2000, as cited in Weller et al., 2013, p.200).

According to the authors, the test has shown robust psychometric properties, such as convergent and discriminant validity across studies on more than 5,000 respondents from 15 counties (Cokely et al., 2012, 2014). It takes about 3 minutes to administer, and different formats may be used.

One of the items is phrased as follows:

> *Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in a choir 100 are men. Out of the 500 inhabitants that are not in a choir 300 are men. What is the probability that a randomly drawn man is a member of the choir?*

While the Lipkus scale has been criticised for being overly easy for tertiary educated populations, the Berlin Numeracy Test is criticised for being overly difficult. In a recent online study of approximately 1,800 participants in the United States aged between 18 and 75 years, 41% of respondents failed to answer any of the 4 items correctly, and a further 31% answered only one of the four items correctly (Millroth et al., 2019). Thus, for respondents with lower statistical numeracy, the test is recommended to be used together with another test of numeracy. One of the test formats combines its items with the three items of the Schwartz scale for use in populations with lower numeracy levels. However, research on the psychometric properties of this scale were not found in the available literature.

Correlations between the Berlin Numeracy Test, versus the Schwartz 3-item scale and the Lipkus 11-item scale, range from .4 to .5, depending on the version of the Berlin Numeracy Test used (Cokely et al., 2014).

3.7.3.4. The Cognitive Reflection Test (CRT) (Frederick, 2005)

Another challenging test that researchers have frequently used with tests of numeracy is the 3-item Cognitive Reflection Test (Frederick, 2005). Frederick (2005) designed the test to measure *cognitive reflection*, or resisting instinctive or intuitive responses. For instance, the first item is:

> *A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?*

The instinctive response to this item is 10 cents. However, to reach the correct response of 5 cents, the individual must reflect on her instinctive answer and correct it. Likewise, the instinctive responses to the other two items are usually wrong and must be reconsidered.

The theoretical rationale of the test is based in dual-process theory. For a correct response, the fast, intuitive System 1 needs to be ignored or inhibited, and thinking switched to the effortful, reflecting System 2 to deliberate and answer the item correctly (Frederick, 2005; Kahneman, 2011; Stanovich & West, 2000). Cognitive reflection is thus a metacognitive trait that describes the preference for reflecting on answers rather than providing intuitive responses (Welsh, Burns, & Delfabbro, 2013).

The Cognitive Reflection Test has stimulated several debates. The first debate is whether the test is merely another IQ test. Frederick (2005) has provided empirical evidence that the test predicts cognitive or decision-making biases better than IQ tests. The counterclaim is that it does not have the range of cognitive skills required under the Cattell-Horn-Carroll model, and so cannot be considered a better predictor of decision-making biases than more relevant intelligent tests (Welsh, Burns, & Delfabbro, 2013).

The second debate, and the most relevant to this thesis, is whether the Cognitive Reflection Test is merely another test of numeracy. There are various positions on the construct validity of the test. All positions hold that the test measures numeracy *to some extent*, as it correlates moderately to highly with tests of objective numeracy (Campitelli & Gerrans, 2014; Cokely et al., 2014; Weller et al., 2013). However, while some researchers view the test as measuring numeracy alone, others see it measuring numeracy as well as rational thinking or insight. Yet other researchers argue that the test measures numeracy ability, as well as a thinking disposition, such as searching for alternatives or actively open-minded thinking (Baron, 2008, as cited in Campitelli & Gerrans, 2014, p. 435; Cokely & Kelley, 2009; Frederick, 2005). It is also said to measure numeracy and insightful problem-solving ability, rather than cognitive reflection (Patel, Baker, & Scherer, 2019). It also has satisfactory internal consistency reliability considering that it has only three items, with reported alpha values of .6 to .74 (Campitelli & Gerrans, 2014).

The third debate is whether the familiarity of the test has reduced its validity. This concern derives from its online accessibility and popular usage (Stieger & Reips, 2016). However, the author of the test and colleagues (Meyer, Zhou, & Frederick, 2018) maintain that respondents who recall exposure are those who succeed on the test. Their evidence derives from of a

longitudinal study of 14,000 Amazon Mechanical Turk subjects whose performance on the test was tracked in a study from 2013 to 2015. Although these subjects took the Cognitive Reflection Test up to 25 times, the effect of multiple exposures on their performance was small. The percentage of respondents who scored each item correctly in their study was 28.3%, 55%, and 53% for the bat and ball, widget, and lake and lily pads items, respectively.

The correlations between the Lipkus, Berlin Numeracy and Cognitive Reflection tests are relatively high in numeracy studies: between the Cognitive Reflection and Berlin Numeracy tests $r = .56$; between the Lipkus and the Cognitive Reflection tests, $r = .41$; between the Lipkus and Berlin Numeracy tests $r = .5$ (Cokely et al., 2012). (Correlations are for the 4-item paper and pencil version of the Berlin Numeracy Test, as used for the thesis.)

There are other tests of numeracy that have been designed using item response theory to measure numeracy in health contexts, for example the Medicine Inventory form (NUMi) (Schapira et al., 2012). The items of this test are all framed in a medical context. In line with the intention of its developers, the test is used mostly in health contexts, and is not considered further in this thesis.

3.7.3.5. Synthesis of the Schwartz, Lipkus, Frederick and Cokely measurement scales

This brief review of the numeracy scales most frequently used in the risky choice literature has indicated that some scales do not discriminate at the lower or upper ends of numeracy ability levels. Thus, a synthesis of scales would be useful to accommodate a range of numeracy levels. Furthermore, the Cognitive Reflection Test (Frederick, 2005), although not intended as a test of numeracy, may complement tests of numeracy. There is thus a need to examine the factor structure of the underlying test items in combination, and to develop a composite scale of numeracy with the objective of predicting risky decision making. Although some authors have already undertaken analyses at the item level, no thorough analysis of the factor structure of all the items of the four tests (the Schwartz, Lipkus, Berlin Numeracy, and Cognitive Reflection tests) was found in the risky choice literature. The limited studies found that were conducted at the item level of the tests are now described.

Lipkus et al. (2001) conducted exploratory factor analysis on the responses of 463 adults to the three general items of Schwartz et al., and their additional eight items. They concluded that all 11 items of their scale were measuring the same construct, based on a single-factor solution. This result suggested that the medical framing of the additional eight items, as opposed to the non-specific framing of the original three items, made little difference to the

assessment of the numeracy construct.

Liberali et al. (2012) included the three items of the Cognitive Reflection Test, as well as the three items of the Schwartz scale and the other eight items of the Lipkus scale, in a factor analysis on the responses of 449 students. They derived four factors, with the cognitive reflection items and the three items of the Schwartz scale loading highly on the first factor which they named 'Verbatim/Monitoring'. The other items of the Lipkus scale loaded variously across the other factors, with a factor structure different from the structure found by Lipkus. However, the dimensionality of the results of Liberali et al. (2012) are questionable as the authors used an orthogonal varimax rotation on items measured on a categorical scale to derive factors that were expected to be correlated.

A more credible analysis at the item level was conducted by Weller et al. (2013). They used a second-generation method of test construction, known as the *Rasch* method of *item response theory*, to develop the Abbreviated Numeracy Scale. They selected items from the Schwartz, Lipkus and Cognitive Reflection Tests (and four items added by Peters, 2007, as cited in Cokely et al., 2014, p.20). Their resultant scale included 5 of the 11 Lipkus scale items, and 2 of the 3 Cognitive Reflection Test items. The scale is reported as having greater predictive power for risk perception than the individual tests do. The combination of items also had improved discrimination ability or sensitivity at the higher end of the numeracy scale, thus overcoming the negatively skewed distributions of numeracy scores characteristic of the Lipkus scale (Cokely et al., 2014). However, their analysis did not include the Berlin Numeracy Test items, which have been shown to have good discriminant validity and to measure additional variation in risk comprehension (Cokely et al., 2014).

This review illustrates the need for a synthesis of the items of the four measurement scales to achieve a single scale of numeracy that discriminates across a range of numeracy levels, with measurement invariance.

As this completes the literature review of decision making, numeracy and their association, the next section addresses the argument of the thesis, followed by the conceptual framework.

3.8. The theoretical framework of the thesis

The theoretical framework of the thesis is presented as a scientific argument based on the literature surveyed in Chapters 2 and 3. A scientific argument comprises a critical evaluation of the reasons and evidence associated with the proposition of a thesis (Van de Ven, 2007). The *Toulmin structure* of argument (Toulmin, 2003, as cited in Van de Ven, 2007, pp. 131-

135) is used here. The Toulmin structure comprises the following elements: (i) *background*, context or problem; (ii) the *claim*, hypothesis or proposition; (iii) *reasons*, major premise or logic of the claim; (iv) *evidence*, minor premise or backing data; (v) *qualifiers*, boundary conditions or assumptions; and (vi) *reservations* or limitations.

Figure 11 presents a graphic display of the theoretical framework of the thesis using the Toulmin structure. It presents the reasons and evidence associated with the proposition of the thesis that the statistical numeracy of decision makers under risk is a meaningful source of heterogeneity in the modelling of risky choices.

(i) The *background* to the proposition refers to risky choice models. The expected utility model sees risk preference as a primitive of the model, according to which individuals make risky choices (Dohmen et al., 2018). By definition, risky choices comprise objectively stated probabilities (Knight, 1921/2006). The model assumes that the decision maker understands the concepts of probability and expected utility. Her cognitive competence is assumed, as are her skill in Bayesian reasoning and statistical numeracy. While the rank-dependent and cumulative prospect theory models acknowledge distortions in probability weightings, they do not formally address the statistical ability of the decision maker under risk. Consequently, policy-focused policies that use these models are made without considering the role of statistical numeracy in risky choices.

Empirical research on risk preference and statistical numeracy presents challenges. Both constructs are latent traits, and therefore measured by their indicators. In the empirical risky choice literature, risk preference is often measured using choices between pairs of financial lotteries. These, and other elicitation methods, lack measurement invariance. Other challenges include the neglect of stochastic error models, compounded by individual variability, and other modelling criteria. The literature on numeracy does not provide a measurement instrument with clear factor structure, discrimination across a wide range of the statistical numeracy constructs, and assurance of measurement invariance. These challenges present some of the methodological difficulties underlying the research of the thesis on the role of statistical numeracy in structural models of risky choice. To the knowledge of the researcher, there is no study that has adequately examined the role of statistical numeracy in the three models of risky choice considered in the thesis, using a parametric modelling approach.

(ii) As previously stated, the *proposition* tested in the thesis is that the statistical numeracy of decision makers under risk is a meaningful source of heterogeneity in the modelling of risky choices.

(iii) The major *reasons* for the proposition are that probabilities are, by Knight's (1921) definition, a part of every risky choice. Further, inductive thinking is common to the latent variables of risky choice and probabilistic reasoning.

(iv) There is ample *evidence* in the empirical literatures on risky choice, cognitive psychology and decision making in health and finance, of an association between risky choice and statistical numeracy. In the domain of health, empirical studies provide the general finding of a positive relation between statistical numeracy and superior health-related decision making (Section 3.4.1). In the financial domain, individuals with superior numeracy tend to take informed risks and make superior investment decisions (Section 3.4.2). In laboratory studies using risky prospects, individuals with higher statistical numeracy levels tend to make normatively superior choices under risk (Section 3.4.3). There are numerous studies that have found a negative relation, although weak, between cognitive ability and risk aversion (Section 3.4.4). In these studies, superior decisions are evaluated either normatively or objectively.

(v) The proposition of the thesis is *qualified* as only applying to the normative model and to the rank-dependent and cumulative prospect models. The findings apply only to the *parametric* modelling of risky choice, using specific functional forms for the parameters. The thesis deals with risk – not uncertainty.

(vi) Finally, the *reservations* are several. It is acknowledged that the relation between risky choice and statistical numeracy is correlational, not causal. Further, other forms of numeracy, for example, subjective numeracy and financial numeracy, are not considered. It may be over-ambitious to obtain reliable and valid statistical numeracy and risky choice data within a limited questionnaire. The researcher also acknowledges that her skills and the respondent data are imperfect, as recognised in the post-positivist paradigm in which the research is set. As noted in the paragraph under 'background' in this section, risky choice research is challenged by model specification and issues of methodology. These reservations may limit the success of empirically testing the proposition of the thesis.

The theoretical framework is reappraised in the conclusions of the final chapter of the thesis (see Section 7.1).

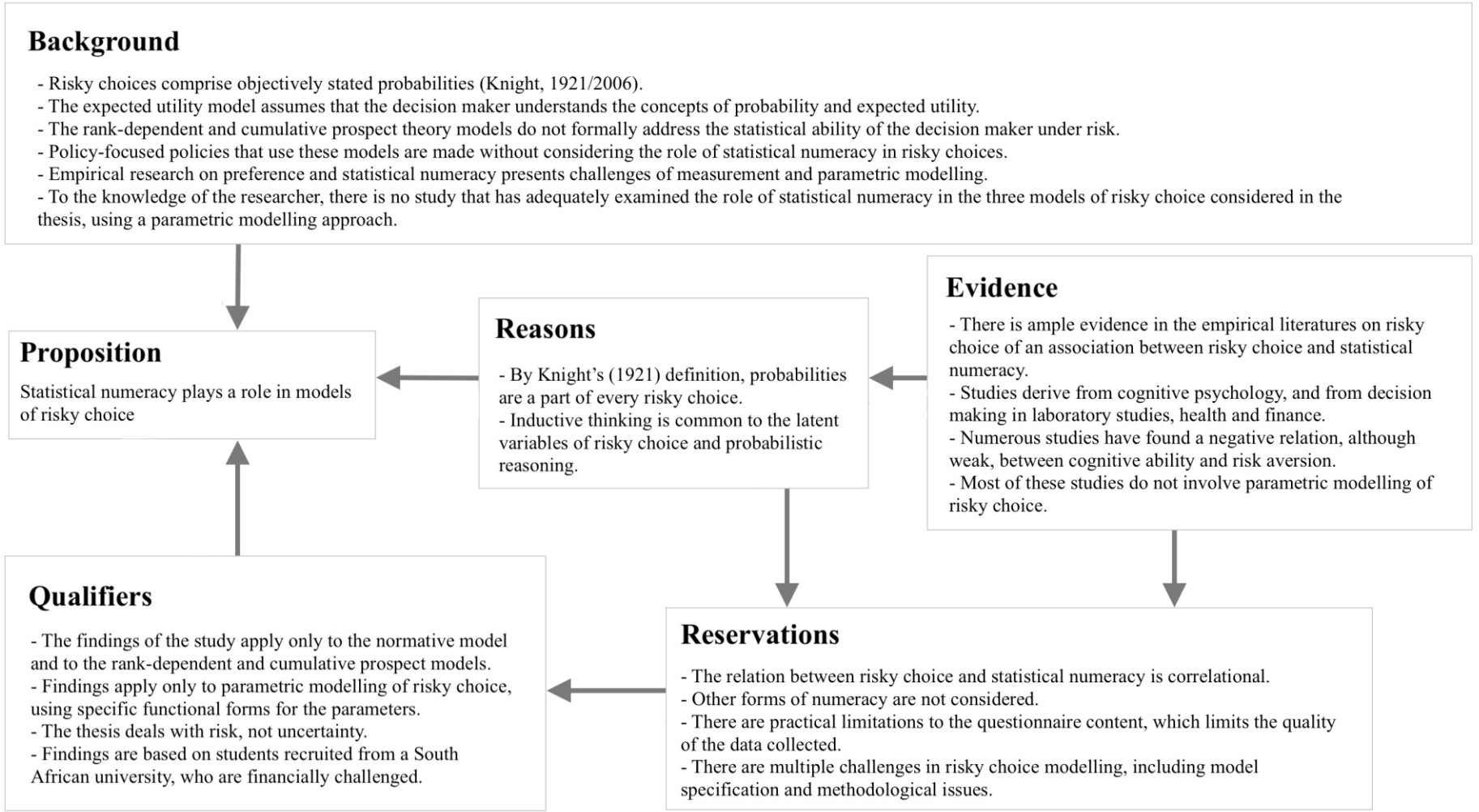


Figure 11: Theoretical framework of the thesis using the structure of a scientific argument
Note: Adapted from Toulmin (2003, as cited in Van de Ven, 2007, p. 130)

3.9. Conceptual framework of the research

As described by Wotela (2017), a conceptual framework is a summary of the literature review process that guides the research conceptualisation and the research methodology.

For the conceptual framework of the thesis, the research adopts the model of a *research project* proposed by Giere (1997, as cited in Van de Ven, 2007). Giere sees scientists as constructing models that are representative of the world. Giere's model of a research project comprises four components: (i) a real-world problem or object; (ii) a theoretical model of this problem or object; (iii) predictions, or similar, derived from the model; and (iv) empirical data from the real world. When these components are mapped onto the corners of a square or rectangle, and the relationships are depicted, the key aspects of the model are shown (Van de Ven, 2007, pp. 31-32).

The researcher presents the conceptual framework of the thesis using Giere's model (Figure 12). The relationships depicted in the figure represent the essential aspects of the literature that guided the study.

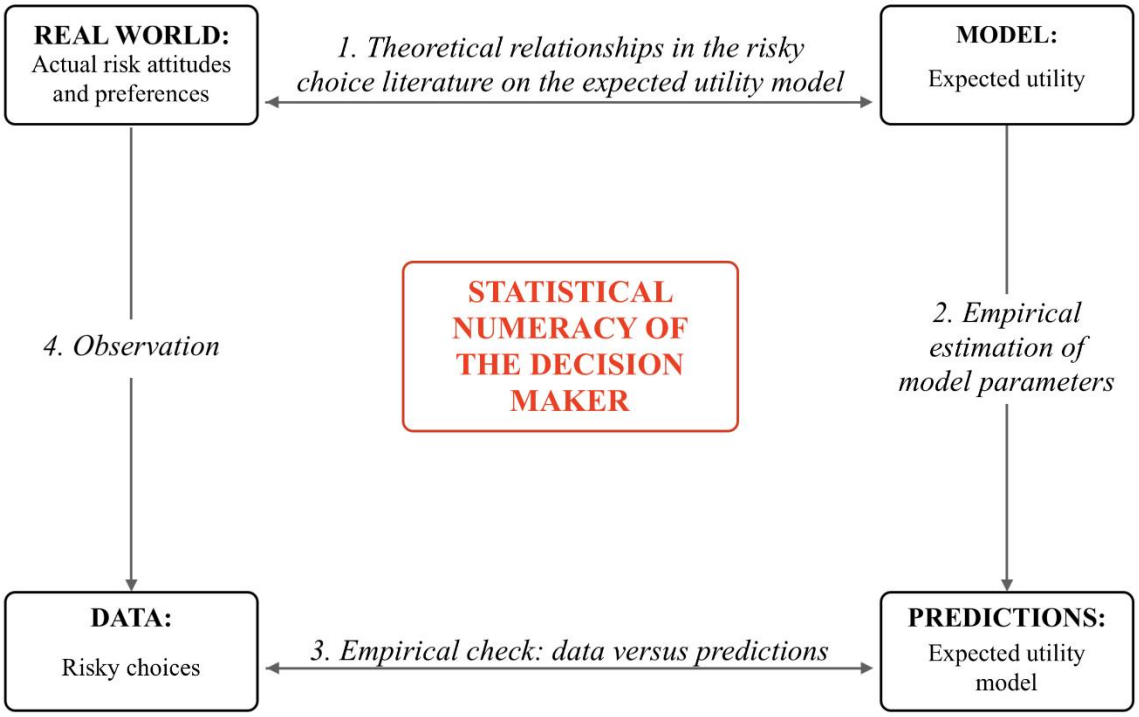


Figure 12: Conceptual framework: Relationships among constructs under the expected utility model based on Giere's (1997) model

Note: Adapted from Giere (1997, as cited in Van de Ven 2007, pp. 32-32)

Consider, first, the arrow from the top left to the top right of the figure, i.e. from the real world (the 'Real World' block) to the model (the 'Model' block), labeled as '1. Theoretical relationships in the risky choice literature on the expected utility model'. For the thesis, this arrow represents the gap between real, or observed, risky choices, and the predictions of the neoclassical economic model (expected utility).

As described in Section 2.1.5, the neoclassical model was formulated as a normative model. It represents a simplified theoretical construction of economic phenomena in a society. Through its abstract, general premises, it explains the risky behaviour of homo economicus, the economic man assumed to be fully informed, and to use the principle of maximisation to make rational decisions. The model is unconcerned with individual variability and psychological processes, and unconcerned with its correspondence with reality. Its "extreme generality makes it a universal principle" (Marinescu, 2016, p. 47).

As described in Sections 2.1.6 - 2.1.9, the gap between what the normative model proposes and the real world, i.e. the normative-descriptive gap, was the impetus for the development of the non-expected utility models of behavioural economics. The two non-expected utility models considered in the thesis are the rank-dependent and cumulative prospect theory models. Thus, the top side of Figure 12 represents the *theory* underlying the decision models of risky choice considered in the thesis.

Second, moving clockwise around the figure to its right side, the arrow labeled as, '2. Empirical estimation of model parameters', depicts the analysis of the empirical choice data under each of the models considered in the thesis, i.e. the estimation of the model parameters. Thus, the right-hand side of Giere's model, as applied in this thesis, represents the *analysis or estimation* component of the study.

Third, the relationship between the bottom-left and bottom-right blocks of Figure 12, labeled as, '3. Empirical check: data versus predictions', represents the relationship between the risky choice data collected (the 'Data' block) and the predictions of the model (the 'Predictions' block). This relationship represents the last stage of the analysis, i.e. the *checking or validation* of the modelling results relative to the observed risky choice data.

Finally, the left side of the figure, labelled as, '4. Observation', represents the relationship between the latent risky preferences of individuals, and the data that is collected, i.e. the risky choices. The strength of this relationship represents the quality of the theory and

methodology of the research. Thus, this side of the figure represents the literature reviewed on the *methodological and design aspects* of the thesis.

Finally, the middle block is inserted in the figure, for the purpose of the thesis, to show that statistical numeracy is regarded as the hub, or centre, of the relationships.

The conceptual framework is completed by the theory in Chapters 2 and 3, and by the guidance on methodology, analysis, and validation provided in Chapter 4.

This chapter ends with the formal statement of the thesis question, the two research questions and their sub-questions.

3.10. Thesis question, research questions and sub-questions

Thesis question: What is the role of statistical numeracy in computational models of risky choice?

Research question 1: What is the structure of the respondent groupings on numeracy?

1.1 What is the factor structure of numeracy?

1.2 Do the factors display measurement invariance?

1.3 What are the latent statistical numeracy subgroups?

Research question 2: What risk attitudes and probability distortions are associated with the risky choices of decision makers of different numeracy groups?

2.1: What risk attitudes are associated with the risky choices of decision makers of different numeracy groups?

2.2: To what extent do the estimated values of the parameters of the normative and descriptive risky choice models considered differ across numeracy groups?

2.3: To what extent do the estimates and predictions of the models hold for the risky choices of decision makers of different numeracy groups?

The following chapter provides details of the methods used in the thesis to answer these questions.

Chapter 4: Method

This chapter proceeds from the most general methodological aspects of the research approach (4.1) and design (4.2). It then describes the sampling method used and the characteristics of the samples (4.3), the measurement considerations of the thesis (4.4), and the construction of the measurement instrument of the thesis (4.5). Thereafter, it describes the procedural aspects of the study, such as the ethical considerations, the research incentive method used and payment protocol, and the online administration of the questionnaire (4.6). The chapter describes the methods used to analyse the statistical numeracy data (4.7), and the risky choice data (4.8). It concludes with a discussion of the reliability and validity of the research and its limitations (4.9).

4.1. Research approach

This section describes the overall approach used to investigate the role of statistical numeracy in mathematical models of risky choice.

The approach to the thesis was entirely quantitative, as opposed to qualitative or a mix of the two. It relied on detailed statistical analysis of the quantitative responses of a large sample of respondents to an online questionnaire designed to measure demographics, statistical numeracy and risky choices. The statistical results were based on a psychometric analysis and modelling of responses reflecting statistical numeracy, and on analyses of the risky choice data designed to estimate model parameters reflecting latent processes of risk preferences. At no stage were these processes measured qualitatively by self-reports or perceptions. This quantitative approach is consistent with the methods used in the economic and behavioural economics literatures on computational modelling of risky choice. It stands in contrast to the qualitative and mixed approaches used in studies on judgement which focus on the processes and heuristics used to make risky choices.

The numeracy and risky choices of the participants constituted the units of analysis of the main parts of this research. The units of analysis are defined as the entities being investigated in the research and the focus of the conclusions (Mouton, 2002).

4.2. Research design

The design, or strategy, used for this research is described in terms of its status as an experiment, and its position as online research. Each is now discussed.

4.2.1. Experimental, quasi-experimental and nonexperimental research designs

In the behavioural sciences, the categorisation of quantitative research design as experiments is relatively clear: those that involve the comparison of scores of individuals in groups are known as experimental, quasi-experimental or nonexperimental; quantitative research strategies that do not involve comparisons are descriptive or correlational (Gravetter & Forzano, 2012).

The three levels of experimental strategies represent a hierarchy of the extent to which the categorising variable can be said to *cause* differences in the measured variable (Gerber & Green, 2011; Gravetter & Forzano, 2012). In turn, the status of the independent variable, i.e. the categorising variable, is determined by whether the participants in the research are randomly assigned to the different levels of this variable (the experimental research strategy), or whether they already belong there with some attempt to control extraneous variables in the research (the quasi-experimental strategy), or with no attempt at experimental control (the nonexperimental research strategy). In the case of the experimental strategy, variables extraneous to the research are controlled by random assignment, and the *internal validity* is considered high, as between-group variability is ascribed to the categorising variable (Gravetter & Forzano, 2012).

This experimental/quasi-experimental/nonexperimental distinction is relevant to the thesis as the research is empirical and quantitative, and its main research question relies on the comparison of the risky choices of respondents in different statistical numeracy groups. As the participants in the study already belonged to their statistical numeracy groups by virtue of their numerical ability, the strict randomisation requirement of the experimental strategy does not apply. However, as described in Section 5.2.3.4, the researcher attempted to guard against threats to the internal validity of the research by considering extraneous variables, or covariates, such as education level, sex, age, and income in the numeracy categorisation. The research strategy is thus quasi-experimental, in the tradition of the *behavioral sciences*. This means that the researcher cannot attribute differences in the dependent variable, risky choice, *causally* to differences in statistical numeracy; she can merely propose or claim *a relation* between the independent variable of statistical numeracy and the dependent variable of risky choice (Gravetter & Forzano, 2012).

However, the definition of an experiment in the *economic sciences* is different. Randomisation is not a necessary condition for research to be called experimental (Gerber &

Green, 2011; Harrison & List, 2004). In the context of research on decision making, empirical work is referred to as *experimental*, whether conducted by statisticians, economists, psychologists or behavioural economists – see the experiments of Mosteller, Marschak and Nogee from the 1950s, Tversky, Kahneman, Edwards, Lichtenstein and Slovic from the 1970s, and the later risky choice studies, for example, the frequently cited study of Hey & Orme (1994) and the studies of Harrison and colleagues from the 1990s to date (Harrison & Ross, 2017). These are some examples of decision-making research termed experimental, but carried out without randomisation (Moscati, 2018). To quote Harrison, "The experimenter has controlled the stimuli, and that is all that is needed" (2013, p. 105). Following this naming tradition for research designs in the economic sciences, the decision-making component of this thesis would be referred to as experimental. This description is adopted in the thesis, as the research context of economics and behavioural economics is dominant over the psychological context.

4.2.2. Field versus laboratory research

There are further terminological differences between the criteria of the behavioural sciences and those of the economic sciences when categorising quantitative research as field or laboratory studies. Again, in the behavioural sciences, there seems to be a clearer distinction, although boundaries are still not absolutely defined and need to be viewed for each case (Gerber & Green, 2011). In the behavioural sciences, laboratory studies refer to research situations in artificial, controlled environments outside of real-world conditions, in which participants are fully aware of the research situation. By contrast, field studies refer to research situations in which participants are observed in a natural setting, or in a simulated environment that the participant perceives as a natural setting, and the participant may not be aware of their participation in the study (Gravetter & Forzano, 2012). By these descriptions, laboratory studies appear to describe most of the empirical research on risky choice, specifically the studies that involve participants choosing between pairs of lotteries in which participants are fully aware of their participation, and where the probabilities and outcomes of the choice tasks are explicitly stated.

Harrison and List (2004) and Harrison (2013) differentiate between laboratory and field experiments as used by experimental economists. As in the behavioural sciences, they define field research as research carried out in the natural environment as opposed to the laboratory or office-like environment, but include the type of sample and the nature of the task as criteria. In their 4-level taxonomy of field experiments, they define the first level as the

laboratory experiment characterised by "a standard subject pool of students, an abstract framing, and an imposed set of rules" (Harrison & List, 2004, p. 1014). This description is typical of the usual risky-choice experimental setting. It is usually student populations who are recruited to choose between risky lotteries, or perform an abstract task in an unfamiliar environment, such as a computer laboratory. The other levels cover nonstandard subject pools and extend to *natural field experiments* in which participants are unaware that they are in an experiment (Gerber & Green, 2011).

Based on the Harrison and List taxonomy of 2004, the present research would fit the description of a laboratory experiment as it involved university students recruited to participate in a study in which they were instructed to answer sets of numeracy and demographic questions, and to choose between risky lotteries. However, this research was administered online which represents an environment different from both laboratory and field environments (Buhrmester, Talaifar, & Gosling, 2018). The strategy of the present research is therefore termed *a risky-choice online experiment*.

4.2.3. Online research

Compared to field versus laboratory research, online research platforms represent a third way of conducting empirical research on large samples using self-administered research instruments. Online platforms, for example, *Amazon's Mechanical Turk (MTurk)*, provide access to demographically diverse, world-wide populations. Evaluation studies of online research results appear positive in terms of psychometric standards, efficiency and cost (Buhrmester, Kwang, & Gosling, 2011; Buhrmester et al., 2018; Horton, Rand, & Zeckhauser, 2011). Horton et al. (2011) have replicated the results of well-known studies using an online platform. Their online research findings replicated those of Tversky and Kahneman's (1981) on the cognitive bias of framing, whereby people reverse their responses to a problem dependent on whether it is framed positively or negatively (Kahneman, 2011). From a practical perspective, online research does not require the physical space and conditions to accommodate respondents in person and has been found to be more efficient to conduct (Horton et al., 2011).

The present study was conducted online, although it involved local students at the University of the Witwatersrand rather than a population of an online research platform. The population of the research, the sampling method and the sample characteristics are described in the following sections.

4.3. Population, sampling method and samples

This section describes the sampling aspects of the study in the context of the sampling approaches of similar types of studies to the thesis.

4.3.1. Population and sampling approach

Sampling in studies on parameter estimation and modelling is traditionally characterised by recruitment of students, in which students are attracted by a potential participation fee or prize and select themselves into the sample of respondents (Harrison & Rutström, 2008). Recruited samples of students featured frequently in the list of empirical studies on cumulative prospect theory modelling, as cited in Harrison (2019). Less often, the participants of research on modelling are members of the general population, and it is unusual to find examples of studies in this area that use rigorous sampling methods, although some do exist (Booij & Van de Kuilen, 2009; Harrison, Lau, & Rutström, 2007; Von Gaudecker, van Soest, & Wengstrom, 2011).

Recruited student samples may be described as a form of nonprobability convenience sampling, or availability sampling, from student populations. Such nonprobability samples cannot be said to represent their underlying populations, and so have poor external validity (Gravetter & Forzano, 2012).

The practice of offering subjects a fixed participation fee to encourage participation is common in experimental economics, and further undermines sample representativeness in terms of the distributions of risk attitudes found. In these samples, biased distributions of risk attitudes, such as risk aversion, arise as the level of the participation fee affects the type of student who volunteers to participate in the study. Experiments advertising a higher fixed participation fee attract samples of subjects who are more risk averse than when a lower fixed fee is advertised, as risk averse subjects require higher incentives in order to take a risk (Harrison & Rutström, 2008). By contrast, a relatively risk-seeking sample of students would be expected to select themselves into a study that offered no participation fee and uncertain payments or prizes. Thus, the distribution of risk types found in student samples of laboratory studies of risk are acknowledged to be determined by self-selection. The actual student population distribution of risk types that is unconditional on self-selection is probably unknown (Wilcox, 2008).

Self-selection of student respondents was used in the empirical research of the thesis. The researcher recruited a large sample from the database of undergraduate and postgraduate

students of the faculties of Humanities, Science, Health Science, Engineering and the Built Environment, and Commerce, Law and Management of the University of the Witwatersrand (Wits), South Africa. Accordingly, the sampling is described as a nonprobability availability sample of students from this university. The sampling thus suffers from the usual weaknesses of non-representativeness, or poor external validity, inherent in laboratory experiments on risky choices of students. However, external validity is neither the claim nor the concern of these studies.

The following section describes the various samples of the study.

4.3.2. Details of samples used

Preliminary sampling was used prior to recruiting students to the main phase of the study. The pilot sampling exercise is described in the next section, followed by details of the main samples used. As detailed in Section 4.5, two questionnaires were constructed for the research: one with gain-framed monetary prospects only (GAINS questionnaire), and the other with gains, losses and mixed monetary prospects (Gain/Loss/Mixed questionnaire or GLM questionnaire).

4.3.2.1. Pilot samples

Before finalising the questionnaires used in the main study, the researcher administered each questionnaire to a small sample of students of the Commerce, Law and Management faculty of the Wits School of Governance (WSG). One sample comprised 9 students of a quantitative analysis course, and the other comprised 15 students of a quantitative research methods course. The students of each group were randomly assigned to one or the other questionnaire, and all students received the same incentives for completing the questionnaires as in the main study. However, this preliminary exercise was administered in the computer laboratory of the school where the students and the researcher were all present. The students were asked to log into their student accounts, access a link to a questionnaire and complete the questionnaire while sitting in the WSG computer laboratory in the presence of the researcher. For both student groups, the session was held in the last 45 minutes of a 3½-hour contact session. Participation counted 5% towards the course mark. From the student perspective, the exercise was used as an example of carrying out quantitative research and was integrated with other course material. From the researcher's perspective, students could be observed completing the questionnaires, their dedicated time recorded to see how long the questionnaires took to complete, and any problems with the software recorded. The researcher could also assess the

difficulty levels of the numeracy tests included. She furthermore received feedback from the students on their experience in completing the questionnaires.

The feedback of the students was unanimously positive, indicating that they had all understood the instructions and required no further elaboration on what was expected from them. Most students said that they had enjoyed the experience. The questionnaire with prospects framed as gains only had 40 risky choice pairs; the questionnaire with prospects with gains, losses and mixed prospects had 48 risky choice pairs. In addition, both questionnaires included 12 numeracy items and 7 demographic items. No student took under 20 minutes to complete either questionnaire, and most students completed the questionnaires within 25 minutes. One student took 40 minutes to complete the questionnaire with the gain prospects. He reported having taken extra care before responding to each item. The researcher calculated the payoffs on a randomly selected risky-choice response of each student, and paid the students in the following contact session of the course.

Based on the feedback from these pilot sessions, the researcher added a further eight risky choice items to the 40-item gains questionnaire and made small changes to the flow of the sections within the questionnaires. The pilot participants were not included in the main sample of the study.

4.3.2.2. Samples of the main study

The recruitment of students for the main sample followed a formal university process that stipulated that all communication to students and staff had to be sent from the office of the registrar (see Section 4.6.1 on ethical procedures). Accordingly, a link to one of the two questionnaires was forwarded through the University's official process to approximately 13,000 postgraduate students. Based on the number of students in each faculty and year, groups of approximately equal size were set to receive one of the links to the two questionnaires at random. However, after 48 hours, only 50 responses to the Gains questionnaire and 122 responses to the GLM questionnaire had been received. The low response rate was indicative either of a problem in the process of sending out the questionnaire links, or an extremely low response rate from the students. Within a week, the recruitment process was repeated, but this time the questionnaire links were sent to both undergraduate and postgraduate students (approximately 25,000 and 13,000, respectively), as the researcher feared a very low response rate. However, the few days between sending out the questionnaires again allowed the researcher to examine the responses and reflect on the

questionnaires, with the result that some changes were made to the questionnaires before the links were sent out the second time (see Section 4.5.2 for details). The second attempt was successful. Within 24 hours, 421 responses had been received to the questionnaire with gain-framed lotteries (GAINS2), and 690 responses were received to the GLM questionnaire (GLM2). At this point, the researcher deactivated the links as the potential payout to the respondents was above her budget.

The researcher then began the process of cleaning the data. She used three criteria for retaining responses. First, the respondent had to have taken at least 10 minutes before submitting the questionnaire, this cut-off time being the quickest feasible time estimated to complete the questionnaire diligently, rather than merely clicking through to the end. Second, the respondent had to have responded to at least 80% of the items of each of the sections of the questionnaire. Third, responses to at least 10% of the items within a section needed to differ from the rest to guard against the response set of merely clicking the same response within a section.

Applying these criteria resulted in a smaller, but much cleaner, set of data. It was found that 152 of the 1283 respondents from both recruitment attempts (11.8%), had clicked on the link but not started, a further 189 (14.7%) had taken less than the minimum time of 10 minutes, and a further 276 (21.5%) had not reached the end of the questionnaire or had little or no variance in their responses. There remained 666 usable responses (51.9% of the original 1283 responses, of which about 15% of these responses had missing values on one or two items. The missing values comprised less than 1% of the data. These values were imputed based on the modal response of that individual to the items within that section of the questionnaire, but meant that these respondents were not eligible for payout, as each section of the questionnaire stated that fully completing the questionnaire was a necessary condition for payout. Before submitting the questionnaire, the respondent had been given a final chance to return and complete any omitted items. The sample sizes of the original and retained respondent groups are provided in Table 1.

According to the definition of a ‘unit of analysis’ as the object under study (Mouton, 2002), the unit of analysis of this research was considered as the 666 respondents in most sections (see Section 5.3 for exception). Throughout the tables of the thesis, percentages that do not sum to 100% reflect rounding error.

Table 1: Sample sizes of respondent groups

Recruitment attempt	Questionnaire	Responses received	Responses retained
1	GAINS1	50	29
1	GLM1	122	98
2	GAINS2	421	240
2	GLM2	690	299
	Total	**1283**	**666**

4.3.2.3. Description of the respondents in the main study

The demographic variables of the 666 respondents are supplied in Table 2. These variables were useful both for describing the respondents, and for serving as covariates in later statistical analyses (see Section 5.2.3.4) (Edwards, Lindman, & Savage, 1963).

Of the total sample of 666 respondents (Table 2), slightly over half were aged 18-21 years (51%), 40% aged 22-30, and 10% older. A small majority (55%) of all respondents had matriculation as their highest qualification, and 57% were female. Almost all respondents (92%) live with others in their households. Approximately 42% indicated their total monthly household income as less than R10,000.

The demographic variables that were considered in later analyses as covariates are: age, highest education qualification completed, sex and household income. There is a strong relation between age and qualification ($\chi^2(2) = 352.97$) with strong effect size (contingency coefficient $C = 0.59$). This result is expected as older students are more qualified. Furthermore, the relation between qualification and household income is significant ($\chi^2(3) = 9.66$, $p < .05$), but with weak effect size (contingency coefficient $C = 0.12$). This relation is likewise expected as more qualified students are likely to earn more.

Table 2: Demographics of respondents in the numeracy groupings

	Demographic	Total (n = 666)
Age	18-21	51%
	22-25	28%
	26-30	12%
	31-40	8%
	41-50	2%
	51 or older	0%
Highest qualification completed	Matriculation	55%
	Undergraduate certificate/diploma	3%
	Bachelors degree	16%
	Honours degree/technology degree	12%
	Postgraduate certificate/diploma	2%
	Masters degree	11%
	Doctorate/post doctorate	1%
Faculty	Science	45%
	Humanities	32%
	Health Sciences	12%
	Engineering and the Built Environment	5%
	Commerce, Law and Management	6%
Your sex?	Male	42%
	Female	57%
	Prefer not to say/other	1%
Total monthly household income (before tax)	R5,000 or less	28%
	R5,001-R10,000	14%
	R10,001-R35,000	27%
	R35,001-R50,000	12%
	R50,001-R100,000	11%
	more than R100,000	8%
People in the household	1	8%
	2-3	29%
	4-6	51%
	7+	12%
	Total	**100%**

4.4. Measurement considerations in the thesis

When measuring an individual's risky choices, it is essential that measurement instruments be carefully constructed to avoid the situation of wrongly ascribing anomalies that are instead due to poor measurement, irrational choices and biases on the part of the decision maker (Wakker, 2010). This section presents the background to the construction and application of the measurement instrument, focusing on the complex measurement of risk attitudes. It considers the practical restrictions of administering the measurement instrument, the options available for measuring risk attitudes and constructing the elicitation method for measuring risky choice, and the type of incentives that could be used in terms of their salience and application.

4.4.1. Practical constraints

The research required the measurement of statistical numeracy and risky choices, as well as a few demographic variables. The measurement format had several restrictions. First, it needed to facilitate online measurement on a single occasion. This restriction ruled out the format used in several laboratory-type risky choice experiments in which large amounts of data were collected from participants over several occasions. The researcher did not consider it feasible to ask the respondents to return to answer a second online questionnaire on another occasion as she feared a high attrition rate.

The questionnaire had additional practical constraints. It needed to be of reasonable length, despite the requirements of it having to include an explanation of the research, the payment protocol, the participants' rights, to provide instructions on the respondents' tasks, as well as to measure the respondents' demographic details, their statistical numeracy and their risky choices, all with adequate power. The statistical numeracy items needed to range in difficulty level and to include the items of the Cognitive Reflection Test (Frederick, 2005). The risky choice items needed to include items in the gain, loss and mixed frames to test the risk attitudes posited in cumulative prospect theory (Tversky & Kahneman, 1992). The researcher aimed to devise a questionnaire that could achieve this content within 20 minutes, although there was no cut-off time imposed on respondents completing the questionnaire.

The measurement of the demographics was straightforward, and the measurement of statistical numeracy involved administering existing tests; the measurement of risky decision making was more challenging, and involved a review of the literature on methods of eliciting risk attitudes.

4.4.2. Risk attitude elicitation methods

In the context of risky choice research, *elicitation methods* refer to how researchers obtain experimental data on risky choice. Researchers have compared and correlated the results of elicitation methods, and carried out meta-analyses on studies that performed cross-method comparisons (Charness, Gneezy, & Imas, 2013; Harrison & Rutström, 2008; Loomes & Pogrebna, 2014; Pedroni et al., 2017a; Zhou & Hey, 2018). In general, these studies provide damning evidence of the lack of invariance of elicitation methods. Different methods indicate inconsistent risk attitudes, different proportions of individuals as risk averse versus risk seeking, and inconsistent rankings of individuals on risk aversion across elicitation methods. In a study on the risk preferences of 1507 adults, with other confounds corrected, the effect of statistical numeracy, measured by the *Berlin Numeracy Test*, was not found to be an explanatory variable for the measurement inconsistencies (Pedroni et al., 2017a, 2017b). The authors concluded that despite all experimental and statistical controls, risk preferences may be constructed by elicitation methods that reflect different cognitive processes. Similarly, in a meta-analysis of research that compared various elicitation methods, Zhou & Hey (2018) found substantial variation in the estimated parameters across elicitation methods.

There are over 20 elicitation methods, each falling into one of four main types: pairwise choices, multi-item choices, continuous choice and rankings (Zhou & Hey, 2018). Pairwise choices are commonly used in the economics and behavioural economics literatures, and so, are examined further for use in this thesis. Within this method are two further methods: the *multiple price list* method, and the method of presenting *individual prospect pairs*, typically in random order, without the use of a list. Each is considered further.

4.4.2.1. The multiple price list method

The multiple price list method requires the subject to choose between two options, presented in a list comprising a sequence of prospect pairs. Within the list, the pairs are ordered either on expected payout value, and/or on risk. The point at which the subject switches from choosing the riskier lottery of a pair to the safer is indicative of her risk aversion. There are different versions of the method. For example, the exercise may be repeated multiple times with different lists, converging on the switch point (Harrison & Rutström, 2009; Harrison & Swarthout, 2016; Loomes & Sugden, 1998; Stott, 2006). The method is associated with frequently cited studies of risky choice in the laboratory, such as the studies of Tversky and Kahneman (1992), and Holt and Laury (2002, 2005). A variation of the method was used in

the influential paper of Gonzalez and Wu (1999). However, the method is time consuming. For example, Tversky and Kahneman's subjects completed 56 choice tasks in three 1-hour sessions over a period of several days, Holt and Laury's experiment comprised a large number of choices, and the experiment of Gonzalez and Wu involved 165 gambles framed as gains only, and required four 1-hour sessions. Moreover, for the method to be used in this thesis, several lists and a large number of choices would have been required to test risk attitudes over a range of risk aversion levels in the gain, loss and mixed frames. Under the onerous time constraints of the questionnaire, the alternative method of presenting individual prospect pairs was adopted.

4.4.2.2. Presenting pairwise choices between prospects

The pairwise choice method of presenting multiple choice pairs individually, rather than within lists, offered an alternative measurement approach for the thesis. The method is referred to as the *random lottery pairs* method when the pairs are presented in a random order, as in the frequently cited study of Hey & Orme (1994), who presented their 80 subjects with 100 pairwise choices over two occasions, approximately a week apart.

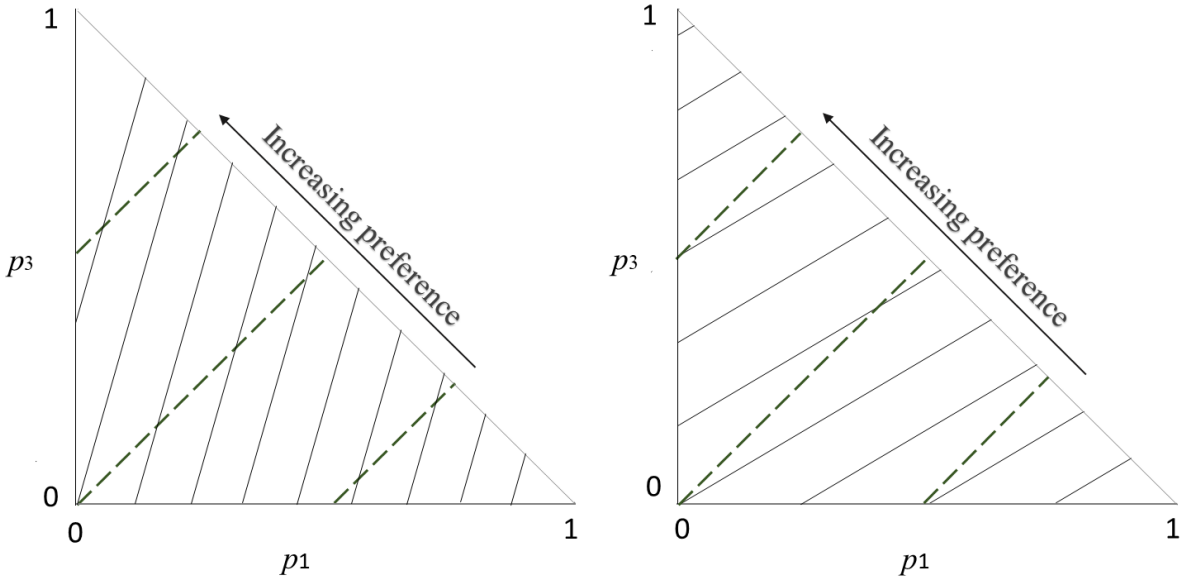
The challenge of this method is to design a set of choice pairs to ensure they represent a gradient of risk attitudes with the potential of identifying decision makers of various risk aversion levels. To do so requires careful selection of the probabilities of the pairwise choices. For example, if the options within a decision pair share the same set of outcomes and have probabilities that are similar, respondents will likely be indifferent to the choice, and so this pair would lack discriminating power. Machina (1987) developed a method that facilitates the selection of decision pairs across a gradient of risk attitudes. The method is known as the *Marschak-Machina triangle*. It was used recently by Harrison and Swarthout (2016) for their selection of random lottery pairs, and has been used by other researchers (Loomes & Sugden, 1998; Xiaoxue, 2017).

The Marschak-Machina triangle

The Marschak-Machina triangle is a graphical method of displaying the probabilities of two competing prospects on the same indifference curve, each with up to three outcomes x_1, x_2 and x_3, with x_1 being the lowest value outcome, x_2 the middle outcome and x_3 the highest outcome ($x_1 < x_2 < x_3$). The method uses a right-angled triangle in which each lottery is plotted as a dot. Within the triangle, the Y axis represents the probability (p_3) of the best outcome of a lottery (x_3), and the X axis represents the probability (p_1) of the worst outcome

(x_1). The probability of the outcome with the middle value is implied. When a prospect lies on the diagonal of the triangle, the probability (p_2) of its middle value (x_2) is 0. Such a prospect has extreme outcomes.

When the two dots of a lottery pair are connected by a straight line, the connecting line represents a *linear indifference curve*, also known as an *iso-expected utility* line, representing the decision maker's preference. Under expected utility theory, indifference curves are linear in the Marschak-Machina triangle. The line represents lottery pairs between which a person is indifferent. Moving along an indifference curve in a north-east direction (towards the diagonal) represents prospects with the same utility but becoming increasingly risky as the probability of the outcome with the middle value decreases. Thus, the riskier lottery of a pair is positioned to the right on the line. An illustration of the triangle for risk averse and risk seeking preferences are shown in the left and right panes of Figure 13, respectively (Machina, 1987, p. 538). Higher parallel indifference curves (moving in a north-west direction within the curve) represent prospects of constant – but higher – utilities, i.e. prospects that are valued more.



Relatively steep indifferences curves of a risk averse individual

Relatively flat indifferences curves of a risk preferring individual

Figure 13: Indifference curves in the Marschak-Machina triangle
Note: Reproduced from Machina (1987, p. 538)

Indifference curves with *steeper slopes* (greater than 45° to the X axis) represent higher risk aversion. The slope of an indifference curve represents the difference between the probabilities of the riskier and safer options' highest outcome, relative to a unit change on the X axis. Correspondingly, with increasing slope, the ratio of the expected values of the riskier to the safer option also increases. Thus, indifference curves with steeper slopes represent choices in which the riskier prospect becomes more attractive, relative to the safer prospect. A respondent who is indifferent to a choice represented by a steep indifference curve, and still selects the safer option to the left of the pair, is highly risk averse. By contrast, curves with slopes with a value of 1 (45° to the X axis) indicate risk neutrality, and those with flatter slopes (less than 45° to the X axis) indicate risk seeking. The flatter slopes represent prospect pair comparisons with relatively small differences between the probabilities of the riskier and safer options' highest outcome, relative to a unit change on the X axis.

In summary, the steeper the slope of the linear indifference curve, the greater the curvature of the utility curve, i.e. the more concave the curve, and the greater the risk aversion. Thus, by selecting lottery pairs represented over a range of indifference curve slopes, a set of lottery pairs can be generated to represent a range of risk aversion levels under expected utility theory. This method was used to select the prospect pairs for this thesis, modelled on the studies of Loomes and Sugden (1998) and Harrison and Swarthout (2016).

The following section considers aspects of incentivising the respondent to participate in the experiment.

4.4.3. Incentives

The aspects of incentives considered here include their salience, the manner in which they are administered and the compatibility thereof with structural models of risky choice.

4.4.3.1. Real versus hypothetical and salient versus non-salient incentives

In the context of most risky choice experiments, incentives are defined as the financial payouts that participants may receive. The issue of incentives for participation in experiments is viewed differently in psychology and other behavioural sciences from the economic sciences. While financial incentives are rare in psychological experiments, they are considered the norm in experimental economics (Camerer & Hogarth, 1999; Hertwig & Ortmann, 2003).

In experimental economics, choices are termed *real* if there are real consequences for the respondent dependent on her choice, for example, a choice in which she may receive the outcome of the prospect she chooses. By contrast, in hypothetical choices, such as in surveys with hypothetical scenarios, there are no real consequences for her choice. Salient rewards change with the performance of the choice task, whereas non-salient rewards do not, for example, in experiments with a show-up fee. In economic theory, real rather than hypothetical choices elicit more honest responses. Salient incentives are also more likely to produce an honest reflection of the respondent's preference than non-salient incentives (Harrison, 2014). Real and salient incentives are, therefore, an important aspect of empirical risky choice research. The researcher thus adopted incentives that were real and salient for South African students.

4.4.3.2. Incentive compatibility

Several researchers in risky choice reject the use of the *random lottery incentive* mechanism, also known as the *random problem selection* mechanism, when assuming a non-expected utility structural model, such as rank-dependent utility theory or cumulative prospect theory (Andersen et al., 2018; Cox, Sadiraj, & Schmidt, 2015; Harrison & Swarthout, 2014). They argue that the random lottery incentive method is *noncompatible* with non-expected utility models, such as rank-dependent utility theory and cumulative prospect theory. A noncompatible incentive is defined as an incentive mechanism or manner of payoff that would create distortions in the choices subjects make when facing multiple decision tasks (Azrieli, Chambers, & Healy, 2018). The reason this problem arises is that non-expected utility theories relax the independence assumption. By contrast, the random lottery incentive assumes that individuals choose from lottery pairs independently, by isolating each pairwise lottery from others, a condition that implies the compound independence axiom. Thus, according to this argument, non-expected utility models that do not assume the independence axiom cannot be estimated without bias using data generated by the random lottery incentive method.

However, other researchers disagree. Azrieli et al. (2018) maintain that the experimenters who claim empirical evidence of independent choice behaviour are assuming some form of reduction process in the choices. The choice behaviours would only imply independence correctly if they involved reduction, in other words, reduction of compound lotteries to simple lotteries. Furthermore, as simple and compound lotteries are evaluated using different mechanisms, support is lacking for the use of the compound independence axiom through

rank-dependent utility (Hajimoladarvish, 2018). Accordingly, Azrieli and colleagues refute the independence assumption for the random incentive lottery method, and argue for the compatibility of the random lottery incentive with the non-expected utility theories that do not assume independence. As reviewed by Dhami (2016, p. 18), other researchers, such as the economists Bardsley, Cubitt, Loomes, Moffat, Starmer, and Sugden (2010), claim no empirical evidence of bias in the results of the random lottery incentive mechanism, despite the rejection of the independence algorithm. Moreover, researchers, for example, Harrison and Swarthout (2016), use the random incentive mechanism pragmatically, while acknowledging its possible shortcomings.

Finally, as the simple lotteries of the questionnaires used in the research of this thesis are unlikely to involve a reduction process, the researcher used this incentive method in motivating her respondents.

4.4.4. Conclusion: The approach to the measurement instrument of the thesis

Based on the considerations outlined on elicitation and incentive methods, the researcher decided to use the pairwise choice method of presenting random lottery pairs, the random lottery incentive mechanism, and the Marschak-Machina triangle for lottery selection, following Loomes and Sugden (1998) and Harrison and Swarthout (2016).

Owing to the required content and length restrictions of the questionnaire, she decided to devise *two* questionnaires with the same content other than for the risky choice items, and to assign about half the sample respondents randomly to each. The one questionnaire measured risky preference on 48 pairwise choices of prospect pairs framed only as gains, while the other questionnaire measured risky preference on three sets of 16 pairwise choices between prospects framed in the gain, loss and mixed contexts, respectively. The questionnaire with gain-framed prospects only (hereafter referred to as GAINS) was designed to include a more complete set of probabilities and choice pairs, while the other questionnaire with gain-, loss- and mixed-framed prospects (GLM) was necessarily more limited. The drawback of the dual questionnaire solution was that separate analyses had to be conducted on the data from the individual questionnaires when respondent-level analyses were required, as opposed to choice-level analyses.

The next section describes the structure and content of the questionnaires.

4.5. Structure of the measurement instrument of the thesis

Following the initial invitation to participate in the study, the first screen of the online questionnaire comprised participant information on the research and the rights of the respondent. The information was presented in line with the *Wits Guidelines for Human Research Ethics Clearance Application (Non-Medical)* (see Section 4.6.1 on ethics). Thereafter, the questionnaires comprised three sections: a brief demographic section, the risky choice section, and the statistical numeracy section. The questionnaire concluded with an optional explanation of the payout mechanism and protocol.

4.5.1. The demographic content of the questionnaires

The first demographic detail requested was the student's email address, accompanied by an explanation that it was required for contacting the respondent if she won a prize. This was the sole reason for the question:

> "To start off, I ask a few questions about you. I ask for your Wits email address for contacting you if you win a prize after completing the Questionnaire."

The other demographic items of the questionnaires were age, highest qualification completed, faculty/faculties of study, sex, total monthly household income (pre-tax), and number of people living in the household. The questionnaires included four items on subjective numeracy which were not analysed for the purposes of the thesis.

4.5.2. The risky choice content of the GAINS and GLM questionnaires

Harrison and Rutström (2009) recommend 60 prospect pairs for reliable parameter estimates in the positive, negative and mixed frames, while not overtaxing the participant. However, as previously outlined, due to the severe time and content restrictions of the questionnaires, the researcher was restricted to 48 prospect pairs in each questionnaire.

For the questionnaire in the gain frame, the researcher modelled her prospects on the prospect pairs of Loomes and Sugden (1998), who used gain-framed prospects only. They used seven Marschak-Machina triangles, six of which were designed with indifference curves of increasing slopes (with gradients of ⅔, 1, 1½, 2, 3 and 4). Their seventh triangle was dedicated to examining dominance and the Allais paradox. (The terms 'slope' and 'gradient' are used synonymously.)

For the thesis, the questionnaire with gain-framed prospects used for the first recruitment of participants (GAINS1) comprised the same prospect pairs as used by Loomes and Sugden

(1998) in their first five triangles, with eight prospect pairs per triangle. The outcomes used for the thesis were R0, R100 and R200 for this set of prospects. The details of the GAINS1 prospects are presented in Table E 1.

The questionnaire with gain-framed prospects used for the second recruitment of participants (GAINS2) was changed in two major ways. First, it included a second context. The researcher made this decision based on the study of Hey and Orme (1994). These authors used four contexts, or combinations of outcomes, each with the same pairs of probability distributions, a design feature praised by Wilcox (2011). Accordingly, she used four prospect pairs from each of the six Marschak-Machina triangles of Loomes and Sugden (1998) for outcomes of R0, R100 and R200, and then repeated these 24 probability combinations for a second context of R0, R100 and R300. Thus, the GAINS2 questionnaire comprised 48 lottery pairs with the same probability distribution over two contexts.

Second, the prospect pairs of each triangle of GAINS2 were designed with the following properties to control extraneous prospect effects as far as possible:

(i) All the pairs within a triangle were plotted with the same distance apart, thereby representing pairs with the same risk differences.

(ii) All four triangles had the same number of *bottom-edge* comparisons. Bottom-edge comparisons have two features: the best outcome of the riskier prospect has a non-zero probability, while the best outcome of the safe prospect has a zero probability; and the probability of receiving the middle outcome of the riskier prospect is only slightly less than the probability of receiving the middle outcome of the safer prospect. In these comparisons, the riskier prospect is chosen more often than in other choices within the same triangle (Loomes & Sugden, 1998).

(iii) Every triangle had three prospect pairs in which the riskier prospect was extreme, i.e. the probability of receiving the middle outcome of the riskier prospect was 0.

(iv) Every triangle had prospect pairs representing high, middle and low attractiveness.

(v) Every triangle included one pair with a safe option (probability of middle outcome = 1).

The details of the GAINS2 prospects are presented in Table E 2, and presented graphically in the top pane of Figure 14.

For the questionnaire with prospects in the gain, loss and mixed frames for the first recruitment (GLM1), the researcher modelled her prospects on those of Harrison and

Swarthout (2016). Harrison and Swarthout had also modelled their study on the prospect selection of Loomes and Sugden (1998), but introduced prospect pairs in loss and mixed frames. For their gain-framed prospects, they used eight prospect pairs from each of the first five triangles of Loomes and Sugden (with gradients of ⅔, 1, 1½, 2, and a combination of 3 and 4), and so had 40 prospect pairs in common. They then selected the eight prospect pairs from the second and fifth triangles (with gradients 1 and 3) to construct 16 loss-framed prospect pairs, and 16 mixed-framed pairs. They rationalised these choices in terms of the information these triangles would provide over others. They repeated 24 of their gain-framed prospect pairs using a second set of outcomes, and included 4 prospect pairs on the Allais paradox. Additionally, they used real, salient incentives, with net prizes for their 177 undergraduate student participants ranging from $0-$70, and from $0-$750 for their 94 MBA students. Their 100 lottery pairs were selected with moderate-to-high power for optimal comparison of the structural theories they considered (Harrison & Swarthout, 2016).

For GLM1, the researcher used 16 prospect pairs for each context, eight from each of the second and fifth triangles of Loomes and Sugden (with gradients of 1 and 3). This selection was guided by the choice of Harrison and Swarthout (2016), who had used 16 loss-framed prospect pairs and 16 mixed-framed pairs from these triangles. Thus, the three sets of 16 pairs had the same probability distributions, i.e. were reflections of each other, based on eight prospect pairs from the first two and the last two of the Marschak-Machina triangles of Loomes and Sugden (1998). The researcher used an upfront ‘endowment’ of R200 for loss-framed prospects, and R100 for prospects in the mixed frame, so that the respondent could not lose money. The details of the GLM1 prospects are presented in Table E 3.

For the second administration of the questionnaire with prospects in all three frames (GLM2), the researcher constructed 16 prospect pairs, with the aim of achieving more power over the range of risk aversion levels, by using four prospect pairs from four triangles (rather than eight pairs from two triangles). As for GAINS2, the prospect pairs of each triangle had the same properties [(i) – (v)]. The first two and the last two triangles of Loomes and Sugden were used (with gradients of ⅔, 1, 3 and 4). Thus, the 3 sets of 16 pairs had the same probability distributions. The details of the GLM2 prospects are presented in Table E 4, and presented graphically in the bottom pane of Figure 14.

The newly constructed prospect pairs added to those used in previous studies are indicated by an asterisk (*) in Table E 1 - Table E 4.

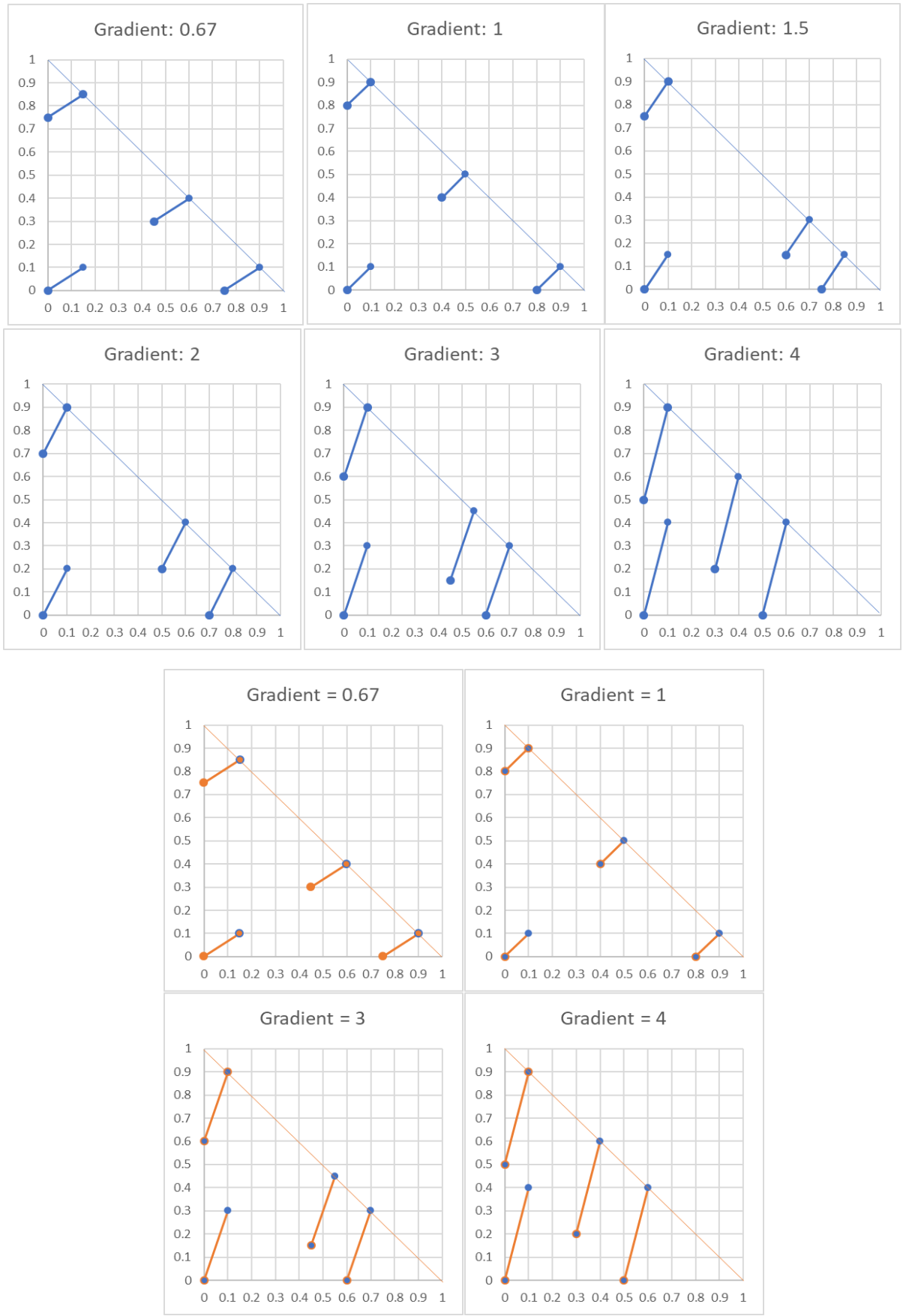


Figure 14: GAINS2 questionnaire with six Marschak-Machina triangles (upper pane) and GLM2 questionnaire with four Marschak-Machina triangles (lower pane)

4.5.3. The statistical numeracy content of the questionnaires

To measure statistical numeracy, the researcher used a combination of the prominent numeracy scales published in empirical studies on numeracy in the context of risky choice. Four scales were considered for the thesis. The first two scales considered were the 3-item scale of Schwartz and colleagues (2011) and its 8-item extension (Lipkus et al., 2001), designed to measure simple mathematical operations on risk magnitudes and conversions of proportions, percentages and probabilities. In view of the extreme skewness of the Lipkus scale in many studies (see Section 3.7.3), four items were selected from the scale. These comprised the original three items of Schwartz et al. (1997) and one additional item from the items added by Lipkus et al. (2001). These four items are referred to from here on as the Schwartz/Lipkus scale and, furthermore, as a single scale as the four items are all part of the 11 items of the Lipkus scale. The third scale used was the 4-item Berlin Numeracy Test (Cokely et al., 2012) designed to measure educated respondents' risk literacy – making good decisions based on information about risk (Cokely et al., 2014). The fourth scale used was the 3-item Cognitive Reflection Test designed to measure cognitive reflection – resisting the first response that comes to mind (Frederick, 2005).

As discussed in Section 3.7.3, the correlations between these tests are moderately high in numeracy studies, correlations ranging between .41 and .56. However, these correlations are not high enough to render any one of the tests redundant. Furthermore, the Schwartz/Lipkus scale was expected to discriminate among respondents of lower numeracy levels, and the Berlin Numeracy and Cognitive Reflection tests were expected to discriminate among respondents of higher numeracy levels. Thus, all these measurement scales were included in the questionnaire. Items that are worded in the original tests in dollars were reworded in Rands, the South African currency. Other minor adjustments made to the item wording are shown in Table F 1. The analysis of the items of the scales, their factor structure and amalgamation into a single measurement instrument with known factor structure and measurement invariance is discussed in Section 5.2.3.

4.6. Procedure used to obtain data

This section presents the procedural issues of the thesis: the ethics process followed, the online administration of the questionnaire and the payment protocol for participating in the research.

4.6.1. Ethics

The researcher followed the *Guidelines for Human Research Ethics Clearance Application (Non-Medical)* of the University of the Witwatersrand, compiled to ensure that human participants are treated fairly, and ethical standards are met. These guidelines require that there be full disclosure of the purpose of the research to participants, no coercion of participants, and that participants are aware of their rights to privacy. The application for this research was approved by the main University Human Research Ethics Committee (Non-Medical) prior to the research (Figure B 1).

As the research instrument was online, this information was conveyed to potential respondents via an email from the Wits University registrar's office. As described in Section 4.3, the deputy registrar sent out invitation emails to students of the university. There were two versions of the email sent out, each with a link to either the GAINS questionnaire or the GLM questionnaire. The student invitations were assigned randomly to one of the two questionnaires.

The invitation emails were worded as follows:

> **Subject:**
> You're invited to participate in my research
>
> **Message body:**
> Dear Student
> I am collecting data for my PhD research on risky decision making and would appreciate your help. By completing my survey, you will receive a payout of up to R200, based on a random selection of one of your choices. The survey is confidential and takes 20-30 minutes. Kindly note that you may only complete the survey once.
> https://su.vc/survey_lh200

Once the student clicked on the link, she was directed to the participant information page, which explained the purpose of the survey and her rights, including her right to withdraw her participation at any time, the confidentiality and anonymity of her responses, and the payment protocol. The researcher's contact details were supplied for students who wished to follow up on the experiment or be debriefed. The student was told that by clicking a button on the screen, she was giving her consent to participate in the study.

Screenshots of the participant information sheet, and examples of risky choices in the gain, loss and mixed frames, are provided in Figure C 1 and Figure D 1, respectively.

4.6.2. Administration

The data was collected via online surveys of a set of decision tasks, measures of statistical numeracy, and demographic items, all programmed for completion by the respondent on a computer or hand-held electronic device without intervention by the researcher.

4.6.2.1. Online survey software

The survey software used was *SurveyAnyplace* (Survey Anyplace, 2020). The software suppliers provide the program free to students in return for feedback on the tool. The respondent was required to log into the questionnaire using her Wits student email. An option in the software was activated that blocked more than one response from the same Internet Protocol (IP) address. This option was intended to stop students from answering the questionnaire more than once, as repetitions of the decision exercise by the same respondent would have compromised the statistical assumption of independent observations. The research was thus cross-sectional rather than longitudinal.

At the start of each section, participants received instructions on the task. In the case of the risky choice task, there were detailed instructions, and the respondent had two initial practice choices. The GLM questionnaire had two initial questions within each of the gains, losses and mixed sections. The choice pairs were presented in a random order within the sections of both questionnaires.

When the student reached the end of the questionnaire, she was asked if she (i) wished to be contacted directly via email to arrange payment of her prize, should she have won a prize, (ii) preferred to receive a *Standard Bank Instant Money voucher* via her mobile phone, or (c) preferred to remain anonymous. If she preferred anonymity, she was required to send an email with her banking details to the researcher. The student was also asked to indicate if she was willing to participate in a follow-up study, and if so, to provide an email address for future contact. This aspect was not considered in the thesis.

4.6.2.2. Payment protocol

The researcher used the random lottery incentive method commonly adopted in experiments with monetary prospects of risky choice. After the student had completed the questionnaire, one of her choices was randomly selected and automatically played out within the program for payment. The respondent received a message on her mobile phone notifying her of the amount she had won (if there was an amount won). Payment was made through Standard

Bank's bulk payment system whereby an automated e-voucher was sent by the bank to the participant's mobile phone number advising her that she may draw the money at a cash terminal or participating retailer of the bank.

Choices between prospects in the loss and mixed frames were supplemented with upfront endowments of R200 and R100, respectively, so that the final outcome of a choice could not be negative. Outcomes ranged from R0 - R300 in the gain frame; a loss of R0-R200 from the initial theoretical endowment of R200 in the loss frame; and a gain or loss of R0 - R100 from the initial endowment of R100 in the mixed frame. These real, salient rewards were expected to motivate the individual and yield realistic responses (Harrison & Swarthout, 2014).

The next part of the chapter presents the data analysis. The main analyses are for the statistical numeracy data, and the risky choice data in the context of statistical numeracy. In view of the complexity of each, they are discussed separately in Sections 4.7 and 4.8, respectively.

4.7. Data analysis for numeracy

A requirement of the thesis was that the statistical numeracy of respondents was measured using a psychometrically sound instrument with a clear factor structure and *measurement invariance.* Measurement invariance refers to whether the underlying dimensions or factors of numeracy have the same meaning or interpretation across different populations (Muthén & Muthén, 1998-2017).

As previously described (section 3.7), the Schwartz/Lipkus, Berlin Numeracy, and Cognitive Reflection tests were used to measure the statistical numeracy of the respondents. However, analysing the 11 items of the tests together was important for the thesis as the individual tests have been found to be overly easy or difficult.

The first objective was to establish a clear factor structure of the items, as the Schwartz/Lipkus and Berlin Numeracy tests were designed to measure statistical numeracy, while the Cognitive Reflection Test (Frederick, 2005) was originally designed to measure cognitive reflection, although often used as a test of statistical numeracy (see Section 3.7.3.4). As it was uncertain whether a single factor or a 2-factor solution would be obtained, an exploratory rather than confirmatory approach to the factor structure of the numeracy construct was indicated. Prior to carrying out the analysis, preliminary descriptive analysis of the items was necessary to identify redundant items.

4.7.1. Preliminary numeracy analysis

The item means (proportion of respondents who answered the items correctly), standard deviations, and cross-tabulations were used to identify items that were potentially redundant. *Item characteristic curves* provided additional information on the ability of each item to discriminate between respondents along a range of levels of an underlying unobservable, or *latent,* variable (Kline, 2011). Item characteristic curves are based on item response theory (IRT). In this study, the 2-parameter model (IRT-2), was used, as computed by Mplus version 12, to examine the difficulty and discrimination of each numeracy item relative to the underlying numeracy latent trait. This model is the second of three IRT models, the first comprising a single parameter that accounts for item difficulty, the second comprising two parameters that account for both item difficulty and discrimination, and the third comprising a third parameters for the effect of item guessing (Bichi & Talib, 2018).

4.7.2. Exploratory factor analysis

Exploratory factor analysis was used to determine the number of dimensions or factors present in the set of numeracy items. Whereas the responses to each of the numeracy items were either right or wrong, and thus measured on a categorical, or binary, scale, the factors are continuous latent variables, or theoretical dimensions of numeracy. The numeracy items are indicators of the factors, or latent variables, that account for the correlations between the items. Exploratory factor analysis using the Mplus software was used to determine the factor structure of the numeracy construct. In doing so, the analysis determined the minimum number of interpretable factors that could explain the correlations among the factor indicators adequately, and also identified items that may have measured the factors poorly (Muthén & Muthén, 2010).

For categorical factor indicators, as in this study, the Mplus software uses four estimator choices: weighted least squares (WLS), robust weighted least squares (WLSM and WLSMV), and unweighted least squares (ULS). Mplus provides several recommendations for deciding on the number of factors: the interpretability of the pattern of factor loadings or factor solution taken in the context of the underlying theoretical framework of the indicators, eigenvalues greater than 1, non-significant model Chi-square goodness of fit values, Steiger-Lind root mean square error of approximation (RMSEA) values less than .06, or Standardized Root Mean Square Residual (RMSR) values less than .05, comparative fit index (CFI) or non-normed fit index (TLI) values greater than .95, and a check that there are no negative

residual variances (Hooper, Coughlan, & Mullen, 2008; Kline, 2011; Muthén & Muthén, 2010).

In this study, the numeracy factors were expected to be correlated, and thus required the default oblique rotation of Geomin for correlated factors. This Geomin rotation was used throughout the factor analyses, as recommended for correlated factors using binary or other ordered categorical dependent variables (Muthén & Muthén, 2010).

4.7.3. Measurement invariance and population homogeneity across groups

When comparisons on a construct or latent variable, such as numeracy, are made across groups, for example, across sex or income groups, it is necessary to check the measurement and structural equivalence of the construct under discussion (Asparouhov & Muthén, 2009). The requirements of measurement invariance and population homogeneity across groups within a study addresses this equivalence. A nontechnical description of measurement invariance in the context of the thesis, is whether the numeracy factors have equivalent meaning, i.e. whether they represent equivalent numeracy-related constructs across different groups (Putnick & Bornstein, 2016). A technical description of measurement invariance in a cross-sectional study, as in the present research, refers to the degree of stability of the factor model across different populations in terms of the number of factors, the same location of the zero factor loadings (configural invariance), equality of the non-zero factor loadings, equality of the intercepts, and equality of the item difficulties (scalar invariance). In item response theory, measurement invariance is known as the lack of item bias (Kline, 2011; Muthén & Muthén, 1998-2017).

Population homogeneity refers to structural invariance, i.e. equality of the factor means, variances and covariances for the sex, education level, and income populations in this study (Muthén & Muthén, 1998-2017). For meaningful comparison of the structural aspects of the numeracy factors across groups, for example, males versus females, their measurement invariance needs to hold, as otherwise the structural comparison across populations would not apply to a common set of latent variables.

In addition to the factor structure and its loadings across the groups, the factor means, variances and covariances need to be compared across the groups. In this context, it was appropriate to regard the types of groups as a fixed mode of variation, i.e. as fixed effects, since the findings of the study are inferred to all the groups in the sample used – in this case to males and females, to all the qualifications of the student groups, and to all the household

incomes groups. This type of fixed effects analysis is consistent with a multiple-group factor analysis, rather than to a multiple-level factor analysis which would have been used had the group effect been random (Muthén & Asparouhov, 2018).

Traditional fixed effects multiple groups analysis uses confirmatory factor analysis (CFA) to analyse measurement invariance. However, a major disadvantage of confirmatory factor analysis is that it requires three degrees of measurement invariance to be checked: configural invariance, which requires the same location of the factor loadings across the groups; metric or weak factor invariance; and scalar or strong factor invariance. For binary variables, only the configural and scalar models are considered.

A second disadvantage of confirmatory factor analysis is that it does not apply rotations to factor structures, and so, patterns of factor loadings cannot be compared without resorting to exploratory factor analysis where rotations are used. In confirmatory factor analysis, the onerous requirement of exact factor loadings for measurement invariance across groups and the many modification indexes that may arise during the analysis of several groups, make it difficult to identify non-invariance without resorting to a more exploratory approach (Asparouhov & Muthén, 2009; Muthén & Asparouhov, 2018). "… [T]echnically appealing, CFA requires strong measurement science that is often not available in practice" (Asparouhov & Muthén, 2009, p. 398).

Since 2009, an alternative method of analysis, coined 'exploratory structural equation modelling' (ESEM), has been made available in the Mplus software (Asparouhov & Muthén, 2009). Among its other uses, the ESEM approach allows exploratory factor analysis with targeted rotations to be used in a structural equation model for tests of measurement and structural invariance across multiple groups. It achieves these comparisons without the parameter restrictions of conventional confirmatory factor analysis.

Accordingly, the ESEM approach was adopted for the numeracy latent variable analysis in this study for its ease of comparisons of rotations and model fit across multiple groups. This method was used to establish whether measurement invariance and population homogeneity could be assumed for the covariates of sex, academic qualification, and household income. In the study, no violation of measurement invariance was found for the covariates (see Section 5.2.3.4). This favourable result meant that comparisons across the groups on factors were meaningful.

4.7.4. Latent groups

Once the researcher had established the factor structure of the numeracy latent variable and its measurement invariance, she could apply this knowledge to embark on the process of establishing the latent numeracy group structure. To establish this structure, she followed the methodology of Clark et al. (2013) who used a factor mixture model to establish an underlying structure for psychological disorders.

The factor mixture model was devised by Muthén in 2006 for situations in which latent groups are formed on a categorical latent variable, but within these groups there exists variation on a continuous latent variable or factor. The model is thus based on both categorical and continuous latent variables. The categorical aspect is used in a latent class analysis to find people who are similar within a sample based on their observed scores, and to group or classify these people into their most likely classes. Thereafter, the model uses continuous latent variables or factors to allow variability of the latent variable within the classes. In the context of this study, the factor mixture model was used to classify respondents into latent numeracy classes with within-group heterogeneity on the underlying continuous numeracy factor scores.

Following Clark et al. (2013), the logical construction of the factor mixture model is explained next, beginning with a depiction of a latent class analysis, followed by a description of factor analysis, a generalised factor mixture model, and finally variations of the factor mixture model.

4.7.4.1. Latent class analysis

A latent class analysis model is shown in the following diagram (Figure 15), in which the boxes represent the observed responses to the 11 numeracy items, the encircled letter '*C*' is the categorical latent class variable with *k* classes, and the arrows pointing from the latent class variable *C* to the boxes mean that the items are measuring, or reflecting, the latent class variable. It is assumed that the latent class variable explains the observed outcomes, and thus there is no residual correlation between the items. This is the conditional or local independence assumption (Clark et al., 2013).

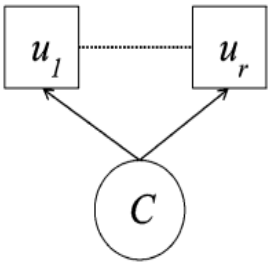


Figure 15: A latent class analysis model diagram
Note: Reprinted from Clark et al. (2013, p. 683)

For each of the latent classes, the latent class analysis provides conditional item probabilities. In the context of this study, each of the item probabilities in a latent numeracy class is the probability that a respondent in that group will respond correctly to each of the corresponding numeracy items. The latent class analysis also provides class probabilities. For this study, the class probabilities are the proportions of respondents in each numeracy class.

4.7.4.2. Factor analysis

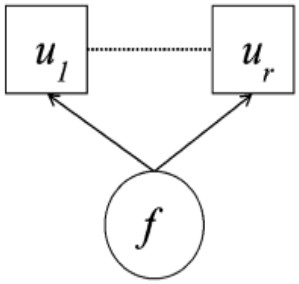


Figure 16: A factor analysis model diagram
Note: Reprinted from Clark et al. (2013, p. 683)

The factor analysis model in Figure 16 is similar to the previous latent class model diagram but contains an encircled letter '*f*' that denotes a continuous underlying factor rather than a categorical latent class variable. Respondents are assumed to differ on the factor. They are assumed to belong to one homogeneous population, rather than come from separate latent classes. As previously, the direction of the arrows means that the items are measuring the factor.

4.7.4.3. The factor mixture model and variations

The factor mixture model in Figure 17 is a hybrid of latent class analysis and factor analysis. Once again, the encircled *C* represents the latent class variable (as in Figure 15) that classifies respondents into the numeracy latent classes, and the encircled *f* (Figure 16) represents the heterogeneity within these latent classes. The solid lines from the factor to the items show

that the continuous factor/latent variable is measured by the numeracy items, and likewise, the solid lines from the categorical latent class variable show that the latent class variable is measured by the numeracy items.

In contrast to the latent class analysis and factor analysis models, all parameters are potentially class specific, thus allowing for different degrees of measurement invariance. Furthermore, the conditional independence assumption of latent class analysis is relaxed, thus allowing for individual differences within classes on the continuous factor using factor scores.

The addition of the arrows with broken lines shows that the factor structure of the model can differ for each of the classes. These differences may be on factor loadings, factor means, the factor covariance matrix, or on item thresholds. The within-class heterogeneity is represented by the continuous factor scores of the respondents, with classes obtained via the latent class variable.

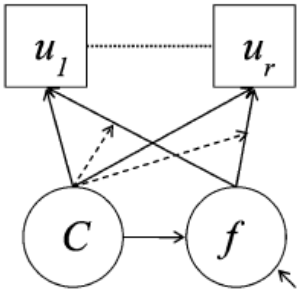


Figure 17: A generalised factor mixture model diagram
Note: Reprinted from Clark et al. (2013, p. 683)

Clark et al. (2013) show four *variations* within the factor mixture model (FMM):

The first model (FMM-1) is the most restrictive of the model variations in terms of measurement invariance. Only the factor means vary across the classes; item thresholds and factor loadings are invariant across classes, meaning that numeracy is measured in the same way in all the classes. As the factor covariance matrix is fixed at 0, there is no within-class heterogeneity in the latent numeracy trait.

The second model (FMM-2) is less restrictive as the factor covariance matrix is not fixed. Thus, there may be within-class heterogeneity in the latent numeracy trait.

In both these factor mixture models, the factor is interpreted the same way in each class. However, the later variations of the factor mixture model are progressively less restrictive. They are summarised in Table H 1. In the final variation, the least restrictive model, the

measurement parameters vary freely across the classes, and thus the factors may be different across the classes.

The researcher followed the 5-step process for factor mixture model building and comparison advocated by Muthén and colleagues and summarised by (Clark et al., 2013). In the first step, the latent class analysis model and the factor analysis model were fitted. Then the factor mixture model was fitted with two classes, increasing the number of factors iteratively under the different factor model restrictions.

Statistical as well as pragmatic criteria were used to compare the models and evaluate their fit to the data (Geiser, 2013; Muthén & Muthén, 1998-2017). The first was the statistical criterion in the form of information criteria (ICs) provided by the Mplus software. The information criteria are descriptive statistical indices that consider the goodness of fit of the model, as the parsimony of the model is based on the number of estimated parameters in the model (Geiser, 2013). The information criteria include the Akaike (AIC), the Bayesian (BIC) and the sample-size adjusted BIC. From a statistical perspective, models with the smallest information criteria values and fewer estimated parameters are preferred.

Pragmatism was the second criterion for deciding on the best factor mixture model to retain. The profiles of item probabilities in the latent classes that do not cross over each other are more interpretable theoretically. Furthermore, the probabilities of respondents' classification into the classes are important as all classes need to be represented adequately. The pragmatic criterion was important in deciding on the numeracy categorisation of the respondents in the thesis.

This concludes the section on analysis methods for the statistical numeracy data. Following the setting up of the numeracy latent classes, the next section describes how the risky choice models were fitted to these classes, and the parameters of these models estimated and compared. However, before conducting any complex analyses, the researcher conducted a rudimentary descriptive analysis of the choice proportions for each of the choice problems, aggregated separately for the respondents in each numeracy group. This preliminary analysis provided an indication of the risk aversion of the respondents. Thereafter, the analysis progressed to more complex decision model-dependent considerations of analysis, as described in the following sections.

4.8. Approach to inferring risk attitudes

There are several approaches that can be used to infer risk attitudes. Fox & Poldrack (2009) group these methods into four broad categories: statistical estimates based on decision makers' certainty (cash) equivalents; nonparametric methods that assess value and decision weights without assuming their functional forms; semi-parametric methods that assess values and decision weights, assuming their functional forms; and parametric methods that estimate values and decision weights assuming their functional forms. The researcher chose to use parametric methods as these methods generally require the participant to make fewer choices than required using nonparametric methods (Fox & Poldrack, 2009).

4.8.1. Structural models, functional forms and error models

As described in Section 2.1, the researcher used structural estimation of three latent risky choice models and compared the risk attitudes of decision makers of different statistical numeracy levels on the core parameters of these models. The categorisation of statistical numeracy was based on the factor mixture modelling of numeracy scores, as described in Section 4.7.

The parametric estimation of the three models required functional forms to be specified for each of the structural models. The models and the functional forms assumed for each are provided in Table 3. The functions were described previously (Sections 2.2.1 - 2.2.2).

Stochastic response error was accommodated by introducing Wilcox's (2008) contextual error. Stochastic parameter error was addressed using Bayesian analysis that estimates the parameters of the model as random variables (Section 4.8.4) (Bhatia & Loomes, 2017).

The two main approaches suitable for estimating the model parameters from sample choice data are *maximum likelihood estimation* and *Bayesian* analysis with its hierarchical extension. Both methods were used in this research, with emphasis on the Bayes approach, and are described in depth in the analysis of the data of the thesis (Sections 4.8.3 - 4.8.4). Thus, they are only briefly described in the following sections beginning with a short history and their underlying assumptions.

Table 3: Structural models and functional forms assumed

Model (frame)	**Functional forms**			
	Value		Probability weighting	
Expected utility (gains)	Power	$x > 0$ $U(x) = x^{\theta}$ θ risk aversion	linear	
Rank-dependent (gains & losses)	Power	$x > 0$ $U(x) = x^{\theta}$ $x = 0$ $U(x) = 0$ $x < 0$ $U(x) = -(-x)^{\theta'}$ θ risk aversion	Tversky and Kahneman (1992)	$w(p) = p^{\gamma} / (p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$ $0 < \gamma < 1$ implies an inverse *S* function γ is assumed the same value for gains/losses in maximum likelihood estimates, but estimated separately in Bayesian analyses.
	Power	linear $\theta = 1$ (assumed)	Lattimore, Baker, and Witte (1992)	$w(p) = \delta p^{\gamma} / (\delta p^{\gamma} + (1 - p)^{\gamma})$ γ discriminability/curvature (γ+ and γ- for gains/losses resp.) δ attractiveness (elevation) (δ+ and δ- for gains/losses resp.)
Cumulative prospect theory (gains, losses, mixed)	Power	$x > 0$ $U(x) = x^{\theta}$ $x = 0$ $U(x) = 0$ $x < 0$ $U(x) = -\lambda(-x)^{\theta'}$ θ risk aversion λ loss aversion	Tversky and Kahneman (1992)	$w(p) = p^{\gamma} / (p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$ $0 < \gamma < 1$ implies inverse *S* function γ is assumed the same value for gains/losses in maximum likelihood estimates, but estimated separately in Bayesian analyses.

4.8.2. Analysis paradigms: Frequentist versus Bayesian paradigms

Maximum likelihood estimation and Bayesian analysis derive from distinctly different analytical paradigms with different assumptions and methods. Maximum likelihood estimation derives from the classical frequentist paradigm, while Bayesian statistics stems from the Bayesian paradigm.

4.8.3. Frequentist modelling

The classical frequentist approach views probability as derived from the *Law of Large Numbers*. According to this law, a stable result for a random event will occur after a very large number of trials – for example, the probability of obtaining a head after a very large, or infinite, number of flips of a coin. This approach is based on defining a probability using this

large-sample limit. In finite sample sizes, however, parameters are unknown fixed constants that must be estimated from data using various methods. Examples of estimation methods include *maximum likelihood* and *least squares* (Rouder et al., 2016). Thus, the frequentists view a parameter as a fixed but unknown physical property or constant of an object, estimable from observing the data of a limitlessly repeated random sample (StataCorp, 2017). Frequentist methods are associated with the Fisher method of null hypothesises significance testing, and the use of sampling distributions and *p* values (Kline, 2013; Vallverdú, 2003).

4.8.3.1. Maximum Likelihood Estimation

Maximum likelihood estimation has been used by several risky-choice researchers (Harrison, 2008; Harrison & Rutström, 2009; Stott, 2006). In principle, the method estimates model parameters by maximising the probability of the observed data (Gujarati, 2003). It does so by selecting the parameter vector that makes the data 'most likely'. This parameter vector is known as the likelihood function. An algorithm is used iteratively to find an optimal set of parameters. It is computationally easier to maximise the natural logarithm of the likelihood function. Maximising the log function implies also maximising the original likelihood function as the log function is monotonic. This process is referred to as *maximising a likelihood function.* However, the maximum likelihood estimates need not be unique, and in some cases no maxima are found, or sub-optimal parameter values may be returned. The latter situation is known as the *local maxima problem* (Myung, 2003; Nilsson et al., 2011).

4.8.3.2. Application to the thesis

In this study, maximum likelihood estimation was used to estimate the parameters of expected utility theory, rank-dependent theory and cumulative prospect theory. The estimates were computed via Stata version 15.1 software (StataCorp, 2017), based on adaptations of the programming code supplied by Harrison (2008).

Following Harrison (2008), the researcher applied a number of steps in an attempt to satisfy the assumptions of the method and ensure the suitability of the data for maximum likelihood estimation in the context of risky choice.

First, the method's assumption of independent observations needed to be met. As each participant responded to multiple pairs of binary choices, responses were not independent due to within-person correlation. By invoking the 'cluster' command in the Stata software, whereby the standard errors of estimates are corrected, the independence assumption was satisfied. Second, the core parameters defining risk attitudes were estimated jointly in view of

their likely interdependence. Similarly, the estimates were interpreted jointly and in the context of the utility and weighting functionals applied in the modelling (Harrison & Rutström, 2008). Third, following Harrison (2008), nonlinear transformations were made to the model parameters with limited range, so the program could maximise the likelihood function over a transformed variable with unconstrained range. Finally, the researcher attempted to incorporate characteristics of the respondents into the model by estimating each parameter as a linear function of these potential covariates (Andersen, Harrison, & Rutström, 2006). This method was an attempt at modelling the parameters as a linear function of observed individual characteristics, such as demographics or numeracy scores, and to provide parameter estimates at the group or individual level (Harrison & Rutström, 2008).

As detailed in Chapter 5, maximum likelihood estimation of the model parameters was applied to a subset of the data, and the estimates used for prior probabilities for the Bayesian estimation on the remainder of the data thereafter.

The Bayesian method is described in the following section.

4.8.4. Bayesian modelling

Bayesian statistics can be used to relate complex models of psychological processes to data (Spiegelhalter, 2004). Thus, models of preference, such as the expected utility model or behavioural models, can be related to risky choice data and Bayesian statistics used to evaluate the models against the data.

In contrast to frequentists, Bayesians view probabilities as statements of observers' subjective beliefs about the occurrence of an event. These subjective beliefs are termed *prior* probabilities. These beliefs may be updated by applying Bayes' rule.

Bayes' rule is stated as follows (Bayes & Price, 1763, as cited in Kruschke & Vanpaemel, 2015, p. 281):

$$\underbrace{p(\alpha|D)}_{\text{posterior}} = \underbrace{p(D|\alpha)}_{\text{likelihood}} \underbrace{p(\alpha)}_{\text{prior}} / p(D)$$

where

$$p(D) = \int d\alpha \, p(D|\alpha) p(\alpha)$$

The updated beliefs are also expressed as probabilities and known as the *posterior* distribution. According to Bayes' rule, it is derived from a prior distribution representing the beliefs about a parameter prior to analysing the data, and a likelihood model providing information about the parameter based on observed data. As the posterior distribution reflects both the prior and the data, Bayesian methods describe changes in uncertainty when new data are considered (Kruschke & Vanpaemel, 2015, p. 298). If the sample size is smaller, the data will contribute less information, and so the prior distribution will have greater influence on the posterior than if the sample size is larger (Rouder et al., 2016). Thus, Bayesians regard probabilities as a construct that the observer learns rationally from data, rather than as an objective property of an object (Rouder et al., 2016).

In Bayesian analysis, parameters are analysed as distributions of values, rather than as one fixed value as in classical frequentist analysis. The Bayesian parameters are conceptualised as random variables with their own probability distributions based on prior knowledge and explicitly stated assumptions. The method of analysis is ideally suited to the requirements of the thesis, since it estimates the model parameters probabilistically, and thus satisfies the objective of accommodating stochastic error in the model parameters. It was therefore used to estimate central tendency and variability statistics for each of the model parameters for the individual respondents in the sample.

The following section describes the approach to the selection of the priors.

4.8.4.1. Bayesian priors

The prior distribution needs to be specified unambiguously, although there are several sources of priors and ways of specifying them, and no prior can be regarded as the correct one (Spiegelhalter, 2004). Kass and Greenhouse (1989, as cited in Spiegelhalter, 2004, p. 159) conceptualise a 'community' of priors, incorporating expert opinion and a synthesis of available evidence. One of the criteria for incorporating prior information is *exchangeability*, whereby the parameters of similar current and past studies are judged similar. If past studies measure the parameters of current interest directly, the parameters of the historical data are assumed *equal* to the parameters of current interest and the exchangeability criterion is assumed (Spiegelhalter, 2004). Priors formed according to previously observed distributions of values are known as 'empirical priors' (Spektor & Kellen, 2018).

The condition of exchangeability was used in the thesis, as the maximum likelihood estimates of the model parameters, computed on a subset of responses to the questionnaires, were used

for the prior distributions of the Bayesian parameter estimates computed on a similar set of responses. Alternately, a less stringent view of the formulation of the priors in the thesis is that they were empirical priors.

4.8.4.2. Hierarchical Bayesian analysis

When parameters are estimated under the Bayesian paradigm, the analysis considers the *similarity* between individuals, rather than considering individual observations as independent as in the frequentist paradigm. Non-hierarchical models accommodate individual differences by making separate inferences for parameters and data per person, and then use post-hoc analyses to compare individuals. Hierarchical procedures provide a further level of control over parameter estimates by partially pooling individual estimates at the group level. The group has its own distribution for the mean and the standard deviation of each of the parameters (Lee, 2011; Scheibehenne & Pachur, 2015). This imposed group-level structure informs the individual level, as the individual estimates can borrow strength from the information available about the other individuals in the group (Gelman, Carlin, Stern, & Rubin, 2004; Scheibehenne & Pachur, 2015). The hierarchical process also 'corrects' individual parameter estimates that are deemed unlikely, given the overall distribution of parameter values, by pulling them closer toward the group mean by the Bayesian principle of *shrinkage*, thereby lessening the influence of such estimates at the group level. Hierarchical solutions are thus generally considered more stable and robust than non-hierarchical ones (Kruschke, 2011).

In the context of the thesis, the statistical numeracy groups formed the structure of the model hierarchy. Thus, the statistical numeracy groups provided the group-level distribution of each parameter. Each of the individual-level parameters was assumed to come from a group-level distribution for that parameter, for example, a normal density function with a theoretically-based mean and standard deviation. Within a group of participants of the same numeracy level, each of the estimated individual-level parameters was constrained by its more restricted parameter distribution for the group of individuals with the same statistical numeracy level (Nilsson, Rieskamp, & Wagenmakers, 2011a; Scheibehenne & Pachur, 2015).

The analysis followed the example of Nilsson et al. (2011a, 2011b) who estimated the cumulative prospect theory parameters at the individual level using hierarchical Bayesian analysis. In the Nilsson et al. study, the constraints of the group-level distribution for each of the parameters of the model were based on a theoretical distribution with mean and standard

deviation derived from the risky choice modelling literature. They showed that hierarchical Bayesian modelling provided more accurate estimates than maximum likelihood (non-hierarchical) estimation. These researchers used the open-source WinBUGS software (Lunn, Thomas, Best, & Spiegelhalter, 2000; Lunn, Spiegelhalter, Thomas, & Best, 2009) for hierarchical Bayesian analysis, and supplied their programming code.

The research of the thesis differed in several ways from the study of Nilsson et al. (2011, 2011a).

First, in the thesis, the parameters of three structural models were estimated, and the group level of the hierarchy was based on the researcher's statistical numeracy categorisation. As three numeracy groups were obtained in the study (Section 5.2.4), three separate hierarchical Bayesian analyses were computed. Thus, for each group of individuals in a particular numeracy group (e.g. the high numeracy group), the group-level parameter distributions that were designed to constrain each of the estimated individual-level parameters (i.e. for each of the individuals with high numeracy), were based on prior knowledge specific to individuals in that (high) numeracy group. The priors for the model parameters for each numeracy group were obtained from the maximum likelihood estimators computed on a (different) subset of the data for each numeracy group (described in detail in Section 5.3.3).

Second, errors in the code of Nilsson et al. (2011b) were corrected: for the rank-dependent and cumulative prospect models, cumulative decision weights were implemented, and the loss aversion parameter was applied to loss gambles in addition to mixed gambles. [The MultiBUGS model code used for the expected utility, rank-dependent and cumulative prospect models is supplied for the high numeracy group in Appendices I, J and K, respectively.]

Third, the code of Nilsson et al. (2011b) was modified for the three models from binary prospects with two outcomes, to binary prospects with three outcomes.

Fourth, Wilcox's (2008) contextual formula was introduced to accommodate stochastic response error instead of using the Luce exponential rule.

Fifth, the lognormal distribution was used at the group level on prior distributions of all parameters with values that could exceed 1, and gamma transformations were used at the group level for the prior distributions of all precision parameters.

Sixth, MultiBUGS, a later version of the WinBUGS software, was used for the hierarchical Bayesian modelling (Goudie, Turner, & De Angelis, 2017). More details on these aspects of modelling and on the program are supplied in Section 5.3.4.

4.8.4.3. Deriving the posterior distribution using Markov Chain Monte Carlo sampling

When models are complex, such as risky choice models, the posterior distribution cannot be derived directly. Instead, it must be approximated by a technique such as the *Markov Chain Monte Carlo* (MCMC) sampling method to obtain probability distributions of the unknown parameters (Geyer, 1992; Van Ravenzwaaij, Cassey, & Brown, 2018). MCMC generates long sequences of samples, known as *chains*, starting from the prior distributions of the parameters to be estimated. According to Bayes' rule, the prior distributions are updated by samples of the data that result in the posterior distributions (Heathcote et al., 2019).

A simple description of the MCMC process for estimating the parameters of the risky choice model is as follows: The sampling algorithm uses a proposal distribution of initial values for each of the parameters. The Gibbs sampler is an MCMC algorithm used by the MultiBUGS software. MCMC sampling algorithms propose a chain of steps, intelligently selecting different random samples of parameters repeatedly from the posterior distribution of all possible combinations of the parameters considered jointly. Each sample in the chain is dependent only on the values of its immediate predecessor, in accordance with the Monte Carlo principle (Van Ravenzwaaij et al., 2018). The chain moves to sets of parameters in higher and higher probability regions of the parameter space in an iterative process that retains sets of parameters that successively increase the probability of convergence with the posterior distribution of the target. If the process is allowed to continue long enough, the distribution of retained parameters will eventually converge to the posterior distribution, so that eventually the chain of samples will reach its stationary, or desired, posterior distribution (SAS Institute Inc., 2014). At this point onwards, the samples drawn are considered to be reliable estimates of the posterior distribution (Heathcote et al., 2019). From the chains of samples, point estimates are calculated for the posterior model parameters, such as means, medians and percentiles. Additionally, interval estimates (referred to as 'credible intervals') for the parameter distributions, and probabilities of the parameter values in the posterior population distribution, are calculated (Heathcote et al., 2019). Thus, Bayesian inference treats parameters as random variables with distributions (Shiffrin, Lee, Kim, & Wagenmakers, 2008).

4.8.4.4. Checking convergence

As previously described, Bayesians view parameters as unknown random variables, and inferences are based on the posterior parameter distributions. Only if the diagnostics of the chains indicate convergence with the posterior target distributions of all model parameters may the samples of the posterior be used for inference purposes. Otherwise, inferences based on the posterior samples cannot be considered valid.

Although several convergence diagnostics were published in the late 1990s, it has since been recognised that there is no absolute criterion for the evaluation of convergence. As explained by Robert and Casella (2010, p. 238), the inherent randomness in the sampling process mitigates against a "categorical guarantee of performance", and there will always be chains that will invalidate a certain diagnostic. Diagnostics cannot prove convergence – only show non-convergence.

As recommended (Robert & Casella, 2010), the researcher ran the MCMC analyses with multiple parallel chains from different starting states, and compared them on multiple convergence diagnostics. She examined the graphic plots and diagnostics provided by the open source *CODA* package in *R* (Plummer, Best, Cowles, & Vines, 2006).

Graphic plots

CODA provides trace plots, density plots, and autocorrelation plots for assessing the mixing and convergence of chains. The researcher examined overlaid plots of the chains (Plummer et al., 2006).

Trace plots examine the mixing of each chain, i.e. the efficiency of the GIBBS sampling algorithm in the MCMC sampling data. The plots show the history of the parameter values generated by the samples over the iterations. The plots are used to examine the number of random draws or iterations it takes for the chains to converge. If the chain has good mixing, the plot should have approximately constant mean and variance across iterations and not show any trends. Density plots provide the distribution of the posterior distribution of each parameter value. These plots should be normally distributed and not show any trends. Autocorrelation plots check for convergence by assessing whether the correlation between samples, or between samples with a number of steps between them, becomes increasingly smaller. Slow mixing, and high correlation between samples, are indicated if the autocorrelation does not drop off.

Summary statistics

The researcher examined the Gelman-Rubin (1992) diagnostic to test whether the chains converged to the same target distribution. The statistic compares parallel chains based on iterations from the second half of the chains onwards. It generates a statistic, called the *estimated potential scale reduction factor*, or the *shrink factor*, based on the variance between sequences, relative to the variance within sequences of samples. This ratio approaches the value of 1 as the chains converge (Cowles & Carlin, 1996).

4.8.5. Validation of model estimates

As advised by Wilcox (2011), two criteria are required for evaluating the performance of model specifications. The first is the model *explanation* criterion that depends primarily on the correct structural model and entails examining the in-sample fit of the model in comparison to other models. For the thesis, the comparisons of fit were made between the numeracy groups. The second criterion is the model *prediction* performance that depends primarily on the correctness of the stochastic model. It entails comparing the predictions of the model on out-of-sample data. This sample is also known as a 'holdout sample' in the context of data mining, also referred to as cross-validation or generalisation. Improvements in the explanatory power of models thus depend on the structure of the model (expected utility versus rank-dependent utility or other), whereas improvements in the predictive power of models depend on stochastic models.

The researcher used the parameter estimates of the models, per numeracy group, separately on the in-sample data and the out-of-sample data, to examine the correspondence between the model estimates versus the observed in-sample choices, and the correspondence between the model predictions versus the observed out-of-sample choices. For each model, mixed repeated measures analyses of variance, were computed on the consistency data of each respondent, to examine the numeracy and gradient effects of the Marschak-Machina triangles, on in-sample and out-of-sample consistency data. Mixed linear models were computed for the most successful of the models, as a check on the results (and underlying assumptions) of the mixed repeated measures analyses of variance (see Section 5.3.5).

The more rigorous validation method of parameter recovery was outside the scope of the thesis (see Section 6.2.2). Murphy and ten Brincke (2017) and Nilsson et al. (2011a) provide examples of the parameter recovery method.

This section completes the methodology on analysis. The following Sections (4.9 and 4.10) conclude the chapter with a discussion of the internal and external validity of the research.

4.9. Internal validity of the research

In the psychological research tradition, the *internal validity* of a quantitative research study refers to whether a single, unambiguous explanation or cause can account for the relationship between the variables of the study (Gravetter & Forzano, 2012, p. 170). Previously (Section 4.2.1), the researcher described the research as quasi-experimental with only partial internal validity. However, there can be no claim of a causal relation between the statistical numeracy and risky choice indicators, as multiple unconsidered variables could have contributed to the analyses comparing the risky choice indicators of preference. Several variables reported to be correlated with numeracy and risk taking were included in the study as covariates, for example, sex, income and highest education level (Skagerlund et al., 2018). However, group differences could be, at best, only partially controlled, as random assignment of participants to numeracy groups was not appropriate. The researcher checked that the factor structure was invariant for the covariates of sex, age and education groups, which ensured the numeracy factors could be interpreted similarly for variables traditionally found to be correlates of numeracy. Notwithstanding, statistical numeracy still cannot be claimed the cause of the differences between the groups on risk attitudes as it is possible that other extraneous variables could have been operating.

Additional issues specific to the online administration of the questionnaires are considered unlikely to have jeopardised the quality of the data, and the relation between numeracy and risk attitudes (Horton et al., 2011). The first is whether the online responses of the participants were unique and independent. As the researcher identified respondents by their student number, any repeat responses were identified and discarded. Additionally, the software did not allow multiple entries from the same IP address. Thus, although these measures did not guarantee unique responses, it is likely to have reduced them, and so this issue was not regarded as a threat to the study. A second concern was the potential of attrition of respondents and the participants providing random responses. Attrition of respondents was not considered problematic as the researcher discarded incomplete responses. As she expected that participants who responded randomly would also have clicked through the questionnaire quickly, she eliminated responses of all participants who had completed their questionnaires in under 10 minutes, as well as those with no variance in their responses to the

risky choice items. These attempts to obviate noise in the data resulted in a cleaner data set for the research.

4.10. External validity of the research

As discussed (see Section 4.3), the researcher recruited a large sample of students from the database of undergraduate and postgraduate students at the University of the Witwatersrand, and so the students selected themselves into the study. The sampling is thus a nonprobability availability sample and suffers from the usual weaknesses of non-representativeness, or poor external validity, inherent in laboratory experiments on students' risky choices. However, the sample was never intended to be representative of any student population. The purpose that these students fulfilled in the study was to ensure that they represented a wide range of numeracy scores, and expressed their risk attitudes. This said, it is feasible that the size of the incentive attracted a risk averse sample as the potential of a R200 or R300 prize is likely to be appealing to students.

The research was intended to investigate the role of statistical numeracy in the three risky choice models, under the most feasible and favourable conditions. Although the research used real, salient incentives to encourage accurate choices under risk, no guarantee can be made of the study's ecological validity, or its applicability to simulate risky choice in real-world economic situations.

Chapter 5: Results

This chapter presents the results of the research. To begin, it provides preliminary descriptive statistics on the three numeracy scales (5.1). Thereafter, the main sections of the chapter are aligned with the research questions and sub-questions: first, questions on the latent structure of the numeracy construct, the latent subgroups of statistical numeracy and the measurement of the construct (5.2); and second, questions on estimation and validation of the parameters of the normative and descriptive models of the risky choices of the latent subgroups (5.3). The implications of the research findings for economic policy related to risky choice are discussed in Chapter 6 (6.3).

5.1. Preliminary analysis of numeracy scales

Descriptive statistics are supplied in Table 4 for the three numeracy scales selected for the study. The table also provides internal consistency estimates for each test, using the Kuder-Richardson Formula 21 (K-R 21). The K-R 21 estimates are appropriate for assessing the internal consistency aspect of reliability for tests comprising dichotomous items, assuming that the scales are unidimensional. [K-R 21 is the equivalent of the Cronbach alpha coefficient for tests comprising continuous items (Ghiselli et al., 1981).] The mean inter-item correlation is also supplied, as its value is unrelated to the number of items in the scale (unlike the K-R 21 and coefficient alpha values). A value of .3 or greater for the mean inter-item correlation is the guideline for acceptable internal consistency (Hair et al., 2010). Histograms of the total score distributions of each scale are supplied in Figure G 1. Ideally, the score distributions should assume the symmetric shape of the normal density function.

Table 4: Descriptive statistics and internal consistency estimates of the numeracy scales (n = 666)

Test	Items	Mean	Median	*SD*	Skewness	Internal consistency	Mean inter-item correlation	Mean (%)
Schwartz/Lipkus	4	3.24	4	0.98	-1.13	0.52	0.21	81%
Berlin Numeracy	4	1.69	2	1.14	0.35	0.51	0.21	42%
Cognitive Reflection	3	0.85	0	1.05	0.90	0.68	0.41	28%

The Schwartz/Lipkus score distribution is strongly negatively skewed, with a mean score of 81% and median value equal to the maximum score. The test was easy for the respondents and provides little, or no, discrimination for high scores. Conversely, the Berlin Numeracy and Cognitive Reflection total score distributions are positively skewed. The Cognitive Reflection Test was particularly difficult for the respondents, whose mean score was 0.85 out of 3 (28%), with median score of 0 out of 4. The score distribution of the Berlin Numeracy Test is the least skewed, with the highest variability. However, the internal consistency of the scale is below the minimum acceptable value.

In Section 5.2, multivariate analyses are conducted on the factor structure of the 11 items of these scales.

5.2. Analysis of statistical numeracy responses

For the convenience of the reader, the research question and sub-questions pertaining to the latent structure of statistical numeracy and its measurement are restated:

Research question 1: What is the structure of the respondent groupings on numeracy?

1.1 What is the factor structure of numeracy?

1.2 Do the factors display measurement invariance?

1.3 What are the latent statistical numeracy subgroups?

As previously outlined (Section 3.7.3), the items of three existing numeracy instruments were used to measure statistical numeracy as follows: four items selected from the Schwartz/Lipkus (SL) Test (Lipkus et al., 2001; Schwartz et al., 1997), the four items of the Berlin Numeracy Test (Cokely et al., 2012), and the three items of the Cognitive Reflection Test (Frederick, 2005).

The psychometric adequacy of these numeracy items was essential for the study. As such, a process of analysis was undertaken, starting with a thorough analysis of the item difficulty levels, variability, intercorrelations and discrimination ability to identify items that would be helpful, versus redundant, to the objective of categorising respondents into numeracy subgroups (Section 5.2.1). Thereafter, the factor structure of the numeracy indicators and examination of its measurement invariance were investigated (Section 5.2.3), thus addressing Research sub-questions 1.1 and 1.2. The factor structure was used to inform the categorisation of respondents into numeracy groupings (Section 5.2.4), thus addressing Research sub-question 1.3.

5.2.1. Item difficulty, variability, cross-tabulations and discrimination ability

The numeracy items are presented in Table 5. The table displays the acronyms used from here on to refer to the items based on the tests from which they are derived, the item wording used in the thesis, the proportion correct responses (item means), and response variability (item standard deviations).

Table 5: Numeracy items, acronyms, percentage correct responses and response variability

Item		Mean	*SD*
SL1	Imagine that we rolled a fair, six-sided die 1,000 times. (The word 'die' is the singular for dice.) Out of 1,000 rolls, about how many times do you think the die would come up even (2, 4, or 6)?	0.77	0.42
SL2	In a lottery, the chance of winning a R10 prize is 1%. About how many people would win a R10 prize if 1,000 people each buy a single lottery ticket?	0.73	0.44
SL3	In a competition, the chance of winning a car is 1 in 1,000. What percent of tickets win the car?	0.79	0.41
SL4	Which of the following numbers represents the biggest risk of getting a disease? (1 in 1,000; 1 in 100; 1 in 10)	0.95	0.21
BNT1	Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir, 100 are men. Out of the 500 inhabitants that are not in the choir, 300 are men. What is the probability that a randomly drawn man is a member of the choir?	0.32	0.47
BNT2a	Imagine we are throwing a five-sided die 50 times. (The word 'die' is the singular for dice.) On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3, or 5)?	0.73	0.44
BNT2b	Imagine we are throwing a loaded (unbalanced) die with six sides. (The word 'die' is the singular for dice.) The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?	0.49	0.50
BNT3	In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?	0.15	0.36
CRT1	A bat and a ball cost R1.10 in total. The bat costs R1.00 more than the ball. How much does the ball cost?	0.28	0.45
CRT2	If it takes 5 machines 5 minutes to make 5 items, how long would it take 100 machines to make 100 items?	0.25	0.43
CRT3	In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	0.32	0.47

The correct responses to all but one item ranged from 15% - 79%, with fair variability as indicated by the standard deviation (*SD*), where the maximum *SD* of a dichotomous item is .5. The exception was item SL4: *Which of the following numbers represents the biggest risk of getting a disease? 1 in 1,000; 1 in 100; 1 in 10*. The respondents found this item very easy – it was answered correctly by 95% of respondents and had low variability (*SD* = 0.21). This result was the first indication of the potential redundancy of the item.

Although the researcher wished to discriminate between respondents at all levels of statistical numeracy, item SL4 was considered unduly easy and non-discriminating for the sample respondents. By contrast, item BNT3 – the most challenging item – was answered correctly by only 15% of the respondents but was retained for further investigation as its variability was higher (*SD* = .36).

Next, the cross-tabulations of the frequencies of responses to the items were considered (Table 6).

5.2.1.1. Item cross-tabulations

The inter-item relationships are important for identifying items with unduly high intercorrelations. Such items are problematic for analyses of the factor structure of the underlying measurement scale, and so should be considered for deletion (Muthén & Muthén, 1998-2017). In Table 6, each cross-tabulation entry shows the number of respondents who answered one item of that pair correctly (the item in the row, indicated as a ‘1’ in the first column of numbers) and the other item wrongly (the item in the column, indicated as a ‘0’ in the first row of numbers). Thus, the frequencies in the main diagonal of the table refer to the same item, and so are necessarily 0.

In the table, a frequency of 0 in the off-diagonal entries indicates a perfect correlation between the items in a pair of non-identical items. Frequencies less than 10 in the off-diagonal cells are highlighted (underlined, italicised and bolded) in the table, as they show high inter-item correlations. The highlighted values for item SL4 (in the column indicated by the selected outlined block) show a disproportionately high overlap with the responses to 6 of the other 10 items, once again indicative of the possible redundancy of the item.

Table 6: Cross-tabulations of responses to item pairs measuring statistical numeracy

		SL1	SL2	SL3	SL4	BNT1	BNT2a	BNT2b	BNT3	CRT1	CRT2	CRT3	
		0	0	0	0	0	0	0	0	0	0	0	Total
SL1	1	0	105	74	16	335	90	239	423	340	362	312	515
SL2	1	78	0	74	13	308	99	219	397	320	338	297	488
SL3	1	83	110	0	18	338	109	240	432	344	369	326	524
SL4	1	135	159	128	0	426	164	319	534	450	471	426	634
BNT1	1	34	34	28	***6***	0	39	81	150	110	122	101	214
BNT2a	1	61	97	71	16	311	0	209	393	317	337	288	486
BNT2b	1	48	55	40	***9***	191	47	0	249	190	211	171	324
BNT3	1	***9***	10	***9***	***1***	37	***8***	26	0	33	51	35	101
CRT1	1	12	19	***7***	***3***	83	18	53	119	0	94	64	187
CRT2	1	14	17	12	***4***	75	18	54	117	74	0	58	167
CRT3	1	***8***	20	13	***3***	98	13	58	145	88	102	0	211
	Total	151	178	142	32	452	180	342	565	479	499	455	666

5.2.1.2. Item difficulty and discrimination levels relative to a numeracy factor

According to item response theory, the discrimination ability of items is based on a user-independent, automated calculation of a standardised, single underlying latent factor (Muthén & Muthén, 2010). In line with IRT-2 theory (Section 4.7.1), two parameters are estimated for each item relative to the standardised factor scores. The first parameter is the item difficulty level, displayed via the position of item characteristic curves plotted relative to the numeracy factor scores on the X axis. These curves show the probability of a correct response to each item as the factor score increases. Curves placed further to the left of the axis indicate easier items; those placed further to the right indicate more difficult items. The measure of item difficulty is the factor score on the X axis that corresponds to a line drawn from the median value of .5 on the Y axis to the item curve. This measure is the factor score value above which at least half of the respondents would be expected to score the item correctly. The second parameter is each item's ability to discriminate between respondents along different levels of the factor, displayed via the slopes of each curve.

The characteristics of the 11 numeracy items were evaluated based on their item characteristic curves plotted relative to the standardised numeracy factor scores on the X axis

(Figure 18). The items display varying difficulty levels, as their positions range widely across the numeracy factor scores. As expected, the leftmost curve corresponds to item SL4, the item already earmarked for exclusion as it was unduly easy for respondents.

The discrimination abilities of the items, displayed via the slopes of the curves, show the ability of the items to discriminate between respondents with statistical numeracy factor scores over the corresponding factor score range. The item characteristic curve of item SL4 has steep slopes over the -4 to -2 standardised factor score range, which would discriminate over the low extreme of the numeracy factor scores. However, even if this item is omitted, the remaining item curves discriminate over a wide range of factor scores, for example, from -2 to +3. By contrast, the rightmost curve corresponds to item BNT3, the most difficult item. This item discriminates well over the higher range of the numeracy factor scores (1 to 4) where few other items discriminate. Overall, the slopes of the curves show that the numeracy items have discrimination ability across a wide range of numeracy factor scores.

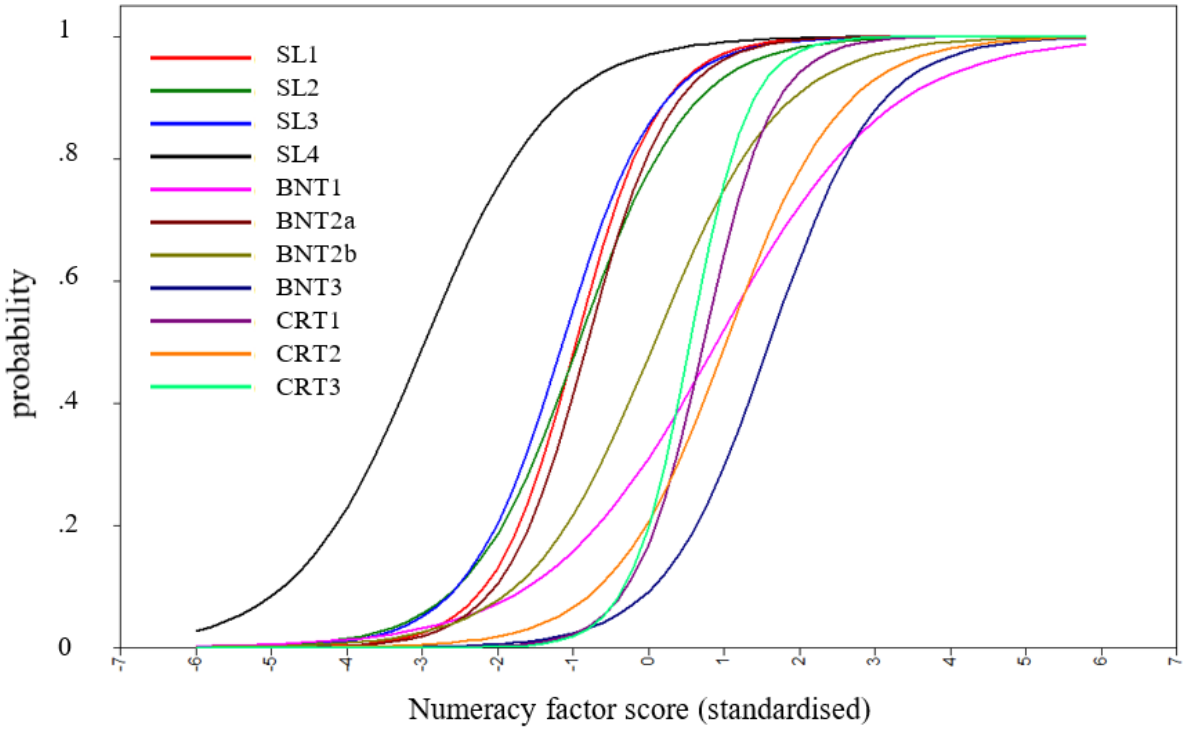


Figure 18: Item characteristic curves for 11 indicators of statistical numeracy

5.2.2. Conclusions based on item-level statistics

The analyses of the numeracy items in terms of difficulty, variability, intercorrelations and discrimination ability indicate that the items are sensitive to individual differences across a

wide range of factor scores. There is also strong evidence that item SL4 is redundant and should be dropped from all subsequent analyses. The next analyses proceeded on this basis.

5.2.3. Factor structure of the statistical numeracy measurement

The IRT-2 analysis assumed a single factor. However, a single-factor solution was not necessarily optimal for the remaining analyses. This section presents the analysis of the factor structure of the remaining 10 statistical numeracy items. The results are presented in three stages.

First, exploratory factor analysis was undertaken to identify the number of factors and item loadings for the numeracy scores of the total sample of 666 respondents (5.2.3.1). This analysis used weighted least squares estimation and Geomin rotation, appropriate for binary item responses (see Section 4.7.2). Thereafter, the measurement invariance of the factor structure was checked using the technique of exploratory structural equation modelling (ESEM) in the Mplus software (5.2.3.4). Finally, the population homogeneity of the factor structure was checked for the covariates of age, highest education qualification completed, sex and household income, based on the factor means and variances of comparison groups (5.2.3.5).

5.2.3.1. Initial exploratory factor analysis: 10 items

Exploratory 1-factor and 2-factor solutions were fitted to the 10 statistical numeracy items previously described.

A summary of the model fit diagnostics is presented in Table 7. As described in Section 4.7.2, the model fit diagnostics considered are RMSEA, CFI, TLI and SRMR.

The 2-factor solution fitted the data better than the single factor solution on all fit indices: First, the Chi-square fit index is not significant (p = .723) for the 2-factor model showing better fit than the significant 1-factor model (p = .0007). Second, the RMSEA value is favourably lower for the 2-factor model, and lower than the customary .05 comparison value. Third, the CFI and TLI values are favourably greater for the 2-factor model and greater than the customary .96 comparison value. Fourth, the SRMR values are favourably lower for the 2-factor model and lower than the customary .05 comparison value.

The results of the Geomin oblique 2-factor rotation are supplied in Table 8. Based on the *Z*-test statistics for the estimated Geomin-rotated factor loadings relative to their standard errors (shown in the two rightmost columns of the table), the loadings of three test items (SL1, SL3

and BNT2a) are clearly higher on Factor 1 than on Factor 2, and those of five items (BNT1, BNT3, CRT1, CRT2, CRT3) are clearly higher on Factor 2. Two items (SL2 and BNT2b) discriminate less well on the factors.

Table 7: Summary of model fit information for 1- and 2-factor solutions for 10 indicators of statistical numeracy

Factors	Indicators	Number of parameters	χ^2	df	*p*-value	RMSEA	CFI	TLI	SRMR
1	10	10	68.073	35	0.0007	0.038	0.985	0.98	0.065
2	10	19	21.361	26	0.7231	0.000	1.000	1.004	0.034

Table 8: Geomin oblique factor rotation on 10 binary indicators of statistical numeracy

	Geomin rotated loadings (* significant at 5% level)		*SE* Geomin rotated loadings		*Z:* Est./*SE* Geomin rotated loadings	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
SL1	0.830*	-0.006	0.062	0.014	**13.418**	-0.458
SL2	0.338*	0.336*	0.110	0.105	3.085	3.210
SL3	0.459*	0.277*	0.115	0.114	**4.005**	2.434
BNT1	-0.244	0.825*	0.196	0.153	-1.246	**5.383**
BNT2a	0.653*	0.149	0.122	0.131	**5.355**	1.136
BNT2b	0.255*	0.389*	0.117	0.102	2.189	3.800
BNT3	-0.002	0.718*	0.011	0.060	-0.157	**12.03**
CRT1	0.216	0.677*	0.133	0.103	1.615	**6.593**
CRT2	0.195	0.571*	0.130	0.105	1.507	**5.436**
CRT3	0.442*	0.533*	0.123	0.103	3.604	**5.151**

* $p < .05$

In a bid to use the most discriminating and efficient set of items for the numeracy construct, items BNT2a and BNT2b (see Table 5) were re-examined. These items are used as filter items in the computer-adaptive test format of the Berlin Numeracy Test for deciding whether item BNT3 is presented to the respondent (Cokely et al., 2012). As item BNT3 was important to the factor structure based on its high *Z*-score value, items BNT2a and BNT2b were not necessary to filter it out here, and thus not considered esssential for the questionnaire. Nevertheless, the decision whether to drop one of these items was based on a comparison of

model fit, factor structure interpretability, and measurement invariance, with both items included versus with one excluded. As the factor loadings of the former item BNT2a were both higher and more discriminating between the factors than those of item BNT2b, the model fit information was reassessed with the latter item excluded.

5.2.3.2. Model fit comparison: 10 versus 9 items

Consistent with the superior model fit of the 2-factor solution over the 1-factor solution for 10 items, the fit of the 2-factor solution was also superior on all fit indices for 9 items, excluding item BNT2b (Table 9). Additionally, the eigenvalues for the 10-item sample correlation matrix were marginally lower at 61.8% of explained variance compared to 64.8% for the 9-item sample correlation matrix. The next step in deciding whether to eliminate item BNT2b was the interpretability of the factor structures with 9 and 10 items.

Table 9: Summary of model fit information for 1- and 2-factor solutions for indicators of statistical numeracy with 9 items

Factors	Indicators	Number of parameters	χ^2	df	*p*-value	RMSEA	CFI	TLI	SRMR
1	9	9	63.297	27	0.0001	0.045	0.981	0.974	0.070
2	9	17	15.178	19	0.7112	0.000	1.000	1.004	0.032

5.2.3.3. Factor structure interpretability: 10 versus 9 items

The factor loadings of both the 9- and 10-item Geomin rotated solutions were similar (Table 8 and Table 10, respectively). Based on the *Z*-test statistics for the estimated Geomin-rotated factor loadings relative to their standard errors, the most dominant loadings for both the 9 and 10 indicators are SL1, SL3 and BNT2a on the first factor, and BNT1, BNT3, CRT1, CRT2 and CRT3 on the second factor. The factor loadings for 9 and 10 indicators are relatively stable. The first factor appears to reflect statistical numeracy involving relatively simple frequency or percentage calculations; the second factor appears to reflect statistical numeracy involving more complex, multistep probability calculations, as in Bayesian-type reasoning. For now, the two factors are said to reflect simple and complex statistical numeracy, respectively. The interpretation of the factors is discussed against the theoretical background of cognitive ability in Section 6.1.1.

Table 10: Geomin oblique factor rotation on nine binary indicators of statistical numeracy

	Geomin rotated loadings (* significant at 5% level)		*SE* Geomin rotated loadings		*Z:* Est./*SE* Geomin rotated loadings	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
SL1	0.852*	-0.005	0.064	0.010	**13.223**	-0.496
SL2	0.333*	0.339*	0.108	0.102	3.095	3.318
SL3	0.452*	0.285*	0.112	0.111	**4.023**	2.567
BNT1	-0.245	0.830*	0.194	0.152	-1.259	**5.442**
BNT2a	0.630*	0.159	0.117	0.125	**5.372**	1.273
BNT3	0.000	0.712*	0.010	0.060	-0.030	**11.906**
CRT1	0.215	0.679*	0.131	0.099	1.640	**6.832**
CRT2	0.194	0.578*	0.128	0.102	1.514	**5.675**
CRT3	0.436*	0.537*	0.121	0.100	3.607	**5.385**

The final decision on the elimination of item BNT2b was the examination of the measurement invariance of the 2-factor structure with this item included, versus excluded.

5.2.3.4. Measurement invariance of the factors: 10 versus 9 items

The results in this section are presented in line with the motivation presented in Section 4.7.3 for using the multiple group ESEM approach of Mplus to investigate the measurement invariance of the numeracy scale (Muthén & Asparouhov, 2018). Measurement invariance was examined for the demographic covariates of this study, i.e. for sex, highest education qualification completed, household income and age. These variables were categorised into meaningful subgroups of approximately equal sizes: males versus females, matriculated versus further qualified respondents, respondents with household monthly income less versus more than R10,000, and respondents aged 18-21 versus older.

According to the ESEM approach, two steps are required to test for measurement invariance of the binary dependent variables, i.e. the numeracy items with correct/wrong responses. In the first step, the *configural* factor model is tested by examining the model fit of a largely unconstrained factor solution. This step tests whether the numeracy factors have the same pattern of loadings for the categories of each of the covariates. The configural model has free factor loadings and thresholds across groups, scale factors fixed at a value of 1 in all groups, and factor means fixed at 0 in all groups. The second step is more stringent, as it tests the *scalar* model by examining whether a more constrained model fits the data significantly

worse than the freer, configural model. This step tests whether the means (intercepts) of the items that loaded on the factors are equal, or equivalent, for the groups compared. The scalar model has factor loadings and thresholds constrained equal across groups, scale factors fixed at a value of 1 in one group and free in the other groups, and factor means fixed at 0 in one group and free in the other groups (Muthén & Muthén, 1998-2017).

Table 11 presents a summary of the tests of configural and scalar measurement invariance for each covariate. In addition, it presents the sample sizes compared, the number of free parameters, the model fit information (the Chi-square goodness of fit test, and the RMSEA, CFI, TLI and SRMR values), and the Chi-square test of the difference. All models used the 2-factor solution identified previously, Geomin oblique factor rotation and weighted least square parameter estimates.

The first covariate considered in the ESEM modelling was sex. Although the model fit and factor structure interpretability of the 2-factor solution was in favour of dropping item BNT2b and using 9 rather than 10 indicators of numeracy, the ESEM modelling was performed for males versus females separately on both the 9 and the 10 sets of items as an additional check. The results confirmed the advantage of the 9-item solution. Both the 9- and 10-item models were invariant with respect to sex: the values of all the Chi-square model fit indices were favourable for both the configural models, for both the structural models, and for both the Chi-square tests for the difference between the constrained versus less constrained models. There were also no modification indices suggested for improving the fit of the models. In all cases, all the fit indices were better with item BNT2b omitted rather than included, yielding after exclusion, for example, the higher *p*-values for the Chi-square fit statistics (.682 versus .272, and .645 versus .318 for the configural and scalar 9- versus 10-item analyses, respectively in Table 11). Thus, these results confirmed that the item could be dropped without penalising the research. Accordingly, the decision was made to run the ESEM analyses on the remaining covariates with nine items only.

The ESEM analyses comparing matriculated versus further qualified respondents, respondents with household monthly income less versus more than R10,000, and 18-21-year-olds versus older respondents, showed similar measurement invariance. There was one modification index that suggested an improvement in the model fit for the measurement invariance analysis of income. This involved introducing a path from SL2 to the more complex factor (Factor 2), and would have reduced the Chi-square value by 10.058. As this

value was only marginally greater than the minimum cutoff value of 10 (Muthén & Muthén, 2010), it seemed to have little substantive merit and was not introduced to the model.

After considering all the evidence summarised in Table 11, the 9-item 2-factor structure was judged to have measurement invariance for the covariates of sex, qualification, income and age. Thus, the interpretation of the numeracy factors or latent constructs was equivalent for males and females, respondents with different qualifications, income levels, and age groups. This equivalence was the necessary condition for meaningful comparisons of the factor scores across sex, qualification, income and age.

For completeness, a histogram of the total scores of the nine items is provided in Figure G 1, together with histograms of the three underlying measurement scales. Descriptive statistics are supplied in Table G 1. The 9-item distribution is more symmetric than any of its three underlying scales, based on the value of its skewness (0.25). The values of its mean and median are close, at 48.3% and 44.4%, respectively. As the scale is multidimensional, its internal consistency is not calculated.

Table 11: Multiple group ESEM tests of measurement invariance of the 2-factor structure for numeracy

Model					Chi-square test of model fit							Chi-square test for difference testing		
Covariate	Sample sizes	Type	Indicators	Free parameters	χ^2	df	*p*-value	Chi-square contribution from each group	RMSEA	CFI	TLI	SRMR	χ^2	df	*p*-value
Sex	Male: 283; Female: 382	Configural	10	58	57.703	52	0.273	Male: 21.406; Female: 36.297	0.018	0.997	0.995	0.050			
		Scalar		41	74.026	69	0.318	Male: 30.171; Female: 43.855	0.015	0.997	0.996	0.059	17.142	17	.445
		Configural	9	52	33.405	38	0.682	Male: 13.282; Female: 20.123	0.000	1.000	1.005	0.043			
		Scalar		37	48.623	53	0.645	Male: 20.999; Female: 27.624	0.000	1.000	1.004	0.054	14.787	15	.467
Qualification	Matric: 368; Qualified: 298	Configural	9	52	28.894	38	0.856	Matric: 19.897; Qualified: 8.997	0.000	1.000	1.009	0.037			
		Scalar		37	46.845	53	0.711	Matric: 27.951; Qualified: 18.894	0.000	1.000	1.004	0.050	16.324	15	.361
Income	<R10k: 280; >=R10k: 386	Configural	9	52	25.778	38	0.935	<R10k: 14.106; >=R10k: 11.672	0.000	1.000	1.013	0.036			
		Scalar		37	40.978	53	0.886	<R10k: 23.590; Qualified: 17.388	0.000	1.000	1.009	0.044	13.588	15	.557
Age	18-21: 338; 22+: 328	Configural	9	52	29.912	38	0.823	18-21: 15.905; 22+: 14.007	0.000	1.000	1.008	0.038			
		Scalar		37	49.592	53	0.608	18-21: 28.445; 22+: 21.147	0.000	1.000	1.003	0.051	17.812	15	.273

5.2.3.5. Population homogeneity of the numeracy factor structure

The group comparisons were again computed using the Mplus multiple group ESEM approach, but with a new set of constraints designed to compare the means and variances of the factors across the two subgroups of each covariate. The analysis used the defaults of group-invariant loadings and thresholds for the binary dependent variables, and group-varying factor variances and means. In so doing, the population homogeneity of the numeracy factor structure was examined for the covariates.

The results are summarised in Table 12, based on nondirectional comparisons. They show that female respondents scored significantly worse than males did on Factor 2 ($p < .001$), that more highly qualified respondents scored significantly worse than matriculated respondents on Factor 1 ($p = .004$), that respondents with higher income scored higher than those with lower income on both factors ($p < .001$ and $p = .007$, respectively), and that older respondents scored lower than younger respondents on Factor 1 ($p < .001$).

Table 12: Mean factor scores of multiple group ESEM analyses for covariates

Covariate	Subgroup	Means	Estimate	*SE*	Z: Est./*SE*	2-tailed *p*-value
Sex	Male	F1	0.000			
		F2	0.000			
	Female	F1	-0.244	0.154	-1.583	.113
		F2	-0.768***	0.155	-4.946	< .001
Qualification	Matriculated	F1	0.000			
		F2	0.000			
	Qualified	F1	-0.365**	0.126	-2.899	.004
		F2	-0.188	0.139	-1.352	.170
Income	<= R10,000pm	F1	0.000			
		F2	0.000			
	> R10,000pm	F1	0.405***	0.123	3.300	< .001
		F2	0.567**	0.210	2.704	.007
Age	18-21	F1	0.000			
		F2	0.000			
	22+	F1	-0.510***	0.134	-3.792	< .001
		F2	-0.133	0.156	-0.856	.392

** $p < .01$; *** $p < .001$

5.2.3.6. Correlation between the numeracy factors

The final step of the numeracy analysis was inspection of the correlation between the factors. As expected both statistically and conceptually, the factors are strongly correlated. Statistically, the correlation was based on the orthogonal Geomin rotation; conceptually, different numerical abilities would be expected to be correlated. Factor scores based on simple linear weighting of the 9-item rotated factor loadings (Table 10) produce a correlation between the factors of .66. The scatter plot of the relation (Figure 19) shows a strong linear relation. The points below the fitted solid line in the figure indicate respondents who scored relatively lower on the more complex items (Factor 2) than expected, based on their responses to the simpler items (Factor 1); those above the line scored relatively higher on the more complex items than expected, based on their responses to the more simple items.

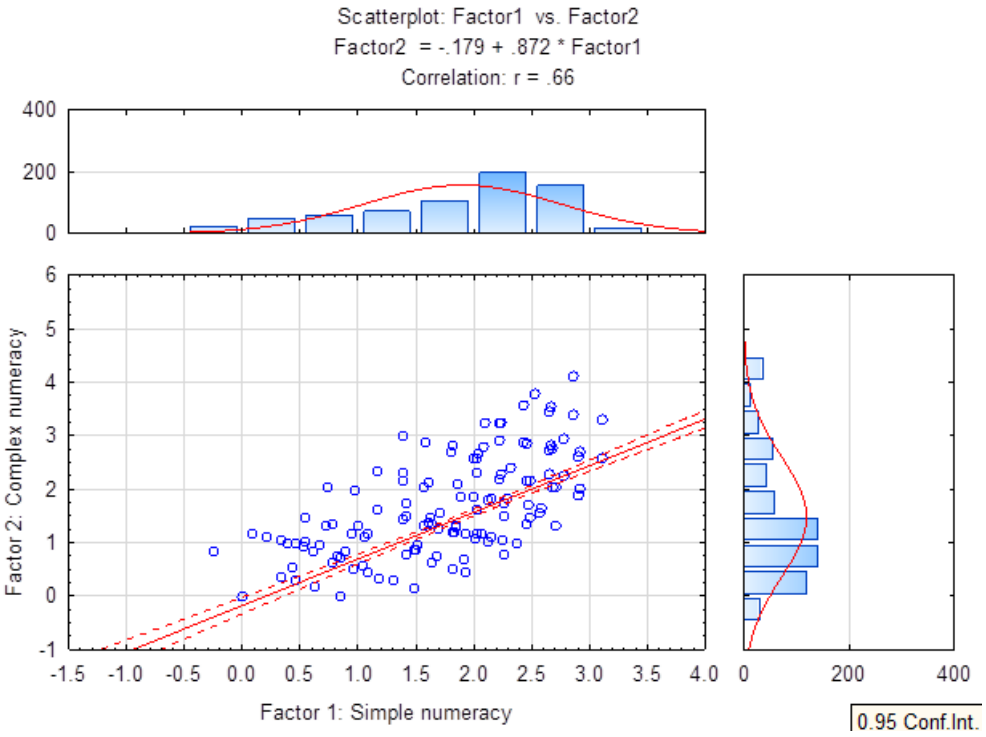


Figure 19: Scatter plot of the relation between the numeracy factors

5.2.3.7. Conclusions based on the factor structure of the numeracy items (sub-questions 1.1 and 1.2)

The first factor measures simpler statistical numeracy based on a single computational step analysis, while the second factor measures more complex statistical numeracy based on more complex probabilistic reasoning and computation. The measurement invariance of these factors ensures the comparability of the factor scores across the demographic groups, even though their factor means are different. Group comparisons computed via the multiple group ESEM approach showed differences between the population groups on the mean scores of the 2-factor solution computed on the nine numeracy items retained. Finally, the factor scores are highly correlated, with both factors contributing to the interpretation of the statistical numeracy of the respondents. The factors are discussed relative to the numeracy literature in Chapter 6.

The results on the numeracy factor scores inform the categorisation of the respondents into the numeracy latent subgroups.

5.2.4. The latent subgroups on statistical numeracy (sub-question 1.3)

As described in Section 4.7.4.3, the factor mixture model was used to categorise the respondents into numeracy groups. The factor mixture algorithm combines the categorical-type latent class analysis in which respondents are assigned to numeracy classes based on categorical latent class scores, with the continuous-type numeracy factor scores which allow for dimensionality and variability (Clark et al., 2013).

5.2.4.1. Selection of the factor mixture model

The researcher followed the systematic approach of Clark et al. (2013) in choosing an interpretable variation of the factor mixture model for categorising the 666 respondents into statistical numeracy groups. Each latent mixture variation is described in Table H 1. The details of these variations are summarised in Table 13, using the criteria for model fit previously defined (Section 4.7.4.3). In addition, in the two right-most columns of Table 13, the final class proportions are presented for the latent classes from each estimated model, followed by an indication of whether the item profiles for the groups of each model are clearly interpretable. Item profiles that were distinct and did not cross over one another were considered as clear.

The criterion of clear item profiles for models with acceptable groupings eliminated most models. The models eliminated on this basis were: first, two of the three models with class invariant thresholds and factor loadings but class varying factor covariance matrices and factor means (FMM-2); second, all of the models with class varying thresholds, class invariant factor loadings, class variant or invariant factor covariance matrices, and factor means fixed at 0 (FMM-3); third, all the model classes defined by class varying thresholds, factor loadings and factor covariance matrices with factor means fixed at 0 (FMM-4). Following these eliminations, 6 of the 14 models remained for consideration of the numeracy groupings.

Thereafter, the criteria of reasonably sized classes and the incorporation of two factors were applied to the remaining models under consideration. The class-size criterion was necessary as very small classes would have limited power for subsequent estimation of model parameters. The 2-factor criterion was adopted in view of the previous evidence supporting a 2-factor solution rather than a single-factor solution. Furthermore, the 3-class 2-factor solution comprising 29%, 19% and 52% of respondents (Table 13) was preferred to the 2-class solution comprising a 22%, 78% split, as it was anticipated that a 3-class solution would provide more variance in the relation between risky choice and numeracy. Finally, the decision to adopt the 3-class 2-factor solution, over the 4-class 2-factor solution comprising an 8.9%, 24.7%, 27%, 39.5% split, was based on two further criteria. First, the 3-class solution was more parsimonious. Second, as outlined in Section 5.3, the analysis of the choice data was conducted separately on the responses obtained from the first and second administrations of the two questionnaires (GAINS1, GLM1, GAINS2 and GLM2). Within these response groups, the representation of the three numeracy classes was considered more adequate for analysis than the four classes (Table 15).

The item profiles of the 3-class 2-factor solution are presented in Figure 20.

5.2.4.2. Interpretation of the numeracy groupings of the selected factor mixture model

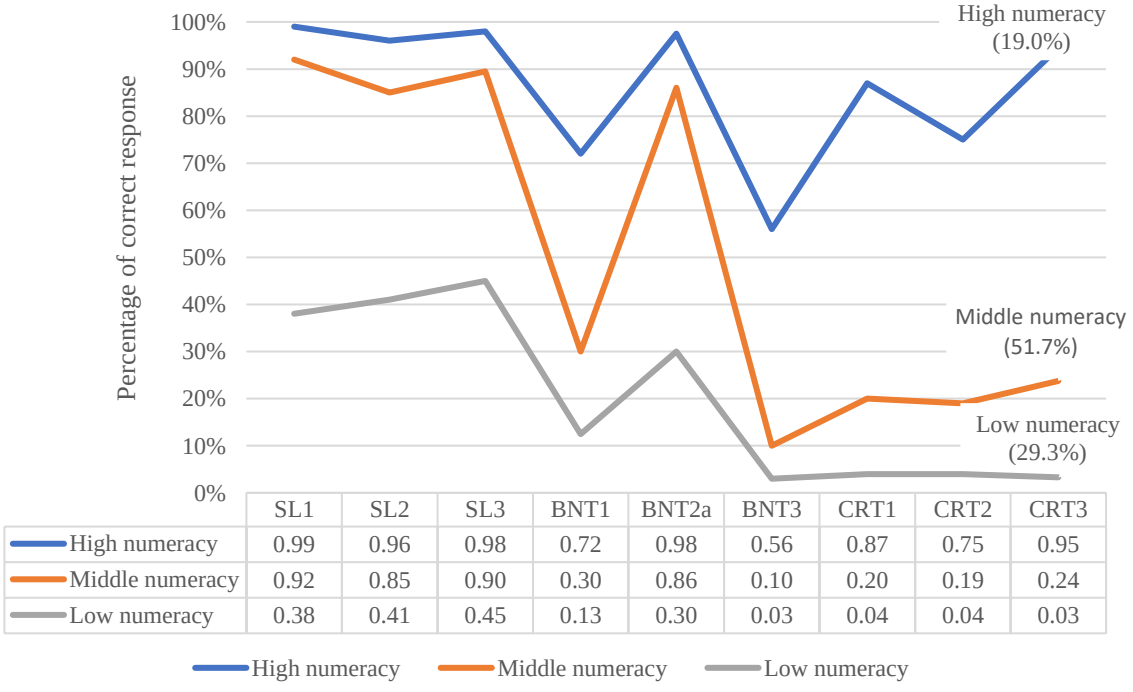
The numeracy groupings are interpreted in terms of the items and their corresponding probabilities of correct responses, per numeracy group. As can be seen from the item response profiles in Figure 20, 19% of respondents were categorised as latent Class 2 (the top line). They scored highest on all items, almost certain to answer correctly the Schwartz/Lipkus Test items, probabilities of 50% or more of correctly answering the Berlin

Numeracy Test items, and probabilities of 75% or more of correctly answering the Cognitive Reflection Test items.

Slightly over half (51.7%) of respondents were categorised as latent Class 3 (the middle line). Their probabilities of correct responses lay between these two classes, with probabilities of at least 85% of correct responses to the relatively simple Schwartz/Lipkus items and the simplest of the Berlin Numeracy Test items (BNT2a), but probabilities ranging from 10% - 30% of correctly answering the two more complex Berlin Numeracy Test items and the Cognitive Reflection Test items.

By contrast, 29.3% of respondents were categorised as latent Class 1 (the lowest line). They scored lowest on all items: on the simplest numeracy items, i.e. those of the Schwartz/Lipkus scale, the probability of their correct responses ranged from 38% - 45%, on the more complex Berlin Numeracy Test items, the probabilities of correct responses peaked at 30%, and were less than 5% on the Cognitive Reflection Test items.

For ease of reference, these groupings are referred to as the high, middle and low numeracy groups corresponding to Class 2, Class 3 and Class1, respectively. This categorisation addresses Research sub-question 1.3, and is examined further in Chapter 6 (see Section 6.1).



	SL1	SL2	SL3	BNT1	BNT2a	BNT3	CRT1	CRT2	CRT3
High numeracy	0.99	0.96	0.98	0.72	0.98	0.56	0.87	0.75	0.95
Middle numeracy	0.92	0.85	0.90	0.30	0.86	0.10	0.20	0.19	0.24
Low numeracy	0.38	0.41	0.45	0.13	0.30	0.03	0.04	0.04	0.03

Figure 20: Item response profiles for numeracy latent classes derived from 3-class 2-factor mixture model (9 items)

Table 13: Summary of latent class, factor analysis and factor mixture models for numeracy groups

Model	Log-likelihood	Free parameters	Akaike (AIC)	Bayesian BIC	Sample-size adjusted BIC	LMR *p* Value	BLRT approx. *p* value	Best log-likelihood value replicated	Entropy	Final class proportions for latent classes from estimated model	Clear item profiles
Latent class analysis											
1-class	-3364.793	16	6747.587	6788.098	6759.523	n/a	n/a	✓			
2-class	-2980.539	19	5999.078	6084.602	6024.276	.000	.000	✓	.767	.366; .633	✓
3-class	**-2895.345**	**29**	**5848.689**	**5979.227**	**5887.150**	**.000**	**.000**	✓	**.724**	**.306; .186; .508**	✓
4-class	-2881.198	39	5840.396	6015.946	5892.119	.705	.000	✓	.690	.105; .246; .240; .408	x
5-class	-2869.929	49	5837.858	6058.421	5902.843	.059	.146	✓	.721	.125; .148; .126; .349; .252	x
Factor analysis											
1-factor	-2911.105	18	5858.209	5939.232	5882.081				.767	.378; .622	
2-factor	-2898.588	26	5849.177	5966.210	5883.659						
Factor mixture analysis											
2-class 2-factors											
FMM-1	-2980.539	22	6005.078	**6104.107**	6034.255	.000	.000	✓	.767	.378; .622	✓
FMM-2	**-2897.272**	**25**	**5844.543**	5957.076	**5877.699**	**.274**		✓	**.484**	**.218; .782**	✓
FMM-3	-2885.002	34	5838.005	5991.049	5883.097			✓	.525	.581; .419	x
FMM-4	-2876.29	39	5830.580	6006.130	5882.303			✓	.640	.267; .734	x
3-class 2-factors											
FMM-1	**-2902.888**	**25**	**5855.775**	**5968.307**	**5888.931**	**.000**	**.000**	✓	**.724**	**.293; .190; .517**	✓
FMM-2	-2891.616	31	5845.232	5984.772	5886.345	.006		✓	.660	.112; .646; .242	x
FMM-3	-2872.698	47	5839.396	6050.956	5901.729			x	.547	.453; .383; .164	x
FMM-4	-2862.163	59	5842.327	6107.903	5920.574			x	.639	.273; .078; .649	x
4-class 2-factors											
FMM-1	**-2890.462**	**28**	**5836.923**	**5962.959**	**5874.058**	**.002**	**.000**	✓	**.659**	**.089; .247; .270; .395**	✓
FMM-2	-2890.491	37	5854.981	6021.529	5904.052	.328		✓	.658	.408; .246; .105; .240	x
FMM-3	-2861.724	60	5843.448	6113.526	5923.022			✓	.636	.375; .046; .199; .380	x
FMM-4	-2850.195	79	5858.390	6213.992	5963.162			✓	.752	.173; .141; .463; .223	x

5.2.4.3. Demographics of the numeracy groupings

The analysis of the demographics of the participants of the latent numeracy groups provided additional information on the groups, and thus responded to the third sub-question (1.3) (Table 14). The biggest difference between the groups on demographic characteristics is on sex. Males represent 42% of the total sample, but are relatively over-represented in the high numeracy group (67%), and relatively under-represented in the low numeracy group (29%). There is a larger representation of participants from the sciences in the middle and high numeracy groups, than in the low numeracy group whose participants are largely reading for the humanities.

Table 14: Demographics of respondents in the numeracy groupings

	Covariate	High numeracy (n = 121)	Middle numeracy (n = 376)	Low numeracy (n = 169)	Total (n=666)
Age	18-21	58%	55%	36%	51%
	22-25	24%	27%	34%	28%
	26-30	12%	11%	15%	12%
	31-40	4%	6%	14%	8%
	41-50	1%	1%	2%	2%
	51 or older	1%	0%	0%	0%
Highest qualification completed	Matriculation	65%	58%	43%	55%
	Undergraduate certificate/diploma	4%	3%	4%	3%
	Bachelors degree	11%	16%	18%	16%
	Honours degree/technology degree	8%	11%	17%	12%
	Postgraduate certificate/ diploma	0%	1%	3%	2%
	Masters degree	10%	10%	14%	11%
	Doctorate/post doctorate	1%	1%	2%	1%
Faculty	Science	56%	50%	27%	45%
	Humanities	14%	23%	62%	32%
	Health Sciences	17%	13%	6%	12%
	Engineering and the Built Environment	10%	6%	1%	5%
	Commerce, Law and Management	3%	9%	4%	6%
Your sex?	Male	67%	40%	29%	42%
	Female	33%	59%	70%	57%
	Prefer not to say/other	0%	2%	1%	1%
Total monthly household income (before tax)	R5,000 or less	13%	30%	33%	28%
	R5,001-R10,000	13%	13%	18%	14%
	R10,001-R35,000	21%	28%	29%	27%
	R35,001-R50,000	13%	13%	9%	12%
	R50,001-R100,000	22%	9%	7%	11%
	more than R100,000	16%	7%	4%	8%
People in the household	1	8%	9%	4%	8%
	2-3	32%	29%	28%	29%
	4-6	50%	50%	54%	51%
	7+	9%	12%	14%	12%
	Total	**18%**	**56%**	**25%**	**100%**

5.2.4.4. Conclusions on the numeracy analysis

The main conclusions pertaining to the research sub-questions on statistical numeracy are:

Sub-question 1.1: The 11 numeracy items originally considered to reflect the numeracy groupings were reduced to 9, to eliminate redundancy and facilitate an efficient set of items for grouping respondents into latent numeracy classes. Factor analysis of these items revealed a 2-factor numeracy structure reflecting a simpler numeracy ability factor and a more complex one.

Sub-question 1.2: The factor structure showed measurement invariance for the covariates of sex, qualification, income and age. Thus, factor scores could be compared across the demographic groups.

Sub-question 1.3: The bi-factor structure of numeracy was incorporated in a factor mixture model analysis which assigned respondents to latent class groupings. The factor mixture model judged as best revealed three distinct latent numeracy classes: a low numeracy class (29%) with most respondents scoring poorly on relatively simple as well as on more complex and difficult numeracy items, i.e. scoring relatively poorly on both numeracy factors; a middle numeracy class (52%) with most respondents scoring well on relatively simple numeracy items (Factor 1), but poorly on the more difficult items (Factor 2); a high numeracy group (19%) with most respondents scoring both simple and more complex items correctly, i.e. scoring relatively highly on both factors.

These numeracy groupings were used throughout the remaining analyses of the study.

This section completes the numeracy analysis. The following sections are devoted to modelling the risky decisions of respondents in the three numeracy groups, and thus address the second research question.

5.3. Analysis of risky choice responses

This section of the chapter presents analyses of the risky choices of the respondents in the three numeracy groups. The research question and sub-questions on risky choice are restated:

Research question 2: What risk attitudes and probability distortions are associated with the risky choices of decision makers of different numeracy groups?

2.1: What risk attitudes are associated with the risky choices of decision makers of different numeracy groups?

2.2: To what extent do the estimated values of the parameters of the normative and descriptive risky choice models considered differ across numeracy groups?

2.3: To what extent do the estimates and predictions of the models hold for the risky choices of decision makers of different numeracy groups?

Sections 5.3.1 and 5.3.2 provide preliminary analyses of the choice data of the participants of the three numeracy groups, without referring to the structural risky choice models. The first of the preliminary analyses was a simple frequency analysis of the number of safe lotteries chosen per numeracy group (5.3.1). The second of the preliminary analyses was a repeated measures analysis of variance of safe choices (5.3.2). These sections thus address Research sub-question 2.1. The remaining sections of the chapter focus on the modelling and validation of the parameters of the structural risky choice models, per numeracy group. They address Research sub-questions 2.2 and 2.3 as follows: maximum likelihood estimation of the parameters computed on a subset of the data (5.3.3); hierarchical Bayesian analysis computed on the remaining data using the maximum likelihood estimates of Section 5.3.3 for the priors of the structural models for each numeracy group, when available (Section 5.3.4); assessment of the convergence of the MCMC chains and the fit statistics of the hierarchical Bayesian models (5.3.4.1); and validation of the estimates and the predictions of the models (5.3.5).

The numbers of useable responses received to the first and second administrations of the questionnaires, per numeracy group, are presented in Table 15.

Table 15: Numbers of useable responses to the first and second administrations of the questionnaires, per numeracy group

Questionnaires	High numeracy	Middle numeracy	Low numeracy	Total
GAINS1	3	15	11	29
GLM1	10	45	43	98
GAINS2	44	146	50	240
GLM2	64	170	65	299
Total	**121**	**376**	**169**	**666**

For convenience, the definitions of risk attitudes provided by Fox and Poldrack (2009) are presented here. An individual is risk neutral if she is indifferent between a gamble and its expected value, risk averse if she prefers a sure payment to a risky prospect with a higher or

equal expected value, and risk seeking if she prefers a risky prospect with a lower or equal expected value, to a sure payment.

5.3.1. Preliminary analyses of risky choices: Frequencies

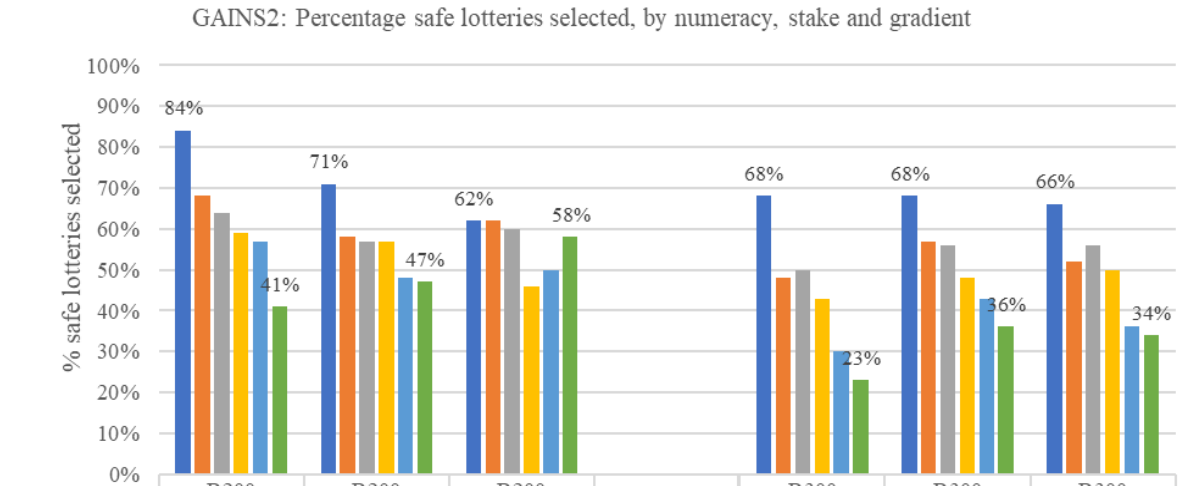
The preliminary analyses of risky choices were carried out on the responses to the GAINS2 and GLM2 questionnaires only, as there were insufficient responses to the GAINS1 and GLM1 questionnaires for analysis per numeracy group (Table 15). In the GAINS2 and GLM2 analyses, the researcher attempted to control for random response error by considering the unit of analysis as the *total number of safe lotteries chosen within the four lottery pairs* of each of the Marschak-Machina triangles. It was assumed that responses aggregated per triangle would be more stable than individual choices. As described in Section 4.5.2, the triangles used to develop the risky choices for each questionnaire differed on the slopes, or gradients, of their linear indifference curves, i.e. on risk aversion under expected utility theory, with other characteristics held constant as far as possible (Loomes & Sugden, 1998; Machina, 1987).

The researcher categorised a respondent's four choices within a triangle collectively as 'safe' if the respondent had selected the safer option within a choice pair at least 3 times out of 4. Each respondent thus received 12 categorised scores: either for 2 payout contexts or stakes (R0, R100, R200 and R0, R100, R300) for 6 gradients in the case of the Gains2 questionnaire, or for 3 frames (gains, losses and mixed prospects) for 4 triangle gradients in the case of the GLM2 questionnaire. Thereafter, the categorised responses were grouped by payout context and gradient, per numeracy group for the GAINS2 frequency analysis, and by frame and gradient, per numeracy group for the GLM2 frequency analysis. The graphic displays of the percentages of safe choices are presented in Figure 21 for the categorised responses to the GAINS2 questionnaire, and in Figure 22 for the categorised responses to the GLM2 questionnaire.

The groups of bars in Figure 21 show a pattern of decreasing choices categorised as safe within Marschak-Machina triangles as the gradients of the linear indifference curves increase from 0.67 to 4. Increasing gradients of indifference curves mean that the riskier options of the choice pairs become increasingly attractive, relative to their counterparts. In spite of the decreasing trend of the percentage of safer choices with increasing slopes in the R200 context, safer choices outweigh riskier ones apart from for the most attractive bets (highest gradients). For choices in triangles with linear indifference curve gradients of 1 - 4 in the

R200 context or stake, the expected value of the riskier option was equal to, or greater than, the expected value of the safer option. Thus, for these choices, selecting the safer option indicated risk aversion, increasingly so with increasing slope. On average, in this context, the respondents were risk averse by selecting the safer option more often than the riskier, other than in the most attractive choices (gradient of 4). In the R300 context, the riskier option of every pair had the higher expected value. Risk aversion was less obvious here (mostly for the least attractive of the R300 bets, with gradient of ⅔).

The decreases in safer choices, and in risk aversion, are most obvious for the high numeracy group, and least obvious for the low numeracy group. The more numerate respondents appear to discriminate more between the attractiveness levels of the bets. The trend is also generally more obvious when the highest payout is R300 rather than R200. A more thorough analysis of these effects is presented in Section 5.3.2.

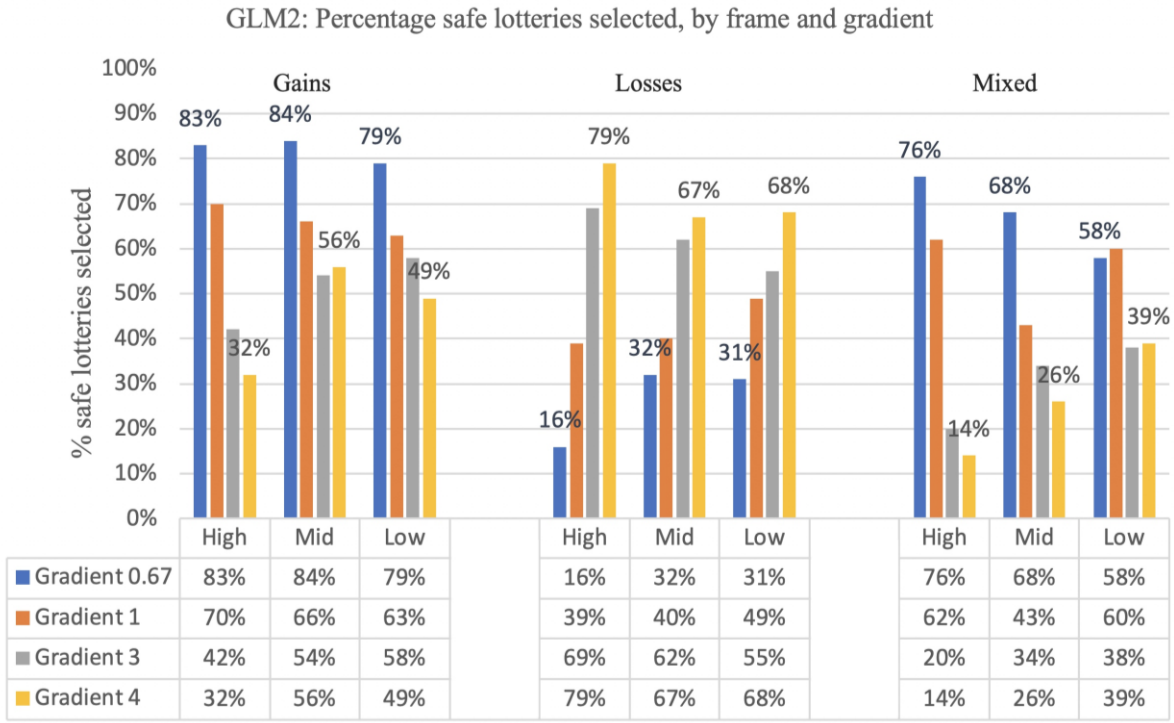


	R200 High	R200 Mid	R200 Low	R300 High	R300 Mid	R300 Low
Gradient 0.67	84%	71%	62%	68%	68%	66%
Gradient 1	68%	58%	62%	48%	57%	52%
Gradient 1.5	64%	57%	60%	50%	56%	56%
Gradient 2	59%	57%	46%	43%	48%	50%
Gradient 3	57%	48%	50%	30%	43%	36%
Gradient 4	41%	47%	58%	23%	36%	34%

Figure 21: Percentage of safe lotteries selected in responses to the GAINS2 questionnaire, per numeracy group and Marschak-Machina triangle

For responses to the GLM2 questionnaire, the groups of bars in Figure 22 show a similar pattern of decreasing percentages of safer bets selected, with increasing indifference curve gradients, for gain-framed and mixed-framed prospect pairs. Once again, this pattern is particularly noticeable for respondents in the high numeracy group. Thus, more risky options are selected with increasing attractiveness of the riskier option within choice pairs. For

choices with linear indifference gradients of 1, 3 and 4, the expected values of the riskier options are at least as high as the expected values of the safer options. Thus, selecting the safer options of these choices indicates risk aversion. For these choices, apart from the neutral choices (gradient of 1), there is little evidence of risk aversion for highly numerate respondents as they selected the riskier options more often, on average. There is some evidence of risk aversion in these choices for the other numeracy groups.



	Gains			Losses			Mixed		
	High	Mid	Low	High	Mid	Low	High	Mid	Low
Gradient 0.67	83%	84%	79%	16%	32%	31%	76%	68%	58%
Gradient 1	70%	66%	63%	39%	40%	49%	62%	43%	60%
Gradient 3	42%	54%	58%	69%	62%	55%	20%	34%	38%
Gradient 4	32%	56%	49%	79%	67%	68%	14%	26%	39%

Figure 22: Percentage of safe lotteries selected in responses to the GLM2 questionnaire, per numeracy group and Marschak-Machina triangle

In the loss frame, choices that involve greater differences in expected *losses* between the riskier and safer options of a pair are associated with increasingly steep indifference curve gradients. As shown in the middle set of bars in Figure 22, the percentage of safer choices increased as the gradients of the indifference curves increased from 0.67 to 4 for losses. Once again, this pattern appears accentuated for respondents in the high numeracy group. Choices with slopes of 1, 3 and 4 have risky options with expected values equal to – or lower than – those of their safer counterparts (i.e. more loss), and so selecting the riskier option for these choices indicates risk seeking. As safer choices outweigh riskier ones for choices with higher slopes (3 and 4), there is little evidence of risk seeking for losses. Evidence of risk seeking is for the neutral loss-framed prospects.

These observations were made based on graphic representation of percentages of safe choices. In the following section (5.3.2), they are re-addressed formally in repeated measures analyses which assume that the dependent variable of the total number of safe options (0 – 4) selected in the choices within a triangle is a continuous variable.

5.3.2. Preliminary analyses of risky choices: Repeated measures

A *2-way repeated measures mixed-model analysis of variance* was used for the analysis of risk aversion. The '2-way' aspect of the model derived from the two treatment effects used in the repeated-measures design: for example, the treatment effects in analysing responses to GLM2 were the *frame* of the prospect (gains, losses, mixed), and the *gradient* or slope of the linear indifference curves of the Marschak-Machina triangles representing the four risk aversion levels. The 'repeated measures' aspect was due to each subject responding to the four levels of risk aversion within each frame (three sets of four responses). The 'mixed-model' aspect derived from the 3-level numeracy factor (high, middle and low) that served as the independent measures factor in the model. This analysis method and the supporting graphics were computed using TIBCO Statistica software version 13.6.0 (TIBCO, 2019). It is acknowledged that mixed-effects logistic regression analysis may have been preferable to the analysis method used, but was not selected in view of the complexity of the repeated measures design. The linear mixed model was also not selected for the preliminary analyses in view of its complexity.

5.3.2.1. Responses to the GAINS2 questionnaire

Before embarking on the repeated measures analysis of the responses to the GAINS2 questionnaire, the assumptions of the analysis model were checked.

Assumption checks

The assumptions of the repeated measures analysis of variance were met on the following criteria: the dependent variable was considered a continuous ratio variable as the respondent could score between 0 and 4 for the choices within a Marschak-Machina triangle, and means were interpretable and could be calculated on this variable across sets of triangles, or across individuals; the independent variables of gradient, highest payout, and numeracy categories were categorical; observations were independent as choice pairs were presented in random order across triangles and the triangles themselves were presented in a random order; and the normality of the sampling distributions of responses was assumed owing to the samples sizes

exceeding the commonly used limit of 30 (44, 146, and 50 for the high, middle and low numeracy groups, respectively). Furthermore, all of the 12 score distributions of the dependent variables (six triangles for two payouts combinations) met the assumption of homogeneity of variance based on the Hartley, Cochran and Bartlett Chi-square tests.

However, the assumption of sphericity was not met, as the Mauchly sphericity tests were significant for the indifference curve gradient repeated measures effect, and the interaction of payout and gradient ($\chi 2(14) = 51.96$ and 36.78 , $p < .0001$, respectively). [Sphericity assumes that the variances of the difference scores between treatment levels – the gradients of the triangles in this case – are equal, i.e. that the correlation between treatment levels are approximately equal (Field, 2013).] To overcome the sphericity violations, the more conservative Greenhouse-Geisser/ Huynh-Feldt corrected results are presented (Table 16).

Risk attitudes to characteristics of gain-framed prospects

The results of the analysis (Table 16) show significant but weak interaction effects for payout*gradient and for gradient*numeracy ($p = .041$ and .033, respectively), with associated effect sizes [partial Eta squared (η_p^2) = .01 and .02, respectively]. The payout*numeracy interaction effect was more highly significant ($p = .007$), with slightly higher, although still weak effect size ($\eta_p^2 = .04$). The main effects of payout and gradient were moderate-to-strong, with η_p^2 values of .11 and .17, respectively. Accordingly, the interpretation is focused on the plots of the least squares means for the payout*numeracy interaction effect (Figure 23), and for the plot of means for the main effect of gradient (Figure 24). The means for the main effect of payout are presented and discussed in the text.

The payout*numeracy interaction plot (Figure 23) compares the mean number of safe lotteries chosen when the highest payout of the choice pairs was R200 versus R300. The difference was greatest for the high numeracy group, meaning that the respondents in the high numeracy group had the greatest change from safer to riskier preferences when the highest stake, or payout, of the lottery pair was increased. The effect is noted, albeit weak.

The mean safe choice scores for the main effect of payout were 2.62 and 2.38 out of 4, showing slightly more safe lotteries chosen when the highest payout was R200 than when the highest payout was R300. Thus, the higher stake (R300 compared to R200) accounted for more choices of the riskier option across respondents.

Finally, the plot of the main effect of gradient (Figure 24) shows that the mean number of safe choices selected decreases, as the attractiveness of the bets increase. There is thus evidence of overall risk aversion which decreases as the bets become more attractive.

Table 16: 2-way repeated measures mixed-model analysis of variance on responses to GAINS2 questionnaire

Effect	SS	df adjusted *	MS	F	*p*	η_p^2	Observed power (alpha = .05)
Intercept	13620.11	1	13620.11	1742.82	<.001	0.88	1
Numeracy	1.04	2	0.52	0.07	0.936	0.00	0.06
Error	1852.15	237	7.81				
Payout	29.99	1	29.99	28.46	<.001	0.11	1
Payout*numeracy	10.68	2	5.34	5.07	0.007	0.04	0.816
Error	249.75	237	1.05				
Gradient	152.88	4.56	30.58	48.22	<.001	0.17	1
Gradient*numeracy	12.85	9.12	1.29	2.03	0.033	0.02	0.893
Error	751.45	1080.23	0.63				
Payout*gradient	6.35	4.7	1.27	2.37	0.041	0.01	0.759
Payout*gradient*numeracy	6.83	9.41	0.68	1.27	0.245	0.01	0.667
Error	636.53	1114.74	0.54				

* df and *p* values adjusted using Greenhouse-Geisser correction

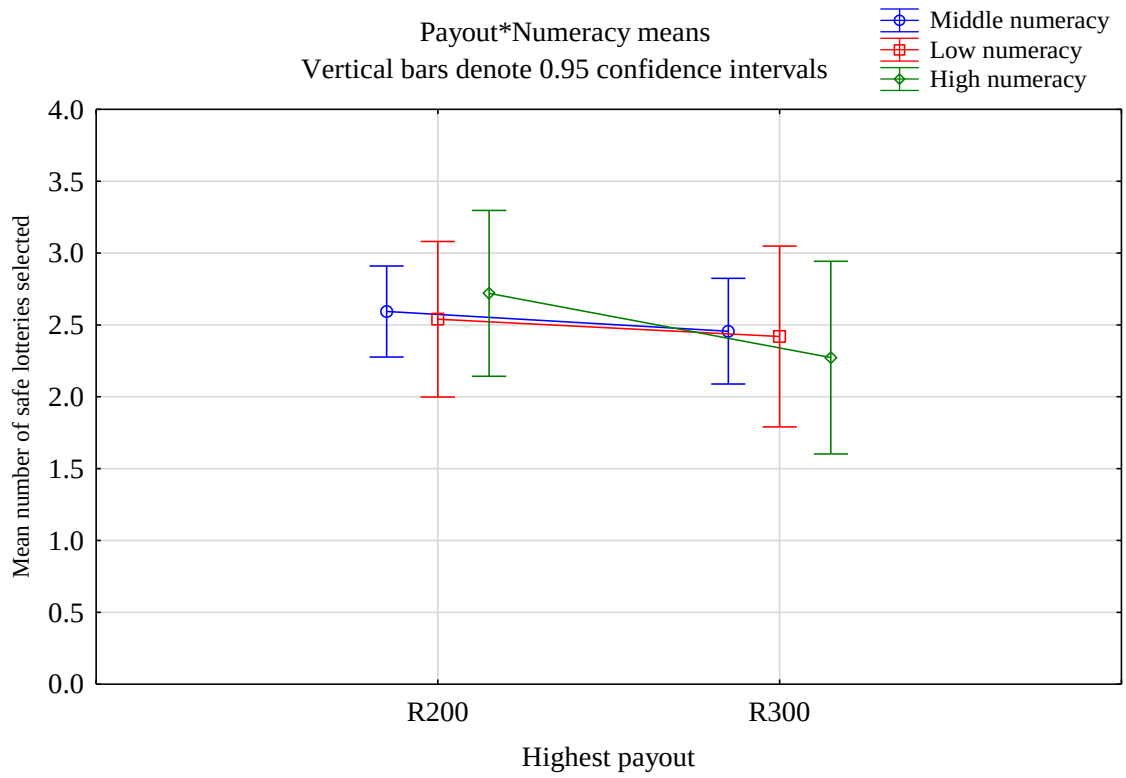


Figure 23: The payout*numeracy effect for GAINS2 responses

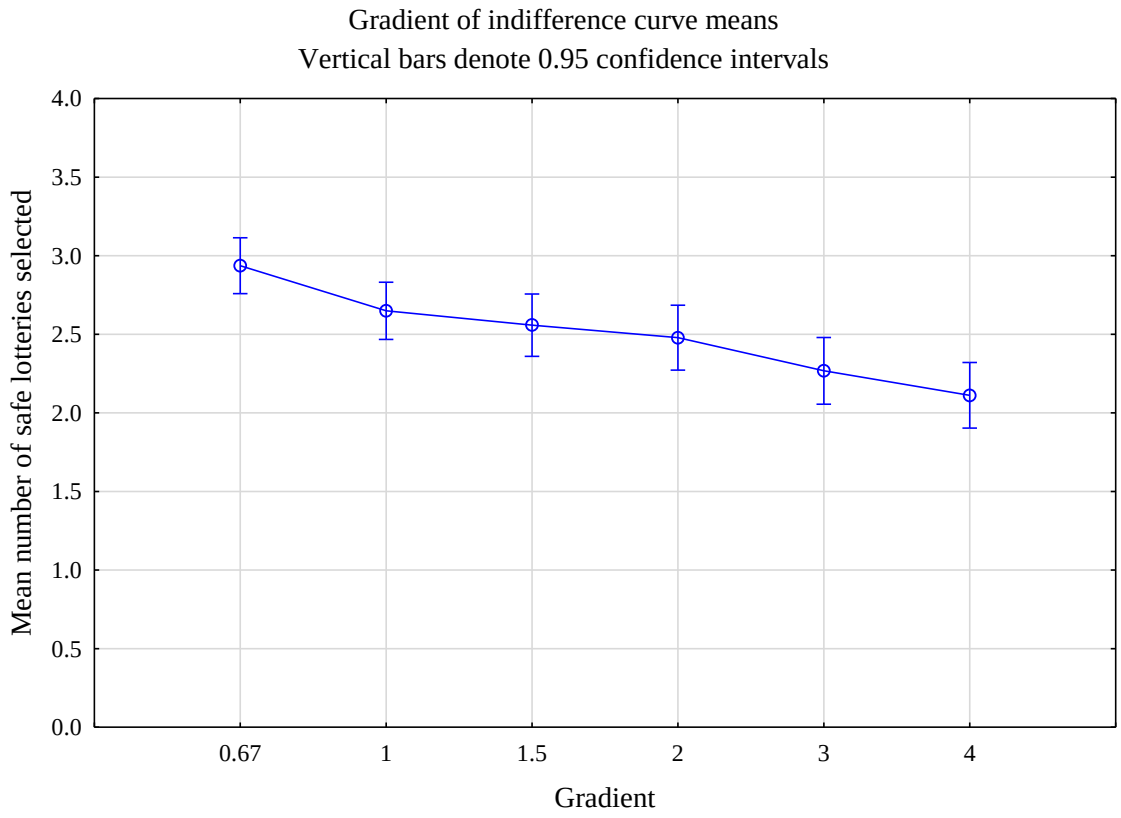


Figure 24: The effect of gradient of the indifferent curves

In summary, on average, respondents preferred the safer option within a choice, reflecting risk aversion. A higher percentage of safer options were selected within choice pairs of triangles from flatter to steep indifference curve gradients. The percentage of times that the safer option of a choice was selected over its riskier counterpart, decreased in two instances: when the highest outcome of the prospect was increased from R200 to R300, and when the gradient of the indifference curves were increased, i.e. when the attractiveness of the bets were increased. These effects were greatest for the highly numerate respondents, although weakly so.

A similar analysis was computed on the responses to the GLM2 questionnaire. The difference between the two analyses was that the repeated measures payout effect (R200, R300) was replaced by the frame effect (gains, losses, mixed).

5.3.2.2. Responses to the GLM2 questionnaire

The researcher carried out the same process of analysis on the responses to the GLM2 questionnaire, before computing the 2-way repeated measures mixed-model analysis of variance. As for responses to the GAINS2 questionnaire, the dependent variable was the total number of times out of 4 that the safe options were chosen in choices represented by

indifference curves with the same gradient. The independent variables were frame, gradient and numeracy.

As before, the assumptions of the analysis model were checked at the outset.

Assumption checks

Once again, the assumptions of the repeated measures analysis of variance were met in terms of the dependent variable assumed to be measured on a continuous scale, the categorical nature of the independent variables, the independence of the observations, and the assumption of normality of the sampling distributions (based on large sample sizes of 64, 170 and 65 for the high, middle and low numeracy groups, respectively).

However, 2 of the 12 score distributions (four triangles for three frames) did not meet the assumption of homogeneity of variance based on the Hartley, Cochran and Bartlett Chi-square tests, and thus the results of this analysis need to be inferred with caution. The two violations were for the gain-framed prospects in the triangle with a gradient of 0.67, and the mixed-framed prospects in the triangle with a gradient of 4. As in the GAINS2 analysis, the assumption of sphericity was violated for the repeated measures effects of frame, gradient of the indifference curves of the triangles, and their interaction ($\chi2(2) = 28.50$, $\chi2(5) = 16.30$ and $\chi2(20) = 84.01$, $p < .0001$, respectively). The more conservative Greenhouse-Geisser corrected results were adopted to overcome the violations, as presented in Table 17.

Risk attitudes to characteristics of prospects in gain, loss and mixed frames

Table 17 shows a significant frame*gradient*numeracy interaction effect ($p < .001$) with high power and weak-to-moderate effect size ($\eta_p^2 = .07$). The corresponding interaction plot in Figure 25 shows that *in the gain and mixed frames*, when prospect pairs had steeper indifference curves, the safer option of a lottery pair was selected less often, reflecting more risky choices. Thus, the pattern of gain- and mixed-frame choices showed that as the expected value difference between the safer and riskier options increased, the riskier prospect became more attractive and the safer prospect was chosen less often. This pattern was most pronounced for high numeracy respondents, and least pronounced for the low numeracy respondents. Furthermore, none of the mean scores of the low numeracy group fell below the midpoint of 2 on the 0 - 4 scale of the safer choice, thus implying that low numeracy respondents were risk averse, on average. For the high numeracy respondents, mean responses all fell below the value of 2, on the 0 - 4 scale of the safer choice, for the choices with the steepest indifference curves, implying lower risk aversion.

Table 17: 2-way repeated measures mixed-model analysis of variance of responses to GLM2 questionnaire

Effect	SS	df adjusted +	MS	F	*p*	η_p^2	Observed power (alpha = .05)
Intercept	17024.77	1	17024.77	4349.24	<.001	0.94	1
Numeracy	17.24	2	8.62	2.20	0.112	0.02	0.448
Error	1158.67	296	3.91				
Frame	93.90	1.83	46.95	23.74	<.001	0.07	1
Frame*numeracy	7.85	3.66	1.96	0.99	0.407	0.01	0.315
Error	1170.66	542.08	1.98				
Gradient	38.32	2.9	12.77	20.26	<.001	0.06	1
Gradient*numeracy	12.32	5.79	2.05	3.26	0.004	0.02	0.933
Error	559.75	857.07	0.63				
Frame*gradient	533.19	5.41	88.87	118.20	<.001	0.29	1
Frame*gradient*numeracy	92.11	10.81	7.68	10.21	<.001	0.07	1
Error	1335.28	1600.45	0.75				

Note: + df and *p* values adjusted using Greenhouse-Geisser correction when the sphericity assumption is violated

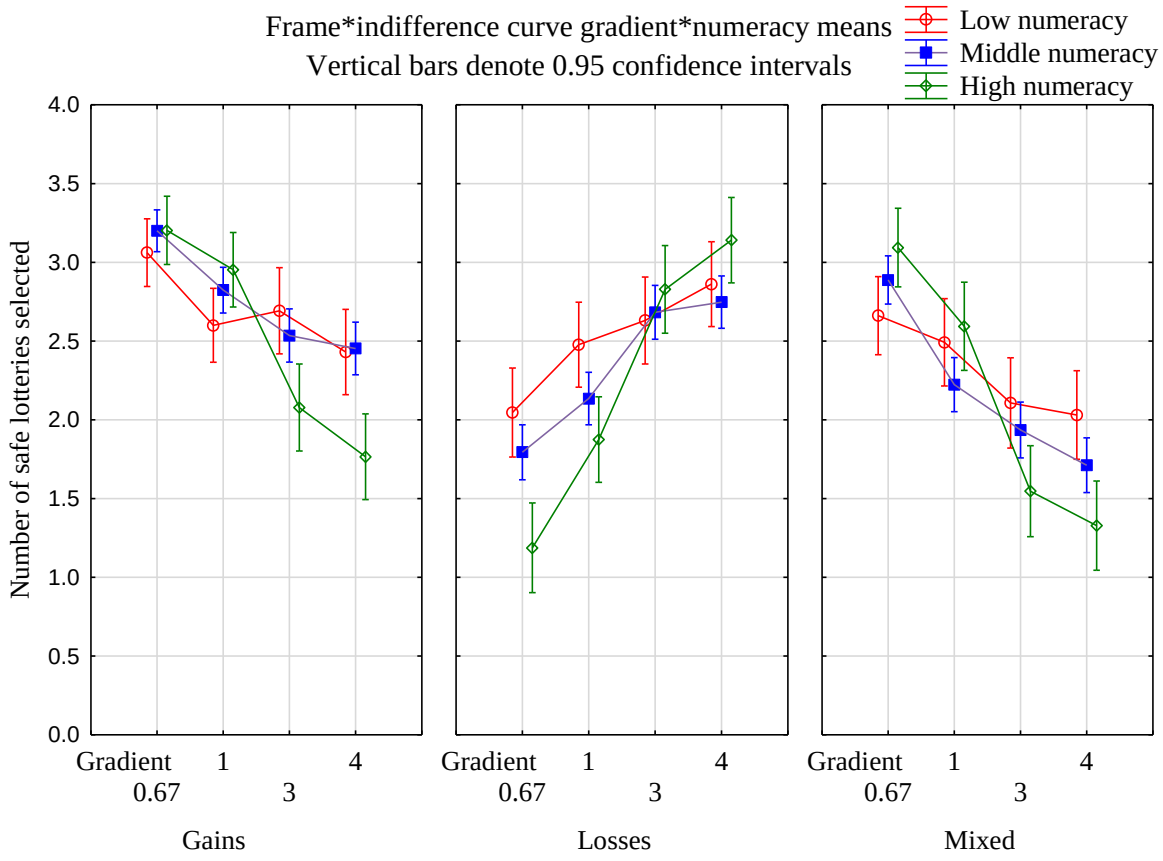


Figure 25: The frame*gradient*numeracy effect for GLM2 responses

By contrast, the pattern of *loss*-framed choices showed that as the difference between the probabilities of losing the highest outcomes of the prospect pair increased, the safer prospect became more attractive and was chosen more often. Once again, the pattern was the most pronounced for high numeracy respondents, less so for the middle numeracy respondents, and least for the lowest numeracy respondents. Based on the means of the safe choice scale, low numeracy respondents, on average, tended to select the safer options in the loss frame, increasingly so with increasingly steep indifference curve gradients. This pattern was most pronounced for the highly numerate respondents (means for safe responses above 2) for choices of higher gradients. Only in the case of the neutral bets, with equal expected values for the riskier and safer options, was there evidence of risk seeking.

In summary, the overall pattern of the safe choice mean values suggests that choices reflecting risk attitudes must be considered in the context of both the choice frame (gains, losses and mixed), and the gradient of the indifference curve of the choice pair (the attractiveness of the bet). The frame*gradient effect is the strongest treatment effect, accounting for almost 30% of the variation in choices, with the other factors controlled (η_p^2 = .29) (Figure 26). Furthermore, although choices generally reflect a preference for the safer option of a bet, the extent of this attitude *depends on the numeracy level of respondents*. On average, for all choices considered, low numeracy respondents prefer the safer option of a bet. By contrast, on average, high numeracy respondents select safer choices in gain- and mixed-framed bets, but change their preference to riskier choices as the attractiveness of the riskier options of these bets are increased. Compared to lower numeracy respondents, they are more sensitive to differences in the attractiveness (probabilities) of the bets. In loss-framed bets, they select the safer option as the difference in the expected losses within a choice increase. They show evidence of risk seeking in the case of neutral bets.

This concludes the exploratory analyses of the risky choice data. The next section presents the first step in the process of estimating the parameters of the risky choice models. For all these analyses, the unit of analysis was the risky choice within each prospect pair.

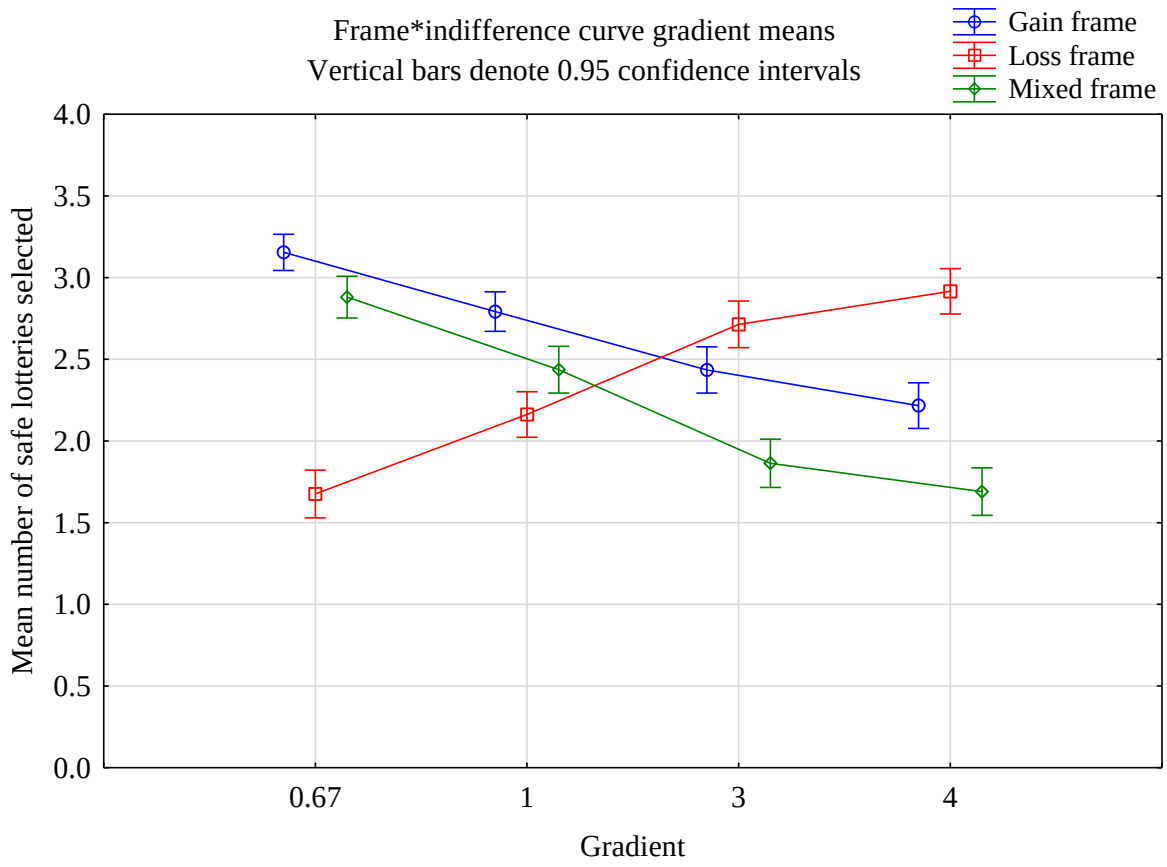


Figure 26: The frame*gradient effect for GLM2 responses

5.3.3. Maximum likelihood estimation of model parameters

As discussed in Section 4.8, the general approach used to estimate the models' parameters was to apply Wilcox's contextual error in the analyses to control random response error, and Bayesian analysis to control random parameter error in the estimation procedures. Ideally, Bayesian analyses per numeracy group are based on sets of informed prior values for the models' parameters per numeracy group. However, in line with the knowledge gap identified and the motivation for the thesis, no study could be found with adequate estimates of the model parameter values for decision makers of different numeracy levels. Thus, the researcher conducted exploratory maximum likelihood estimates on subsamples of the risky choice responses of the high, middle and low numeracy groups. Where possible, the resultant estimates were used for the prior parameter distributions in the Bayesian analyses conducted on the remainder of the responses.

The researcher was careful to ensure that there were no respondents whose scores were common to both the maximum likelihood estimates and the Bayesian estimates of the same model parameters. For each numeracy group, the maximum likelihood estimates computed

on the data from the responses to the gain-framed prospects of the GAINS1, GLM1 and GLM2 questionnaires were used as Bayesian priors for estimating the corresponding model parameters based on the GAINS2 responses. The maximum likelihood estimates, derived from all the responses to the GAINS1, GLM1 and GAINS2 questionnaires, were used as Bayesian priors for estimating the corresponding model parameters based on the GLM2 responses.

Maximum likelihood parameter estimates were carried out for each model and functional form specified in Table 3. The analyses were computed using Stata software (StataCorp, 2017), following the guidelines and adapting the Stata code supplied by Harrison (2008). As the residuals per respondent were assumed correlated, the robust standard errors were adjusted for clusters.

The researcher included the probability weighting function of Lattimore et al. (1992) under the rank-dependent model, with the dimensionality of the parameters reduced by assuming linear utility. This modelling followed the recommendation of Wakker (2010) to reduce the dimensionality of the parameters when decision models only approximate the data, rather than fit it well. Following this advice (P.P. Wakker, personal communication, June 20, 2019), the rank-dependent model was tested using both the Lattimore et al. (1992) probability weighting function with linear utility, and the Tversky and Kahneman (1992) function with power transformed utility.

The results of the maximum likelihood analyses were generally disappointing. In the analyses for the middle and low numeracy groups under the cumulative prospect model with the power utility function and the Tversky and Kahneman (1992) probability weighting function, parameter solutions were deemed unfeasible, or numerical derivatives could not be computed owing to discontinuous regions with missing values. Likewise, maximum likelihood estimates were not found under the rank-dependent model with linear utility and the probability weighting function of Lattimore et al. (1992). Furthermore, in several analyses, the Wilcox contextual error needed to be excluded from the estimates due to convergence problems. The exclusion of this error model from the maximum likelihood estimates was not considered overly serious, as the maximum likelihood estimates were used as the priors of the subsequent Bayesian analyses, which incorporated the contextual error model using different sets of data (responses to the GAINS2 or GLM2 questionnaires). Details of the models, prospect frames used, functional forms, and the results obtained are provided in Table 18.

Table 18: Maximum likelihood estimates per numeracy group

Model	Log pseudolikelihood	functional form	numeracy group	n	clusters	parameter	value	robust std. error adjusted for clusters	[95% conf. interval]	
EUT Gains	-905.97	Power	high	1556	77	alpha	0.439	0.012	0.415	0.462
	-2880.5		middle	4700	229		0.348	0.013	0.323	0.372
	-1570.45		low	2422	118		0.231	0.038	0.157	0.304
RDU Gains and losses	-1682.94	(Tversky & Kahneman, 1992)	high	2634	57	alpha	0.347	0.018	0.312	0.383
						gamma	0.851	0.042	0.768	0.934
	-6126.73		middle	9317	206	alpha	0.255	0.013	0.229	0.281
						gamma	0.896	0.040	0.817	0.974
	-2910.96		low	4305	104	alpha	0.098	0.045	0.009	0.187
						gamma	0.875	0.077	0.724	1.027
RDU Gains	-900.18	(Tversky & Kahneman, 1992)	high	1556	77	alpha	0.450	0.014	0.334	0.391
						gamma	0.900	0.042	0.757	0.920
	-2877.74		middle	4700	229	alpha	0.351	0.013	0.326	0.376
						gamma	0.939	0.029	0.883	0.995
	-1569.78		low	2422	118	alpha	0.234	0.037	0.161	0.308
						gamma	0.928	0.064	0.803	1.052
CPT Gains, losses and mixed	-1879.59	(Tversky & Kahneman, 1992)	high	2795	57	alpha	0.063	0.065	-0.065	0.191
						lambda	1.262	0.501	.280	2.244
						gamma	0.791	0.135	0.526	1.056

Notes:

i. CPT for middle and low numeracy groups, and RDU (Lattimore): could not calculate numerical derivatives – discontinuous region with missing values encountered.

ii. EUT and RDU analyses for gains computed on responses to GAINS1, GLM1 (gains prospects), GLM2 (gains prospects) questionnaires.

iii. RDU analyses for gains and losses computed on responses to GAINS1, GLM1 (gains and losses), GAINS2 questionnaires.

Several of the maximum likelihood results in Table 18 are noted: First, for all the numeracy groups and models (other than the Lattimore function with assumed linear utility), the estimated values of the alpha parameter of the power utility function are consistently low relative to the value of 1 which represents linear utility. As alpha values less than 1 represent risk aversion, there is consistent evidence of risk aversion in the responses of the participants. Furthermore, for the analyses of models in which estimates were obtained for all numeracy

groups, risk aversion is relatively lowest for the high numeracy group, slightly higher for the middle numeracy group, and highest for the lowest numeracy group, with all pairwise differences significant ($p < .001$).

Second, the estimated probability weighting parameter (gamma) values for the rank-dependent utility (RDU) and cumulative prospect models (CPT) using the Tversky and Kahneman (1992) probability weighting function, approximate the expected *S*-shaped pattern ($\gamma < 1$) for all the numeracy groups for which maximum likelihood estimates were obtained. For the rank-dependent model, the differences in gamma values between the numeracy groups is not significant ($p > .05$), although the gamma values for the high numeracy group are marginally lower than those of the other numeracy groups. Examples of the weighting curves are shown for gamma values of 0.8 and 0.95 in the left and right panes of Figure 27, respectively, relative to the linear function. As discussed in Section 2.2.2, under rank-dependence a probability weighting power function with inverse *S*-shaped curvature reflects the risk attitudes of optimism for low probabilities ($p < .4$), pessimism for high probabilities ($p > .6$).

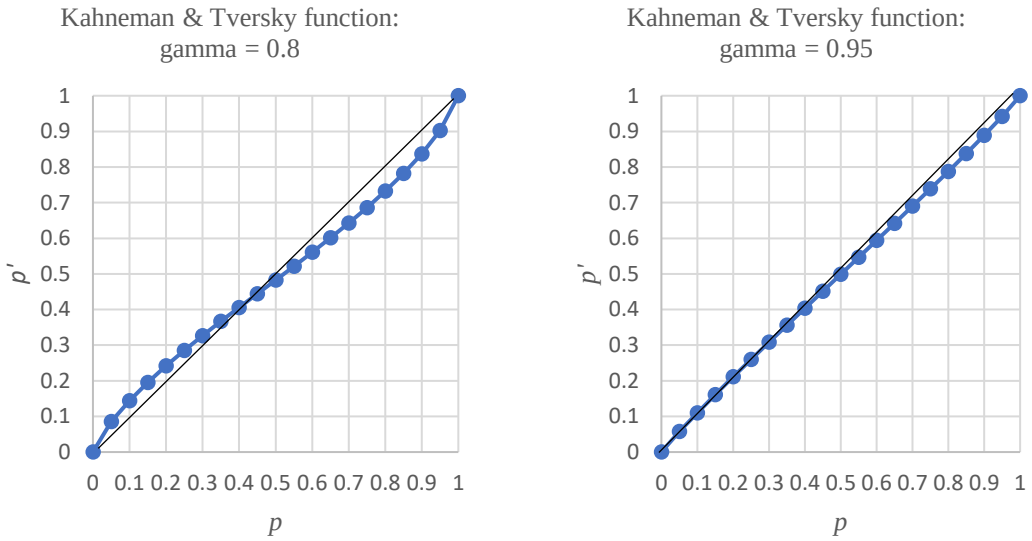


Figure 27: Probability weighting functions using Kahneman and Tversky (1992) with maximum likelihood estimates of gamma for high, middle and low numeracy groups

Third, the estimated loss aversion parameter (lambda) for the cumulative prospect model is greater than 1 for the high numeracy group, indicating the expected loss aversion phenomenon of losses weighing more heavily on the decision maker than gains of the equivalent amounts. Maximum likelihood solutions were not obtained for the lower numeracy groups.

In the next section, the maximum likelihood estimates are used as the parameters of the prior distributions of hierarchical Bayesian analysis, computed via the MultiBUGS software (Goudie et al., 2017).

5.3.4. Hierarchical Bayesian estimation of model parameters

As previously noted (Section 4.8.4.2), Bayesian modelling treats parameters as random variables with their own parameter distributions for each respondent, thus rendering the risky choice models as probabilistic rather than deterministic. Furthermore, hierarchical Bayesian modelling enables the phenomenon of shrinkage. With shrinkage, parameter estimates at the individual level are constrained by higher-level parent distributions for the corresponding parameters that effectively pull wayward individual estimates towards the central tendency estimates of other respondents. An example, specific to the context of the thesis, is that the estimated curvature parameter (gamma) of the weighting function for a respondent with low numeracy would be constrained by the gamma parameter distribution of the group of low numeracy respondents.

In practice, the implementation of the hierarchical Bayesian estimation was challenging. The researcher followed the example of hierarchical Bayesian analysis supplied by Nilsson et al. (2011a) and their code for the WinBUGS software. Her revisions included extending the Nilsson et al. modelling code from 2-outcome prospects to 3-outcomes prospects, substituted the Wilcox error model (Wilcox, 2008) for the Luce exponential rule, used the lognormal distributions on prior distributions of all parameters with values that could exceed one, and used gamma transformations for the prior distributions of all precision parameters. She also used MultiBUGS, the latest version of the WinBUGS software (Goudie et al., 2017) (Section 4.8.4.2). MultiBUGS is substantially more efficient than WinBUGS, as MultiBUGS invokes *parallel processing* whereby the number of iterations is divided by the number of processors of the computer. She hired time on a virtual or *cloud server* (Kamatera, 2020). The cloud server offered an option of 32 cores and rendered the Bayesian computations feasible. All but 1 of the 32 cores was set to run a single chain of GIBBs sampling. Initial values were supplied for the first of the chains, as doing so is advised (D.A. Burger, personal communication, June 9, 2019). The initial values for the other chains were generated randomly.

Several MultiBUGS runs were undertaken as each structural model was estimated for each numeracy group, using one or two functional forms for probability weighting. For some

models, the runs were repeated for different frames. For example, the rank-dependent model was computed for the set of gain-framed prospects from the GAINS2 questionnaire, and thereafter for those framed in both the gain and loss frames from the GLM2 questionnaire. For each run, the objective was to allow the GIBBS sampler as many iterations as necessary for the chains of the MCMC analyses to converge, enabling the results of each analysis to be credible. This meant that some runs were repeated with increasing numbers of iterations to achieve convergence. However, the time required to produce convergent solutions was challenging. Depending on the number of iterations required and the information quality of the priors, simpler runs that estimated only one parameter, as for expected utility model, took 1-3 hours to complete, whereas more complex runs, as for cumulative prospect theory, took up to 10 hours to complete. Furthermore, complex runs often terminated in a computational trap of 'undefined real values', which entailed abandoning the run or refining the initial starting values (i.e. not the priors) of the first chain sampler, based on the states of the chain values immediately before the trap. The final set of results from the MCMC analyses are summarised in Table 19 - Table 23. The Bayesian estimation approach used, and the results on convergence, model fit, parameter estimates and intercorrelations are discussed next.

5.3.4.1. MCMC Bayesian estimation

As described in Section 5.3.3, maximum likelihood parameter estimates were used for the prior distributions of the hierarchical Bayesian analyses. No respondents had scores common to both the maximum likelihood estimates and the Bayesian estimates of the same model parameters. In cases where the maximum likelihood estimates had not been derived, parameter estimates from the literature on those functions were used as priors. For example, for the weighting function of Lattimore et al. (1992), the researcher used past research findings summarised by Fox & Poldrack (2009, p. 158) on the function, although these findings were not based on groups differentiated on statistical numeracy levels. (See Section 6.2.2.3 for details on the alternate selection of prior values when maximum likelihood estimates were not available.)

The researcher used CODA (Plummer et al., 2006), a set of R functions, to inspect the MultiBUGS output. The results are presented as follows: the Bayesian output per numeracy group for each structural model and functional form are summarised under expected utility theory with power utility function (Table 19), rank-dependent utility with Tversky and Kahneman's (1992) probability weighting function (Table 20), rank-dependent utility with the Lattimore et al. (1992) probability weighting function (Table 21), and cumulative

prospect theory with Tversky and Kahneman's (1992) weighting function (Table 22). Each table summarises indicators of model fit in the form of the deviance information criterion (DIC) across the chains of the model (Spiegelhalter, Best, Carlin, & van der Linde, 2002), summary statistics of the parameter distributions across the chains, and a summary of the convergence of the chains with the posterior target distribution of each model parameter, as well as at the multivariate level for the overall model. Finally, Table 23 provides a high-level summary of the Bayesian analyses under all the models and functional forms considered.

5.3.4.2. Convergence

In Bayesian statistics, the samples of the posterior can only be used for inference purposes if the chains' simulations converge with the posterior target distributions for all model parameters. Before convergence may be assumed, the results of multiple convergence criteria need to be considered, and the weight of evidence assessed, as the individual criteria are necessary but not sufficient assurance of convergence (Plummer et al., 2006). The results presented in this section are viewed, conservatively, as indicative of *approximate* convergence, rather than of guaranteed convergence.

The statistical (rather than visual) convergence criterion considered in the thesis was the Gelman-Rubin index, also known as the 'potential scale reduction factor'. The Gelman-Rubin index compares the between and within variances of the simulation sequences of multiple parallel chains (Brooks & Gelman, 1998; Brooks & Gelman, 1992). For potential convergence, the parallel Markov chains overlap, and the index approaches a value of 1. Ideally, the values of the upper confidence limit of the Gelman-Rubin index should be less than 1.1 for all parameters (Heathcote et al., 2019), although some authors seem to prefer a more stringent criterion of 1.05. MultiBUGS provides the upper confidence limit of the index for the individual model parameters, as well as the multivariate scale reduction factor as an index of fit of the overall model. The multivariate value considers all the linear combinations of the individual values (Plummer et al., 2006).

In addition to the statistical criterion considered, the researcher visually inspected multi-frame sets of four plots supplied by CODA for each parameter by numeracy group (Figure L 1 - Figure L 10). Each set of four plots provides the following information: First, the bottom frame of each figure presents overlaid time series trace plots of simulated values versus the iteration number. For each parameter, these plots should have approximately constant means and variances across the iterations, indicative of well-mixing chains, and a necessary

condition for valid parameter estimates (Plummer et al., 2006, StataCorp, 2017). Second, the top left-hand frame of each figure presents the overlaid smoothed density plots of the parameter estimates of the posterior distribution. The shape of the plots should approximate the normal density function and be largely overlapping. Third, the top right-hand frame of each figure presents an autocorrelation plot showing the extent of the dependence between the samples drawn iteratively from the posterior distributions. Ideally, the sampler selects samples that are approximately independent, thus progressively accessing increased information from the posterior distribution of the parameters. Finally, the plot beneath the autocorrelation shows the tapering of the parameter estimates of the chains with increasing iterations.

MultiBUGS uses the Gibbs sampler, which is unfortunately inefficient, as it produces highly autocorrelated samples (Omori, 2007). Furthermore, the models considered in this thesis had connected parameters, which may have increased the autocorrelation problem of the sampler (D.A. Burger and S. Van Der Merwe, personal communication, June 7, 2019). Additionally, correlated parameters are known to slow the convergence of sampling chains significantly, and may result in impractical lengths of sampling (Van Ravenzwaaij et al., 2018). Accordingly, the researcher used chains with long series of iterations and a process of 'thinning' to select samples at fixed intervals to overcome the autocorrelation of successive samples.

The univariate and multivariate Gelman-Rubin potential scale reduction factors for each model and functional form per numeracy group are presented in Table 19 - Table 23. These tables provide statistical evidence of the approximate convergence of the chains of each analysis, among other summary information. In a few instances, the upper limits of the index were marginally greater than the 1.1 criterion, at 1.11 or 1.12, but practical constraints of time and budget prevented further modelling for these small margins. Nevertheless, the corresponding sets of diagnostic plots present graphic evidence of approximate convergence.

The autocorrelation plots of all the series dropped to approximately 0. The eventual reduction of autocorrelation, together with the well-mixed chains, thus satisfied additional conditions for convergent solutions.

The information in these tables and plots is considered indicative of the approximate convergence of the chains for all the models by numeracy groups. However, the results of estimates with upper limits of the Gelman-Rubin index of 1.11 or 1.12 are interpreted with

more caution. The next step was to examine the extent to which the choice data of each numeracy group supported each of the models.

5.3.4.3. Model fit

The models were compared across the numeracy groups on the deviance information criterion. This criterion is a trade-off between the fit of the model and its complexity, with smaller DIC values indicative of models that are better supported by the data. The actual magnitude of the DIC value is not considered a meaningful quantity. However, in comparing models, a DIC difference of more than 10 is used as a 'rough' criterion for eliminating models with higher DIC values, and differences between 5 and 10 are considered as models with 'substantially' better correspondence with the data (MRC Biostatistics Unit, n.d.). On this basis, each of the models is better supported by the risky choice data of the high numeracy group than by the data of the middle or low numeracy groups. Furthermore, the DIC differences between the middle and low numeracy groups are negligibly different for the expected utility model, and for the rank-dependent model with the Lattimore et al. (1992) function. The fit of the rank-dependent model with the Tversky and Kahneman (1992) function is substantially better for the middle numeracy group than for the lower numeracy group, and the cumulative prospect model with Tversky and Kahneman's (1992) probability function is also better for the middle numeracy group than for the low numeracy group.

5.3.4.4. Parameter estimates

Bayesian inference provides several summary statistics for the parameter estimates from the samples at the individual respondent level, and across individuals. Summary statistics across individuals are presented in Table 19 - Table 23. The statistics include the mean and median point estimates of the parameters of the posterior distributions, as well as the uncertainty in the parameter estimates in the form of the standard deviation and the 95% credible interval. The 95% credible interval provides the range within which the true parameter value of the posterior distribution lies with a 95% probability (Heathcote et al., 2019).

For all models, the pattern observed from the summary statistics is that of increasing risk aversion (alpha values decreasing) with lower levels of statistical numeracy, although the increases are small and risk aversion is nevertheless high for the high numeracy respondents.

Under the expected utility model, the 95% credible intervals of the high and low numeracy groups do not overlap on the utility parameter. Based on the direction of the group-level

utility means, risk aversion is lowest in the high numeracy group and highest in the low numeracy group.

Under the rank dependent model with the Tversky and Kahneman probability function, the same pattern of higher risk aversion with lower numeracy is observed, although the credible intervals do overlap. However, the direction of the means of the curvature parameters are not in the expected direction of increased probability distortion.

Under the rank dependent model with the Lattimore function with linear utility assumed, the values of the parameters are in the direction of increasingly distorted risk attitudes with lower numeracy. Once again, there is strong risk aversion for all groups (low alpha values), with the risk aversion of the low numeracy group greater than the high numeracy group. Furthermore, bets are more attractive for high numeracy respondents (highest delta parameter value), and decreasingly so for respondents of middle and low levels. Finally, probability sensitivity away from the extreme probabilities of 0 and 1, is highest for high numeracy respondents (highest gamma value), lower for respondents in the middle numeracy group, and lowest for respondents in the low numeracy group. Plots of the Lattimore weighting functions for the high and low numeracy groups illustrate the steeper gradient of the curve for the high numeracy respondents, i.e. greater probability sensitivity across almost the entire range of probability values, and the higher elevation of the curve, i.e. greater attraction to the gambles (Figure 28).

Under the cumulative prospect model, the alpha values are less than .1 for the three numeracy groups. Probability distortion is consistently lowest in the responses of the high numeracy group. There is a similar small amount of loss aversion observed for all groups. Plots of the Tversky and Kahneman (1992) weighting functions for the high and low numeracy groups show the traditional inverse *S* shape, with the curve of the high numeracy group more linear, and thus less distorted than the curve of the low numeracy group (Figure 29).

Table 19: Summarised results of MCMC analyses under expected utility theory with power function (31 chains)

Expected utility theory; Power function			Fit			Summary statistics							Convergence		
			DIC (chains)			Parameter summary statistics							Gelman-Rubin potential scale reduction factors		
frame	numeracy group	sample n	mean	median	range	parameter	mean	median	*SD*	val 2.5%	val 97.5%	sample		lower est.	upper c.i.
Gains	high	44	2902.9	2903	2902 - 2903	deviance	2899	2899	3.997	2890	2906	372000	deviance	1.04	1.07
						mu.alpha	0.377	0.381	0.05	0.269	0.466	372000	mu.alpha	1.04	1.07
													multivariate 1.05		
	middle	50	3297	3297	3293 - 3298	deviance	3285	3286	6.694	3271	3297	558000	deviance	1.04	1.06
						mu.alpha	0.242	0.246	0.051	0.131	0.332	558000	mu.alpha	1.04	1.07
													multivariate 1.04		
	low	50	3294	3294	3293 - 3294	deviance	3290	3290	3.684	3283	3297	542500	deviance	1.02	1.03
						mu.alpha	0.205	0.206	0.034	0.134	0.27	542500	mu.alpha	1.02	1.03
													multivariate 1.02		

Table 20: Summarised results of MCMC analyses under rank-dependent utility with Tversky and Kahneman (1992) function (31 chains)

Rank-dependent theory; Tversky and Kahneman (1992)			Fit			Summary statistics						Convergence		
			DIC (chains)			Parameter summary statistics						Gelman-Rubin potential scale reduction factors		
frame	numeracy group	sample n	mean	median	range	parameter	mean	median	*SD*	val 2.5%	val 97.5%	sample		lower est.	upper c.i.
Gains and losses	high	64	2824	2824	2824	deviance	2820	2821	3.923	2812	2828	341000	deviance	1.03	1.05
						mu.alpha	0.309	0.310	0.040	0.227	0.384		mu.alpha	1.05	1.08
						mu.delta	0.799	0.791	0.087	0.653	0.992		mu.delta	1.05	1.08
						mu.gamma	0.681	0.679	0.074	0.541	0.833		mu.gamma	1.05	1.08
													multivariate 1.08		
	middle	65	2851	2851	2851	deviance	2850	2850	1.789	2846	2853	341000	deviance	1.05	1.07
						mu.alpha	0.255	0.255	0.011	0.232	0.276		mu.alpha	1.03	1.05
						mu.delta	0.854	0.853	0.028	0.799	0.911		mu.delta	1.07	1.11
						mu.gamma	0.836	0.837	0.030	0.777	0.893		mu.gamma	1.04	1.06
													multivariate 1.09		
	low	65	2845	2845	2845	deviance	2844.0	2844.0	1.7	2841.0	2848.0	341000	deviance	1.08	1.12
						mu.alpha	0.224	0.224	0.010	0.203	0.244		mu.alpha	1.08	1.12
						mu.delta	0.919	0.919	0.032	0.854	0.981		mu.delta	1.06	1.09
						mu.gamma	0.924	0.924	0.032	0.861	0.986		mu.gamma	1.06	1.09
													multivariate 1.09		

Table 21: Summarised results of MCMC analyses under rank-dependent utility with Lattimore et al. (1992) function (31 chains)

Rank-dependence utility; Lattimore et al. (1992); linear utility assumed			Fit			Summary statistics							Convergence		
			DIC (chains)			Parameter summary statistics							Gelman-Rubin potential scale reduction factors		
frame	numeracy group	sample n	mean	median	range	parameter	mean	median	*SD*	val 2.5%	val 97.5%	sample		lower est.	upper c.i.
gains	high	44	2920	2920	2920	deviance	2918	2919	2.551	2913	2923	124000	deviance	1.05	1.08
						mu.delta	0.496	0.496	0.017	0.463	0.53		mu.delta	1.05	1.08
						mu.gamma	0.743	0.745	0.028	0.685	0.795		mu.gamma	1.05	1.08
													multivariate 1.07		
gains and losses	high	64	2793	2793	2793	deviance	2793	2793	0.999	2791	2795	248000	deviance	1.04	1.06
						mu.delta-	0.503	0.502	0.017	0.470	0.537		mu.delta-	1.04	1.07
						mu.delta+	0.499	0.499	0.017	0.466	0.533		mu.delta+	1.03	1.05
						mu.gamma-	0.780	0.779	0.026	0.730	0.835		mu.gamma-	1.05	1.08
						mu.gamma+	0.764	0.764	0.025	0.713	0.814		mu.gamma+	1.05	1.08
													multivariate 1.11		
gains and losses	middle	65	2843	2843	2843	deviance	2843	2843	1.607	2840	2846	310000	deviance	1.02	1.02
						mu.delta-	0.369	0.369	0.012	0.345	0.394		mu.delta-	1.02	1.03
						mu.delta+	0.362	0.363	0.012	0.337	0.387		mu.delta+	1.03	1.04
						mu.gamma-	0.703	0.702	0.024	0.658	0.751		mu.gamma-	1.02	1.04
						mu.gamma+	0.685	0.685	0.023	0.638	0.730		mu.gamma+	1.01	1.02
													multivariate 1.05		
gains and losses	low	65	2843	2843	2843	deviance	2847	2847	1.012	2845	2849	356500	deviance	1.02	1.03
						mu.delta-	0.302	0.301	0.010	0.282	0.322		mu.delta-	1.02	1.04
						mu.delta+	0.298	0.298	0.010	0.278	0.318		mu.delta+	1.02	1.04
						mu.gamma-	0.501	0.501	0.017	0.469	0.536		mu.gamma-	1.03	1.05
						mu.gamma+	0.494	0.493	0.016	0.462	0.526		mu.gamma+	1.03	1.04
													multivariate 1.05		

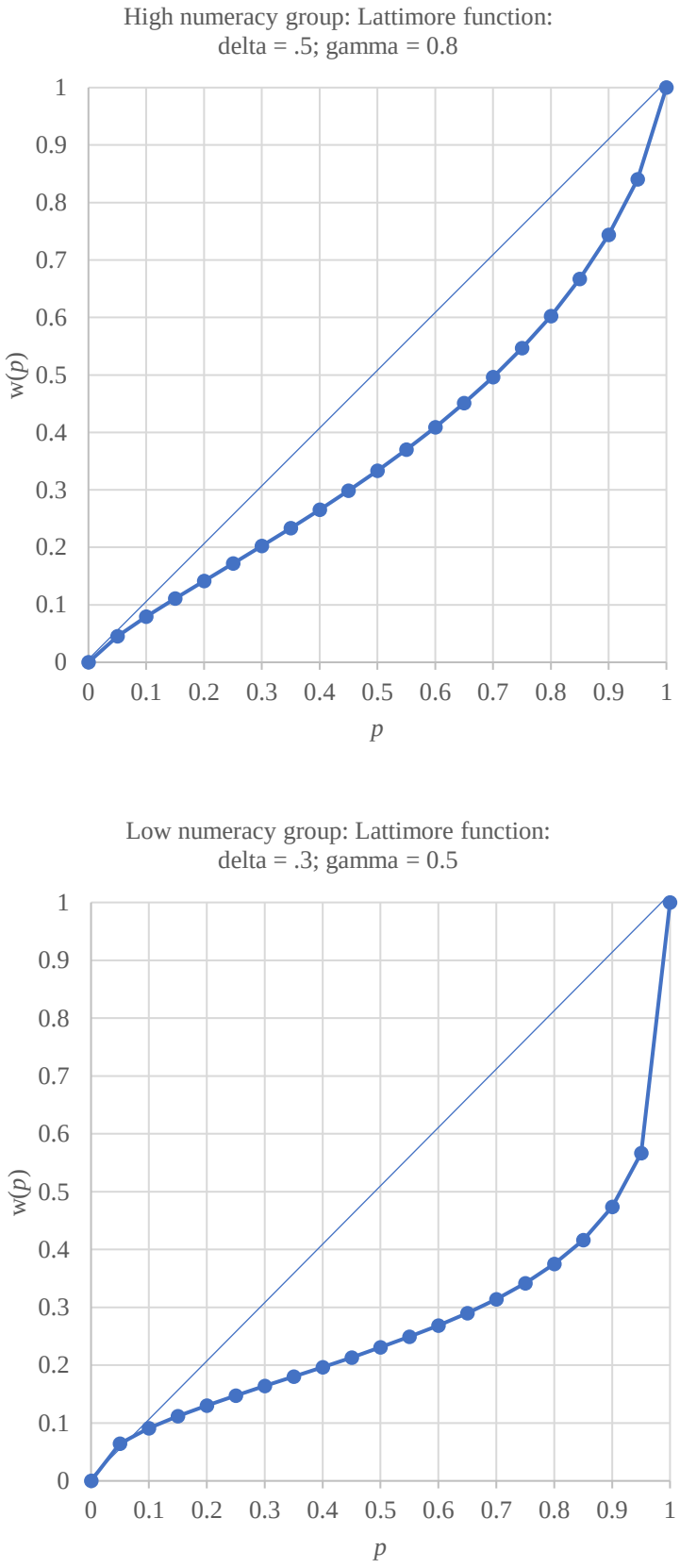


Figure 28: Lattimore probability weighting function with parameter estimates for high and low numeracy groups

Table 22: Summarised results of MCMC analyses under cumulative prospect theory; Tversky and Kahneman (1992) function (31 chains)

Cumulative prospect theory; Tversky and Kahneman (1992)			Fit			Summary statistics							Convergence		
			DIC (chains)			Parameter summary statistics							Gelman-Rubin potential scale reduction factors		
frame	numeracy group	sample n	Mean	Median	Range	parameter	mean	median	*SD*	val 2.5%	val 97.5%	sample		lower est.	upper c.i.
gains, losses and mixed	high	64	4208	4208	4204 - 4205	deviance	4205.0	4205.0	3.719	4197.0	4212.0	341000	deviance	1.03	1.05
						mu.alpha	0.089	0.088	0.017	0.058	0.124		mu.alpha	1.07	1.11
						mu.delta-	0.919	0.909	0.099	0.755	1.142		mu.delta-	1.06	1.09
						mu.gamma+	0.736	0.734	0.072	0.602	0.884		mu.gamma+	1.04	1.06
						mu.lambda	1.187	1.181	0.124	0.960	1.447		mu.lambda	1.07	1.10
													multivariate 1.09		
gains, losses and mixed	middle	65	4250	4250	4249 - 4251	deviance	4240.0	4241.0	6.238	4228.0	4253.0	341000	deviance	1.04	1.07
						mu.alpha	0.073	0.072	0.014	0.047	0.102		mu.alpha	1.05	1.09
						mu.delta-	0.674	0.669	0.063	0.566	0.813		mu.delta-	1.06	1.10
						mu.gamma+	0.466	0.466	0.058	0.353	0.58		mu.gamma+	1.05	1.08
						mu.lambda	1.158	1.148	0.124	0.940	1.426		mu.lambda	1.06	1.10
													multivariate 1.1		
gains, losses and mixed	low	65	4238	4238	4238 - 4239	deviance	4232.0	4232.0	5.075	4222.0	4242.0	341000	deviance	1.04	1.06
						mu.alpha	0.072	0.072	0.014	0.047	0.102		mu.alpha	1.02	1.04
						mu.delta-	0.516	0.515	0.051	0.419	0.620		mu.delta-	1.06	1.10
						mu.gamma+	0.536	0.535	0.056	0.430	0.649		mu.gamma+	1.05	1.07
						mu.lambda	1.18	1.169	0.1302	0.9538	1.469		mu.lambda	1.05	1.08
													multivariate 1.09		

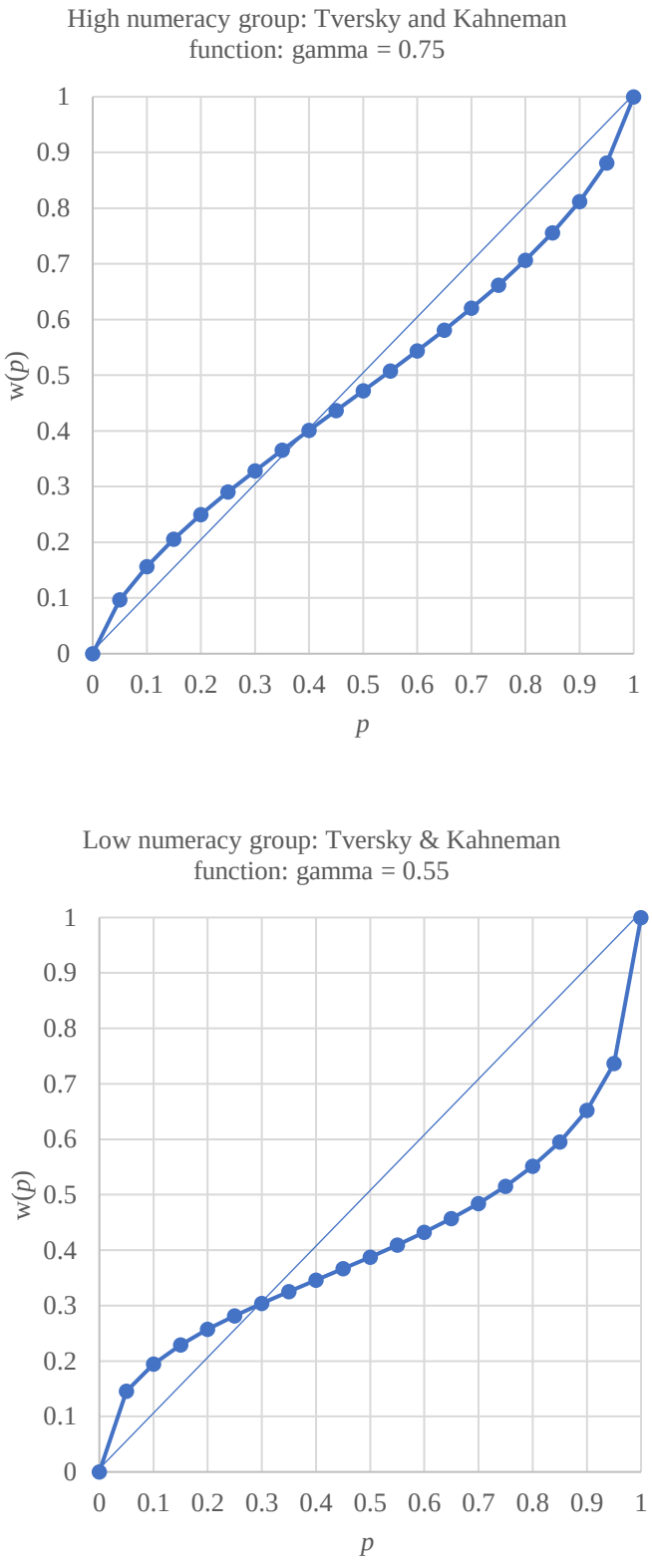


Figure 29: Cumulative prospect theory with Tversky and Kahneman's probability weighting function using parameter estimates for high and low numeracy groups, respectively

Table 23: Comparisons of the numeracy groups on model fit, point estimates of model parameters, and indices of chain convergence

Model description		Fit (median DIC)			Summary statistics for estimates (means)				Convergence (Gelman-Rubin upper confidence limit)			
Model (frame)	functionals	high	middle	low		high	middle	low		high	middle	low
EUT (gains)	power	2903	3297	3294	deviance	2899	3285	3290	deviance	1.07	1.06	1.03
					mu.alpha	0.377	0.242	0.205	mu.alpha	1.07	1.07	1.03
					iterations	372000	558000	542500	multivariate	1.05	1.04	1.02
					n	44	50	50	n	44	50	50
RDU (gains and losses)	T&K; power	2824	2851	2845	deviance	2820	2850	2844	deviance	1.05	1.07	1.12
					mu.alpha	0.309	0.255	0.224	mu.alpha	1.08	1.05	1.12
					mu.delta	0.799	0.854	0.919	mu.delta	1.08	1.11	1.09
					mu.gamma	0.681	0.836	0.924	mu.gamma	1.08	1.06	1.09
					iterations	341000	341000	341000	multivariate	1.08	1.09	1.09
					n	64	65	65	n	64	65	65
RDU (gains and losses)	Lattimore et al.; linear utility assumed	2793	2843	2843	deviance	2793	2843	2847	deviance	1.06	1.02	1.03
					mu.delta-	0.503	0.369	0.302	mu.delta-	1.07	1.03	1.04
					mu.delta+	0.499	0.362	0.298	mu.delta+	1.05	1.04	1.04
					mu.gamma-	0.780	0.703	0.501	mu.gamma-	1.08	1.04	1.05
					mu.gamma+	0.764	0.685	0.494	mu.gamma+	1.08	1.02	1.04
					iterations	248000	310000	356500	multivariate	1.11	1.05	1.05
					n	64	65	65	n	64	65	65
CPT (gains, losses, and mixed)	T&K; power	4208	4250	4238	deviance	4205	4240	4232	deviance	1.05	1.07	1.06
					mu.alpha	0.089	0.073	0.072	mu.alpha	1.11	1.09	1.04
					mu.delta-	0.919	0.674	0.516	mu.delta-	1.09	1.10	1.10
					mu.gamma+	0.736	0.466	0.536	mu.gamma+	1.06	1.08	1.07
					mu.lambda	1.188	1.158	1.180	mu.lambda	1.10	1.10	1.08
					iterations	356500	341000	341000	multivariate	1.09	1.10	1.09
					n	64	65	65	n	64	65	65

5.3.4.5. Intercorrelations of parameter estimates at individual level

This section provides the Pearson product moment correlation coefficients between all pairs of estimated parameters for each multi-parameter model at the individual level. The individual-level estimates of each model parameter are the means of the posterior distributions derived from the hierarchical Bayesian analyses per numeracy group (Table 24). The correlations assume bivariate normal distributions of the underlying variables and, furthermore, that the relations are linear.

It is noted that these individual-level correlations do not reflect the algorithmic structure of the model, but rather arise from associations between the psychological capacities that are measured by the parameters, for example between response noise and probability sensitivity (T. Pachur, personal communication, November 24, 2020). Strong correlations may lead to untrustworthy parameter estimates, and difficulties in assessing the psychological processes represented by the parameters.

The individual-level linear correlations between the parameters are in the order of .5 - .7, in absolute value (Table 24). These correlations imply that the parameters per model and numeracy group should be evaluated jointly, as the magnitude of the separate parameter estimates are not necessarily meaningful. Specifically, there are moderate-to-high correlations between the following pairs of parameters: first, between the utility and curvature parameters for all the multi-parameter models; second, between the curvature and elevation parameters of the rank-dependent model with the Lattimore probability function, within positively-framed prospects and within negatively-framed prospects; third, between the value and loss aversion parameters of the value and loss aversion parameters of the cumulative prospect model, these correlations being negative.

These parameter intercorrelations complicate the interpretation of the psychological processes represented by the individual parameters. Nevertheless, alpha values are lowest for the low numeracy group, implying highest risk aversion for these respondents. This is a consistent finding for the rank-dependent and cumulative prospect models with the Tversky and Kahneman probability weighting function. (The same ordering was observed for the of utility parameter estimates of the numeracy groups for the expected utility theory.)

The following section presents the validation of the models.

Table 24: Individual-level intercorrelations of MCMC parameter estimates for rank-dependent and cumulative prospect models

RDU Tversky and Kahneman (1992)	alpha mean	gamma mean	delta mean
High numeracy group			
alpha mean	1.00		
gamma mean	*0.60**	1.00	
delta mean	*0.60**	-0.05	1.00
Middle numeracy group			
alpha mean	1.00		
gamma mean	*0.42**	1.00	
delta mean	*0.46**	-0.12	1.00
Low numeracy group			
alpha mean	1.00		
gamma mean	*0.47**	1.00	
delta mean	*0.51**	0.19	1.00

RDU Lattimore	delta- mean	delta+ mean	gamma- mean	gamma+ mean
High numeracy group				
delta- mean	1.00			
delta+ mean	*0.47**	1.00		
gamma- mean	*0.26**	*0.43**	1.00	
gamma+ mean	0.18	*0.77**	*0.48**	1.00
Middle numeracy group				
delta- mean	1.00			
delta+ mean	0.17	1.00		
gamma- mean	*0.61**	0.11	1.00	
gamma+ mean	0.03	*0.83**	0.00	1.00
Low numeracy group				
delta- mean	1.00			
delta+ mean	*0.29**	1.00		
gamma- mean	*0.69**	0.13	1.00	
gamma+ mean	0.22	*0.67**	0.21	1.00

CPT Tversky and Kahneman (1992)	alpha mean	delta- mean	gamma+ mean	lambda mean
High numeracy group				
alpha mean	1.00			
delta- mean	*0.55**	1.00		
gamma+ mean	*0.57**	0.00	1.00	
lambda mean	*-0.53**	*-0.49**	*-0.43**	1.00
Middle numeracy group				
alpha mean	1.00			
delta- mean	*0.56**	1.00		
gamma+ mean	*0.62**	-0.01	1.00	
lambda mean	-0.19	-0.10	*-0.45**	1.00
Low numeracy group				
alpha mean	1.00			
delta- mean	*0.74**	1.00		
gamma+ mean	*0.66**	0.15	1.00	
lambda mean	-0.24	0.00	*-0.51**	1.00

Note: italicised values with * denote $p < .05$

5.3.5. Validation of the Bayesian estimates

The final sections of the chapter concern the accuracy of the estimates of the computational models and the validity of their predictions for the risky choices of decision makers in the different numeracy groups. Following Wilcox (2008), the consistency analysis is divided into *estimating* the success of the models based on *in-sample data*, i.e. the data used for arriving at the MCMC parameter estimates, and *predicting* the success of the models based on data not used for the estimates, i.e. *out-of-sample data*. As such, they address sub-question 2.3.

As in the preliminary analyses of the choice data (Section 5.3), the unit of analysis considered was the choices per Marschak-Machina triangle. The choices of each respondent per triangle were scored based on their consistency with the predictions of each of the structural models. For each numeracy group, the predictions of each model by functional form were computed using the parameters identified in the MCMC Bayesian analyses in the previous section (5.3.4.1).

As in Section 5.3, mixed repeated measures analyses of variance were computed on the consistency data of each respondent, by model and functional form, numeracy group, on in-sample and out-of-sample data. For the rank-dependent models, the positive and negative frames were analysed separately, as were the positive, negative and mixed frames of the cumulative prospect model. Furthermore, each analysis was based on responses to a common set of lotteries. For example, the rank-dependent in-sample analysis for the gain frame comprised responses of individuals in the three numeracy groups who had all responded to the same positively-framed choices of the GLM2 questionnaire, whereas the corresponding out-of-sample analysis comprised responses of the individuals in the numeracy groups who had all responded to the same choices of the GAINS2 questionnaire. The common set of choices ensured that there was internal consistency of choice characteristics in the comparisons of the numeracy groups.

The results of models that shared the same set of choices are presented in the same table – for instance, the Tversky and Kahneman function, and the Lattimore function for the rank dependence, are presented side-by-side. Finally, in all analyses *the emphasis is on the between numeracy group comparison* of the models' successful predictions of observed choices, as opposed to a comparison of the models.

In the following sections, descriptive statistics are presented on consistency per numeracy group, with the between numeracy group effect of the mixed repeated measures analyses of

variance. Post-hoc pairwise analyses of significant between numeracy group effects ($\alpha = .05$) were examined, using the Bonferroni correction (Field, 2013), to identify significant numeracy group mean differences. For analyses involving two contexts (R0, R100, R200 and R0, R100, R300), the payout*group interaction effect and payout effect are given. For all analyses, the effects of the gradient of the triangle, i.e. risk aversion, and the interaction of the triangle gradient with the numeracy main effect are considered only when significant and with moderately large effect sizes (η_p^2 values greater than .09).

As described in Section 5.3.2.1, the assumptions of the mixed repeated measures analyses were examined, and the more conservative Greenhouse-Geisser and Huynh-Feldt corrections used when the sphericity assumption was violated. In all cases when Levene's test for homogeneity of variances was significant, the violation of the homogeneity assumption derived from responses to choices in the triangle with the steepest gradient of the GLM2 questionnaire. In such cases, the analyses were repeated excluding these choices. The analyses are presented in Appendix L (Table M 2 - Table M 8). The results are summarised in Sections 5.3.5.2 - 5.3.5.4.

At the outset (Section 5.3.5.1), the percentages of observed choices consistent with the predictions of the *expected value model* are presented for each of the numeracy groups, for gains, losses and mixed prospects for respondents to the GAINS2 and GLM2 questionnaires (Table M 1). As in all the analyses, the emphasis is on the between numeracy group comparison of the model's successful predictions of observed choices, rather than on comparison between the models.

5.3.5.1. Expected value theory

In all cases, for both the GAINS2 and GLM2 responses, the mean consistency levels of expected value predictions with the observed responses were highest for the high numeracy group and lowest for the low numeracy group, whether or not the differences between the numeracy groups were significant..

No significant differences were found between the numeracy groups on the mean consistency levels of expected value predictions with observed responses to the GAINS2 questionnaire. This finding held for prospects with maximum payouts of both R200 and R300, respectively [$F(2,141) = 0.370$, $p = .692$; $F(2,141) = 0.823$, $p = .441$] (Table M 1). However, significant differences were found between the numeracy groups on the mean consistency levels of expected value predictions with observed responses to the GLM2 for prospects in the gain,

loss and mixed frames, respectively [$F(2,191) = 5.575$, $p = .004$; $F(2,191) = 16.826$, $p < .001$; $F(2,191) = 6.691$, $p = .002$], with the effect size for losses moderate to strong, and the others weaker ($\eta_p^2 = .06, .15, .07$, respectively). On average, consistency levels for the GAINS2 questionnaire ranged from 43% - 45% in the R200 context, and 39% - 41% in the R300 context; and for the GLM2 questionnaire they ranged from 46% - 53%, 56% - 68% and 50% - 60%, respectively, for responses to gains, losses and mixed prospects (Table M 1).

5.3.5.2. Expected utility theory

As the parameters of the power function under the expected utility model were estimated for gain-framed lotteries only, the assessment of the model's in-sample consistency applies to responses to the GAINS2 questionnaire, and its out-of-sample consistency applies to responses to the gain-framed lotteries in the two contexts of the GLM2 questionnaire (maximum payouts of R200 and R300) (Table M 2).

In-sample estimates

For respondents in the high, middle and low numeracy groups, the mean consistency levels between actual choices and the in-sample estimates of the expected utility model are 68%, 67% and 63%, respectively, in the R0, R100, R200 context, and significantly lower at 58%, 59% and 61%, respectively, in the R0, R100, R300 context ($F(1,141) = 27.561$, $p < .001$). The switch in the order of the numeracy group means between the two contexts accounts for the significant payout*numeracy interaction effect ($F(2,141) = 3.683$, $p = .028$) shown in Table M 2.

Out-of-sample predictions

When responses to the choices of the triangle with steepest gradient are discounted from the analysis, the out-of-sample consistency means are similar to the in-sample means (65%, 69% and 67%). As previously mentioned, the homogeneity of variance assumption of the analyses was violated when the consistency scores of choices derived from the triangle with the steepest gradient were included, thereby justifying their exclusion from the analysis. The exclusion did not change the significance of the analysis of variance results (Table M 2).

There is insufficient evidence of a significant difference between the consistency means of the three numeracy groups, whether for the in-sample means based on the R200 or R300 maximum payout contexts or for the out-of-sample means. The similarity in the results of the in-sample and out-of-sample analyses is interpreted as the generalisability of the results to other prospects in the gain frame.

5.3.5.3. Rank-dependent theory

The consistency analyses for the rank dependent model using the probability weighting functions of Tversky and Kahneman (1992) and Lattimore et al. (1992) are presented side by side as they pertain to the same prospects. The analysis of gain-framed choices is presented for in-sample estimates in Table M 3 and for out-of-sample predictions in Table M 4. The in-sample analysis is presented for prospects in the loss frame in Table M 5. There were insufficient out-of-sample responses for a consistency analysis in the loss frame.

Gain frame: In-sample estimates

As previously, the analyses of the gain-framed lotteries of the GLM2 questionnaire were computed by first including and then excluding the consistency scores of choices derived from the triangle with indifference curve with steepest slope, and then presenting the latter case in Table M 3. With these results excluded, the mean consistency levels between observed choices and the in-sample estimates of the rank-dependent model using the Tversky and Kahneman (1992) function are 68%, 70% and 67%, respectively. Using the Lattimore et al. (1992) function with linear utility assumed, the consistency means are lower at 60%, 56% and 53%, respectively. Under the Lattimore et al. function, there is a significant numeracy group effect ($F(2,191) = 5.32$, $p = .006$) with a small η_p^2 effect size of .053. The post-hoc Bonferroni test shows a significant difference between the observed versus estimated choice consistency of the high and low numeracy groups ($p = .004$).

Gain frame: Out-of-sample predictions

No significance was found between the numeracy groups on the consistency of the observed versus predicted choices of the rank-dependent model under either the Tversky and Kahneman (1992) or Lattimore et al. (1992) functions, in either the R200 or R300 maximum payout contexts (Table M 4). For the Tversky and Kahneman (1992) function, the mean consistency levels range from 64% - 69% for the context with maximum payout of R200, and from 60% - 62% for the context with maximum payout of R300. For the Lattimore et al. function, the corresponding means are lower: 53% - 55% and 45% - 52% in the R200 and R300 contexts, respectively.

Loss frame: In-sample estimates

The mean consistency levels between observed and predicted choices are lower than 50% under the Tversky and Kahneman function, with significant between numeracy group effect ($F(2,191) = 4.271$, $p = .015$) (Table M 5). The mean level is lowest for the low numeracy

group, and highest for the high numeracy group, with all pairwise differences significant ($p < .001$ for all pairwise mean differences) based on Bonferroni post-hoc tests. The consistency means are higher for the high and middle numeracy groups (67% and 55%, respectively) using the Lattimore et al. function, with all post hoc comparisons of pairwise group means once again significant ($p < .001$).

5.3.5.4. Cumulative prospect theory

The cumulative prospect model in the gain frame is presented for in-sample estimates and out-of-sample predictions in Table M 6. The in-sample analysis for the loss frame and the mixed frame are presented in Table M 7 and Table M 8, respectively. There were insufficient out-of-sample responses for a consistency analysis in the loss and mixed frames.

Gain frame: In-sample estimates

As for the rank-dependent model, no significant difference was found between the numeracy groups on the consistency of the cumulative prospect theory estimates, with mean consistency scores of 63% - 70% (Table M 6).

Gain frame: Out-of-sample predictions

Once again, no significant numeracy group effect was found on the consistency of the predictions of the model and the observed choices in the gain frame. The consistency means are significantly lower in the R300 context than in the R200 context ($F(1,237) = 27.442$, $p < .001$) with a moderate η_p^2 effect size of .1. The mean consistency levels are at 64% - 68% in the R200 context, and 57% - 62% in the R300 context (Table M 6).

Loss and mixed frames: In-sample estimates

The mean consistency levels for the estimates of loss-framed choices with observed choices, and mixed-framed choices with observed choices, are less than 50% (Table M 7 and Table M 8, respectively).

5.3.5.5. Linear mixed models

Finally, the mixed repeated measures analyses on consistency were checked using linear mixed models, which allow greater flexibility in their assumptions of independence between observations, among other advantages (West, Welch, & Galecki, 2007). In the context of the thesis, they provide broader and more efficient methods of estimating the repeated measures effect of the gradient of the indifference curves of the Marschak-Machina triangles. They also allow the effects of the numeracy clustering variable and the repeated measures gradient

effect to be modelled as random factors with differing intercepts and slopes, as opposed to fixed factors without this variability. As in the hierarchical Bayesian analysis, the models allow a hierarchical structure whereby the responses of individuals within numeracy groups can be related to one another. However, the greater flexibility of these models is associated with greater complexity. These models rely on (variants of) maximum likelihood estimation and do not always produce solutions. Other assumptions include the nature of the effects and the correlations among parameters. Thus, for the purpose of the thesis, linear mixed models were used as a check on the extent of the successful predictions of only the most successful decision models, per numeracy group.

Accordingly, linear mixed model analyses were conducted on the predictions in the gain frame of the expected utility model, and in the gain frame of the rank-dependent model with the Tversky and Kahneman function, on the out-of-sample data. Respondents were considered nested within numeracy groups, completely crossed with the repeated measures effect of triangle gradient.

For the predictions of both the expected utility and rank-dependent models, no significant differences were obtained among the numeracy groups on the percentages of out-of-sample estimates that were consistent with observed choices. This finding held whether or not the effects of the group and triangle gradient were considered as fixed or random. Furthermore, the linear mixed model solutions based on random effects were uncertain as they did not achieve convergence. The software used for the modelling, IBM SPSS Statistics for Windows Version 25, nevertheless provided results for the nonconvergent estimates based on the last iteration processed. For fixed effects, the results of the linear mixed models and repeated measures analyses of variance were consistent.

5.3.5.6. Conclusions on the validity of the models considered

Several conclusions are drawn from the detailed analyses of the previous section on the validation of the models, aided by the summaries of Table 23 and Table 25.

First, excluding the expected value model, the estimation and prediction successes of the expected utility, rank-dependent and cumulative prospect models in the gain frame exceeded the chance prediction level of 50%. Relative to observed risky choices in the gain frame, these models' outcomes were consistent in 60% - 70%, on average.

Second, the consistency levels of these models' estimates and predictions in the gain frame with observed choices, were not found to differ significantly between the numeracy groups.

However, the estimates and predictions of these models were based on sets of parameters specific to each numeracy group.

Third, the consistency levels of the models are generally lower when the stake of the highest possible prize is raised.

The fourth conclusion is that the models considered in the loss and mixed frames failed to estimate or predict the observed risky choices of the participants satisfactorily. The estimates and predictions of the models in the loss and mixed frames were frequently worse than chance. This result occurred despite the choices of the GLM2 questionnaire having been constructed to be sensitive to distortions of value and probability, commonly found in the risky choice literature. Furthermore, in these situations of poor model performance, the performance was always worst in the case of the low numeracy group. For example, under rank dependence using the weighting function of Lattimore et al. and linear utility, estimates are consistent for 67% of the actual choices of respondents in the high numeracy group, and 55% and 46% for respondents in the middle and low numeracy groups, respectively.

Fifth, choices of the highly numerate respondents were the most consistent with expected value theory, and those of the low numeracy respondents were always the least consistent. Finally, no added value was found to be derived from substituting the linear mixed model analysis for the mixed repeated measures analysis of variance.

It is stressed that the main objective of this section is to compare the successes of each of the models' estimates and predictions of the risky choices of respondents in the three numeracy groups. This objective is in line with the main research question of the thesis, i.e. to examine the role of statistical numeracy in the modelling of risky choice. The comparisons are *not between* the models, but rather *within* the models, as each model involves a different set of choices and a different sample of respondents. The choices of the questionnaires were not constructed to be equally powerful tests of the models, and so comparisons between them may be biased. By contrast, comparison of the numeracy groups on the same model with a repeated measures analysis uses the same set of prospects for the comparison groups, and the repeated measures analysis controls for differences between the respondents.

These findings address Research sub-question 2.3 on the validity of the models' estimates and predictions for the risky choices of decision makers in different numeracy groups. They are discussed in more detail in Chapter 6.

Table 25: Parameter estimates, consistency of in-sample estimates, and out-of-sample predictions with observed choices, per numeracy group

Model estimates (Research question 2.2)				Consistency of model estimates and predictions with observed choices (Research question 2.3)													
Model				Frame	In-sample/ out-of-sample	Data	Context	Mean consistency of models with actual choices			Significance of consistency		Z scores relative to chance (50%)			Cohen's (1992) effect size relative to chance (5%,15%,25%)		
parameter	high	middle	low					high	middle	low	numeracy	context	high	middle	low	high	middle	low
Expected utility: Power (gains)				gains	in	GAINS2	R200	67%	67%	63%	ns	***	6.34	5.97	4.5	med	med	small
mu.alpha	0.38	0.24	0.21				R300	58%	59%	61%	ns		2.66	3.14	3.8	small	small	small
					out	GLM2	R200	65%	69%	67%	ns	n/a	5.44	7.06	6.38	med	med	med
Rank-dependence: Tversky & Kahneman (1992) (gains and losses)				gains	in	GLM2	R200	68%	70%	67%	ns	n/a	6.52	7.73	6.38	med	med	med
mu.alpha	0.31	0.26	0.22		out	GAINS2	R200	69%	65%	64%	ns	*	7.11	5.27	4.74	med	med	small
mu.delta-	0.80	0.85	0.92				R300	60%	61%	62%	ns		3.37	3.69	4.26	small	small	small
mu.gamma+	0.68	0.84	0.92	losses	in	GLM2	-	48%	42%	37%	*	n/a	-0.58	-2.68	-4.42	-	-	-
					out	small n												
Rank-dependence: Lattimore et al. (1992) (gains and losses)				gains	in	GLM2	R200	60%	56%	53%	**	n/a	3.30	2.06	0.84	small	small	neg
mu.delta-	0.50	0.37	0.30		out	GAINS2	R200	55%	53%	53%	ns	ns	1.52	1.1	1.1	small	neg	neg
mu.delta+	0.50	0.36	0.30				R300	45%	48%	52%	ns		-1.52	-0.78	0.78	-	-	neg
mu.gamma-	0.78	0.70	0.50	losses	in	GLM2	-	67%	55%	46%	***	n/a	6.24	1.6	-1.27	med	small	-
mu.gamma+	0.76	0.69	0.49		out	small n												
Cumulative prospect theory: Tversky & Kahneman (1992) (gains, losses and mixed)				gains	in	GLM2	R200	63%	70%	67%	ns	n/a	4.27	7.73	6.38	small	med	med
mu.alpha	0.09	0.07	0.07		out	GAINS2	R200	68%	65%	64%	ns	***	6.71	5.27	4.74	med	med	small
mu.delta-	0.92	0.67	0.52				R300	57%	62%	61%	ns		2.25	3.92	3.58	small	small	small
mu.gamma+	0.74	0.47	0.54	losses	in	GLM2	-	44%	42%	37%	ns	n/a	-2.08	-2.68	-4.42	-	-	-
mu.lambda	1.19	1.16	1.18		out	small n												
				mixed	in	GLM2	-	47%	44%	42%	ns	n/a	-1.13	-1.83	-2.72	-	-	-
					out	small n												

Notes: *ns* indicates non-significance ($p > .05$); n/a is not applicable; * $p < .05$; ** $p < .01$; *** $p < .001$; R200 context: R0, R100, R200; R300 context: R0, R100, R300

Chapter 6: Discussion

This chapter discusses the empirical approach taken by the thesis in modelling the relation between statistical numeracy and risky choice, and the findings of the research, in the context of the theoretical literature. The first section (6.1), addresses Research question 1. It discusses findings on the underlying factor structure of statistical numeracy, the categorisation of the respondents into groups that differ on statistical numeracy, and the key attributes of the groups. The second section (6.2), addresses Research question 2. It first discusses the results of the preliminary model-free analyses of the risky choice responses (Sub-question 2.1). It then discusses the approach of the thesis to some of the key aspects of modelling. This discussion provides the context for the results on the modelling (Sub-questions 2.2 and 2.3). Finally, the third section (6.3) addresses policy implications of the results of the research.

6.1. Statistical numeracy

The research question on statistical numeracy has three parts: the factor structure of numeracy (6.1.1), the measurement invariance of the factors (6.1.2), and the description of the numeracy groups (6.1.3). These sections address Research sub-questions 1.1 - 1.3, respectively.

6.1.1. Factor structure of the numeracy responses of the thesis (Research sub-question 1.1)

This section discusses the analyses that addressed the first sub-question on numeracy – Research sub-question 1.1: What is the factor structure of numeracy?

The researcher considered various options for deriving the statistical numeracy factor structure. As described in Section 4.7, the exploratory structural equation modelling technique of the Mplus software was used in preference to a confirmatory factor analysis approach. The confirmatory approach would have been more appropriate had the factors been expected to reflect each of the constructs measured by the underlying measurement scales, and had the analysis been proposed to confirm this. However, the researcher preferred an exploratory approach which provided a factor solution that would account for the correlations between the items adequately and efficiently, without the restrictions of the items' allegiances to their original scales. Furthermore, the exploratory analysis approach allows for the rotation of the factor structure to derive a more interpretable solution, which the confirmatory approach does not. The Geomin oblique rotation was appropriate for the numeracy factors

that were expected to be correlated, and based on dichotomous item scores (Muthén & Muthén, 2010).

As detailed in Section 5.2.3, 2 of the 11 numeracy items were dropped from the final exploratory factor analysis owing to their lack of variance and weak factor loadings. The remaining nine items were drawn from the Schwartz and Lipkus scales, the Berlin Numeracy Test and the Cognitive Reflection Test. An inspection of the item characteristic curves showed that the nine items spanned a wide range of numeracy ability levels (Figure 18), and that there was no obvious need to include the other two items, thus permitting a more efficient set of items.

The exploratory modelling technique revealed three items with dominant loadings on the first factor. Two of these items required calculating how many times the rolls of a fair die with a specified number of sides would be expected to have a particular outcome (items SL1 and BNT2a); the third item required transforming a ratio of 1 in 1,000 to a percentage (item SL3). Correct responses to these items reflect the ability to perform the operations of proportions and percentages. By contrast, the items with dominant loadings on the second factor were the two items of the Berlin Numeracy Test that require Bayesian-type reasoning (items BNT1 and BNT3), and the three items of the Cognitive Reflection Test that require the respondent to engage in some form of rational checking of her initial responses (Table 5 and Table 8).

The content of both sets of items fit the definition of statistical numeracy adopted in the thesis, i.e. the practical understanding of probabilities, percentages and statistics for use in quantitative problem solving (Garcia-Retamero et al., 2019). This definition also partially accounts for the high correlation found between the two factors. However, for the researcher to differentiate between the factors, a more detailed description of each factor is required. The descriptions are based on the content of the items with highly significant factor loadings, and on the results of tests of population homogeneity (Table 10 and Table 12).

The first factor reflects numerical ability that is more characteristic of younger, more recently matriculated respondents, rather than older, more qualified respondents. The items with the highest loadings on this factor represent a simple level of *numerical operations* involving ratio concepts, fractions, proportions, percentages and probabilities (Reyna et al., 2009). This basic level of numerical ability is likely to reflect mathematics at secondary school level. Compared to older students, younger recently matriculated respondents are likely to be more familiar with these types of basic numerical problems. Students who have recently left school

are also likely to have encountered similar problems in general intelligence (IQ) tests at school. Furthermore, students with a higher family income are likely to have received superior schooling and be more familiar and comfortable solving these types of problems, compared with those of a lower family income who are likely to have received poorer schooling. The familiarity of the younger students with this type of school-level task, particularly those with higher family incomes, could explain the income-based difference on this factor. This factor is named *probabilistic operations.*

By contrast, the second factor reflects more sophiticated numerical ability. The items that load highest on this factor are more complex and difficult. Their numerical tasks involve multiple, interdependent steps and some items require cognitive reflection (Peng, Wang, Wang, & Lin, 2019; Frederick, 2005). They thus require a higher level of *reasoning* involving mathematical operations (Cokely et al., 2012). They also have a greater verbal component, and thus rely on more sophisticated comprehension skills. Students of higher socio-economic levels who have superior schooling, and students with higher qualification levels, are more likely to have developed the higher numeracy and other cognitive reasoning skills required to answer these complex items correctly. This factor is named *probabilistic reasoning*.

The distinction between the two factors may be related to the sharp distinction that Stanovich and colleagues draw between the constructs that intelligence tests measure, versus the construct of rationality (Stanovich , 2016) (see Section 3.2.2). Interestingly, the Lipkus, Cognitive Reflection and Berlin Numeracy tests are all part of the Comprehensive Assessment of Rational Thinking measurement instrument, compiled by Stanovich and colleagues to measure rational thinking (Stanovich et al., 2016) (see Section 3.2.2). The links between the numeracy factors of the thesis and the rational thinking scale of Stanovich et al. is cited as an area for future study (see Section 7.4).

As presented in Section 5.2.3.6, the derived factors are highly correlated at $r = .66$ ($p < .001$) (Figure 19), based on factor scores calculated as simple linear combinations of the items weighted by their respective factor loadings. The majority of respondents' factor scores tend to covary, with scores on the probabilistic reasoning factor increasing linearly with scores on the probabilistic operations factor. The few low scores on the probabilistic operations factor but high on the probabilistic reasoning factor may reflect unreliability or measurement error in the data. Another possibility is that individuals with such scores are compromised on basic numeracy ability but use smart heuristics to solve problems, as in the tradition of Gigerenzer

and colleagues (Gigerenzer, 2001) (see Section 3.1.3). These possibilities are mere speculation however, and beyond the scope of the thesis which did not address numeracy processes empirically.

Two further observations are made on the factor structure presented in Section 5.2.3: First, each factor represents items or indicators from more than one of the existing measurement scales of numeracy considered in the thesis. Second, the two factors represent a simpler, and a more difficult level of statistical numeracy ability, respectively. Thus, more than one test is important for deriving the numeracy factors. Each test contributes mostly to one or the other factor – whereas the items of the Schwartz/Lipkus scale have high standardised loadings on the first factor, those of the Berlin Numeracy and Cognitive Reflection tests have high loadings on the second factor. However, to the knowledge of the researcher, no study has investigated the latent factor structure of numeracy, as reflected by items of these frequently-used measurement scales of numeracy. In general, studies that involve multiple scales do not examine the constituent items, but rather consider correlations calculated at the test total level. Conversely, studies that analyse responses at the item level generally involve single measurement scales, or items of a few scales only.

The most thorough research endeavour that involved multiple numeracy scales was conducted by Cokely et al. (2012), in a series of studies on the convergent and discriminant validity of the Berlin Numeracy Test. For the purpose of convergent validity, they included the Schwartz and Lipkus numeracy scales, the Cognitive Reflection Test and tests of cognitive abilities; for the purpose of discriminant validity they incorporated measures of different traits, including motivation and personality. The analysis was conducted at the level of total test scores and did not analyse item-level responses. The results revealed correlations of moderate strength among the numeracy test total scores and the total score of the Cognitive Reflection Test, and low correlations between the Berlin Numeracy Test with measures of other traits. Hierarchical regression analyses of risk comprehension on the total scores of the Berlin Numeracy Test, the Raven's Test of fluid intelligence, and the Cognitive Reflection Test showed that both the Berlin Numeracy Test and the Cognitive Reflection Test contributed additional unique variance after controlling for the variance explained by the other two tests. The Raven's Test of fluid intelligence did not contribute more variance than explained by the other two. These findings imply that the Berlin Numeracy and Cognitive Reflection tests explain all the variance in fluid intelligence, but each has a unique role in explaining individual differences in numeracy.

Studies of responses to the items of individual numeracy scales are limited. Lipkus et al. (2001) factor analysed the 11 items of the Lipkus scale only, and found a single factor. Cokely et al. (2012) conducted principal component analysis on the four items of the Berlin Numeracy Test. They found that the items loaded on one factor that explained 45% of the total variance. Likewise, factor analyses of the items of multiple numeracy scales are limited. Campitelli and Gerrans (2014) used a mathematical modelling approach to analyse the responses of over 2,000 individuals to the items of the Cognitive Reflection Test. They sought to determine whether the test measures mathematical ability only, both mathematical ability and rational thinking, or a combination of mathematical ability, rational thinking and open-minded thinking. Their analysis included three items of the Lipkus scale. Weller et al. (2013) incorporated the items of the Schwartz scale, selected items of the Lipkus scale and the cognitive reflection items into the Abbreviated Numeracy Scale, but did not include the Berlin Numeracy items.

The practice of considering the factor structure of the individual numeracy scales in isolation extends to *using* the scales separately in studies of risky choice. The problem here is that the individual measurement scales are generally either too easy (for example, the Schwartz and Lipkus scales), or too difficult (for example, the Berlin Numeracy and Cognitive Reflection tests), to show variance in the numeracy abilities of respondents (Section 3.7.3). Researchers who use only an easy or a difficult scale are faced with low scale reliability and low discriminability in scores at the lower or upper ends of the range of numerical ability. Cokely et al. (2012) provide several examples of studies with numeracy score distributions of the Lipkus et al. scale with extreme negative skew. Tests that produce scores that approach the measurement ceiling may produce biased results. An example is the study of Patalano et al. (2015) whose categorisation of respondents to high versus low numeracy groups on the very easy Lipkus scale was based on a median score of 10.5/11 (see Section 3.6.1).

Using a very difficult scale is similarly problematic, as scores approach the measurement floor effect with little variability at the low end. For example, 73% of the respondents in the study of Millroth et al. (2019) answered at most one of the four Berlin Numeracy Test items correctly. These researchers constructed five numeracy groups based on the distribution of total correct responses, with 2% of their respondents in their top numeracy group. The accuracy of such groupings depends on the reliability of the underlying measurement scale, as random measurement error disturbs the true ranking of the individuals if the test has low reliability (for example, internal consistency as measured by K-R 21). Although Millroth and

colleagues do not provide a measure of the reliability of the scale, the alpha coefficient based on the responses to the Berlin Numeracy items in the thesis was low at .51, with unacceptably low average inter-item correlation of .21 (Hair et al., 2010). Thus, Millroth's method of categorising respondents into five numeracy groups based on the 4-item Berlin Numeracy Test would not have been used successfully in the thesis. In this situation, other researchers have attempted to use multiple numeracy scales, separately. However, this strategy presents the challenge of categorising respondents into numeracy groups based on two or more measures, both of which may display low variance separately. An example is the study of Xing et al. (2019), who used the total scores on the Lipkus and Weller scales separately.

In summary, the thesis has shown the need for a test of statistical numeracy with a clear factor structure that discriminates across a wider range of ability levels than any one of the current numeracy tests does individually. In response, the thesis has combined selected items of four of the commonly used numeracy scales in the risky choice literature into a single measure, and provided a comprehensive item-level analysis of their factor structure. The set of nine items selected from these tests discriminated across a wide range of the statistical numeracy abilities of a university student population. Thus, the thesis has answered the first of the sub-questions on statistical numeracy, by demonstrating a clear factor structure of statistical numeracy, namely a probabilistic operations factor, and a probabilistic reasoning factor (Sub-question 1.1).

The next sub-question on statistical numeracy is focused on the invariance of the derived factors.

6.1.2. Measurement invariance (Research sub-question 1.2)

This section discusses the analyses that addressed the second sub-question on numeracy – Research sub-question 1.2: Do the factors display measurement invariance?

The researcher established the measurement invariance of the numeracy factors for the four covariates considered. For the thesis, measurement invariance implied that the numeracy factors of probabilistic operations and reasoning had the same meaning across groups categorised on sex, qualification, income and age levels. The invariance property was highly desirable for the study, as a lack of invariance would imply the complications of conceptually different latent variables for the construct of numeracy, dependent on a grouping variable. An additional reason for assessing the invariance property is that the Cognitive Reflection Test, integral to the numeracy factor structure, has been found to lack invariance (Campitelli &

Gerrans, 2014). These researchers claim that the test taps three types of abilities, namely, mathematical ability, rational thinking, and the disposition of open-minded thinking. In their analysis, performance on the test was associated with open-mindedness for males, but not for females. Although open-mindedness was not considered in the thesis, it was important to examine the invariance of the Cognitive Reflection Test items on the covariates considered in the context of risky choice.

By displaying the measurement invariance of the factor structure of statistical numeracy, the thesis has answered the second of the sub-questions on statistical numeracy (Sub-question 1.2). The meaning of the factors across the demographic groups was not found to differ, although the demographic groups differed on the means of the factors, i.e. on the population homogeneity of the numeracy factor structure, as discussed in Section 6.1.1.

As described in Section 4.7.4, the factor structure of the numeracy items contributed directly to the categorisation of respondents into statistical numeracy groups.

6.1.3. The statistical numeracy groups (Research sub-question 1.3)

This section discusses the analyses that addressed the third sub-question on numeracy – Research sub-question 1.3: What are the latent statistical numeracy subgroups?

The two numeracy factors provided useful input to the factor mixture model that produced the latent statistical numeracy subgroups, and contributed to the interpretation of what these groups meant for the study. As described in Section 5.2.4, the factor mixture model uses an algorithm that simultaneously conducts a categorical-type latent class analysis with a continuous factor-type analysis. In the thesis, the combined analyses allowed respondents to be assigned to latent numeracy classes while considering the dimensionality and variability of the continuous factor-type latent variables (Clark et al., 2013). Guided by the 2-factor structure found for statistical numeracy, the researcher selected a 2-factor mixture model. She further selected a 3-group variation of the factor mixture model based on the interpretability of the item profiles and the size of the groups that resulted when the algorithm of the mixture model was applied (see Section 6.1.3).

The three clearly defined numeracy groups set the condition necessary for assessing the risky decision models on each one. The groups comprised a high numeracy latent class (19%), with most respondents scoring both simple and more complex items correctly (probabilistic operations and probabilistic reasoning, respectively); a middle numeracy class (52%) with most respondents scoring well on relatively simple numeracy items (probabilistic operations),

but poorly on the more difficult items (probabilistic reasoning); and the lowest numeracy class (29%) with most respondents scoring poorly on both the more complex probabilistic operations and reasoning items. Consistent with the naming of the numeracy factors, the high numeracy group is named the *numerically rational group*, the middle numeracy group is named the *numerically operational group*, and the low numeracy group is named the *non-numerical group.*

The numeracy groups differed on demographic variables. The high and middle numeracy groups were slightly younger, with more of their respondents aged 18-21 (58% and 55%, respectively), compared to 36% in the low numeracy group. Similarly, they had a higher percentage of matriculants (65% and 58%, respectively), compared to 43% in the low numeracy group. There was also a higher representation of males in the high and middle numeracy groups (67% and 40%, respectively), compared to 27% in the low numeracy group.

These numeracy groups could not have been derived successfully by using any of the numeracy scales considered separately. A retrospective analysis, in which the total scores of the four items of the Schwartz/Lipkus scales were categorised and cross-tabulated with the three numeracy groups of the factor mixture model, showed that, at best, 43% of respondents could have been categorised correctly into the numeracy groups of the study, with multiple possible allocations of the other respondents. On the Berlin Numeracy Test, depending on the categorisation used on the 0 - 4 range of total scores, 61% - 71% of the respondents could have been categorised correctly into the three groups with the remaining categorisations ambiguous. Similarly, 60% - 62% of the respondents could have been categorised correctly using the Cognitive Reflection Test alone. These analyses show the benefit of using a composite of the items of the numeracy tests and the Cognitive Reflection Test for categorising respondents into numeracy groups.

By producing, describing and naming three clear latent groups on numeracy, the thesis has answered the last of the sub-questions on statistical numeracy (Sub-question 1.3). In so doing, the risky choices of the three numeracy groups could be modelled.

6.2. The relation between risky choice and statistical numeracy

This section of the chapter addresses Research question 2: What risk attitudes and probability distortions are associated with the risky choices of decision makers of different numeracy groups? It does so by answering each of the supporting sub-questions on the relation between numeracy and risky choice. Section 6.2.1 discusses the model-free results on the risky

choices of the respondents in the three numeracy groups (Sub-question 2.1). Section 6.2.2 discusses various modelling criteria necessary for evaluating the modelling results of the thesis. Section 6.2.3 evaluates and compares the results of each of the models across the numeracy groups (Sub-question 2.2), and the accuracy of their estimates and predictions (Sub-question 2.3).

6.2.1. Model-free, preliminary analyses of risky choices of numeracy groups (Research question 2.1)

This section discusses the analyses that addressed the first sub-question on risky choice – Research sub-question 2.1: What risk attitudes are associated with the risky choices of decision makers of different numeracy groups?

This question was framed to examine the risky choices of respondents in the numeracy groups in the absence of any decision model. These analyses are limited to a discussion of individuals' propensity to select safer versus riskier options, using the definition of risk aversion.

As presented in Section 4.4.2.2, linear indifference curves with steeper slopes in Marschak-Machina triangles depict utility functions that indicate more risk aversion under expected utility theory (Machina, 1987). With steep slopes, the differences between the expected values of the riskier options and their safer counterparts is large, and thus the riskier option is attractive. The slopes used to construct the risky choice pairs of the questionnaires covered a wide range, from 0.67 to 4, and thus the choices were sensitive to the risk attitudes across a range of risk aversion levels (Harrison & Swarthout, 2016; Loomes & Sugden, 1998; Machina, 1987; Xiaoxue, 2017). Respondents who selected the safer options in triangles with steeper slopes were more risk averse than those who selected the riskier options. (From here on, the term 'triangles' refers to Marschak-Machina triangles.)

The low consistency levels with expected value predictions may be due to the construction of the choice pairs. As a consequence of construction of choices on indifference curves with increasing slopes, most of which have slopes greater than neutral (1), there were relatively few choices in which the safer option corresponded with the predictions of expected value theory. This imbalance is a feature of measuring risk aversion via the Marschak-Machina method, and was a feature of the questionnaires of the thesis, as well a feature of the elicitation instruments of the studies on which the thesis is based (Harrison & Swarthout, 2016; Loomes & Sugden, 1998). The imbalance undermined the potential success of the

expected value model for risk averse respondents. Stated differently, the choice architecture of the elicitation method partly determined the success of the expected value model, and may also have partly determined the success of the other models considered in the thesis (for example, Andersson et al., 2016, 2018) (see Section 3.5.2). Owing to the many practical requirements of the questionnaires of the thesis, and their time and space limitations, the set of questions could not be equally fair and powerful for all the models considered. Thus, as stated repeatedly, the thesis does not compare the performance of the models to one another – it rather compares the performance of each model in predicting the choices of respondents in the three numeracy groups. Even then, however, different amounts of stochastic error in the responses of the individuals in the numeracy groups may disturb the intergroup comparisons of each model.

In an attempt to control some of the potentially extraneous characteristics in the choice pairs of the GAINS2 and GLM2 questionnaires, the researcher constructed the characteristics of the choices within and between triangles to be as similar as possible, other than in the slopes of their linear indifference curves. For example, she ensured that each triangle comprised four choice pairs with the same number of bottom-edge choices (see Section 4.5.2). The lengths of the chords of all choices within a triangle were kept constant so that the differences between the expected values of the prospects of each choice pair within a triangle were always the same.

As with all the stages of the risky choice analysis, the researcher took measures to control for random response error. In the preliminary analyses, the aggregated responses per Marschak Machina triangle formed the unit of analysis (Machina, 1987). Although analysis at the aggregated level of choices within a triangle disallowed the analysis of bottom edge, and other such effects, the analysis of these prospect characteristics was not among the research questions of the thesis.

In general, low numeracy respondents tended to be more risk averse than high numeracy respondents in the gain frame. There were distinct patterns of decreasing risk aversion for gains, as higher proportions of riskier options were chosen as the attractiveness of the riskier option within a choice pair increased, and also chosen when the highest outcome of the prospects was increased from R200 to R300. All these patterns were stronger for the high numeracy group and weaker for the low numeracy group. The choices in the loss frame represented low risk seeking, contrary to the expectation under cumulative prospect theory (Tversky & Kahneman, 1992). Rather than selecting the riskier option with lower expected

value, respondents generally selected the safer option as the risk of expected losses increased. This pattern was particularly obvious for high numeracy respondents, where risk seeking decreased to a minimum. The pattern was present for the low numeracy group, although to a weaker extent.

In the thesis, the finding of the propensity for selecting safe choices, i.e. for risk aversion, more so in the gain frame by respondents with low numeracy than high, is consistent with the findings of some empirical studies (e.g. Benjamin et al., 2013; Burks et al., 2009; Dohmen et al., 2010; L'Haridon & Vieider, 2019a). However, the relation between risk aversion and cognitive ability or numeracy in the gain frame may be weaker than often claimed, judging by a meta-analysis of relevant studies (Lilleholt, 2019) (see Section 3.4.4). Furthermore, Lilleholt's (2019) meta-analysis found no evidence of this relationship in the loss frame.

Dual process theory may account for the negative relation between risk aversion and cognitive ability (see Section 3.3.1.2). Individuals with higher levels of cognitive ability who engage in Type 2 deliberation may recognise the irrationality of risk aversion with increasing attractiveness of the riskier prospect of the choice pair (Lilleholt, 2019).

The affect mechanism provides an alternative explanation for risk averse attitudes (see Section 3.3.1.3). Pachur et al. (2017a) found that risk aversion in older adults was more strongly linked to affect than to cognitive factors. Despite having lower cognitive ability than the younger participants in the study, older adults were less risk averse than younger adults in gain-framed choices. The older adults' positive affect and motivation may explain their risk taking to win. The affect-choice relation may be related to numeracy as affect appears to take precedence over probabilities in the risky choices of low numeracy individuals (Petrova, van der Pligt, & Garcia-Retamero, 2013). Petrova et al. (2013) found that less numerate respondents were not as sensitive to the probabilities of an affect-rich outcome compared with more numerate respondents.

Affect appears to have been an important motivating factor in the students' participation in the empirical aspect of the thesis. Judging by the students' response rates to the second administration of the questionnaire, and their messages of gratitude for their prizes, they were highly incentivised to win the prizes. In particular, the disproportionately high percentage of safe choices made by the students in the low numeracy group indicated that they typically selected the safer option when the three possible outcomes were R0, R100 and R200. The R200 maximum prize seemed not to have been sufficiently attractive for them to select the

riskier option. As in the study of Millroth et al. (2019), they tended to minimise the risk of receiving the lowest outcome of the options. The high level of risk aversion displayed in the choices may indicate the severity of the financial distress of South African students, particularly in the low numeracy group where the mean income level was lowest. On a much larger scale, the financial challenge of many South African students was amplified with the #FeesMustFall student revolt that began in October 2015. The revolt challenged financial exclusion from South African higher education institutions, among other student grievances (Booysen, 2016).

Before concluding this section, it is restated that the relation between all aspects of risky choice and numeracy are assumed non-causal. Differences in risky choices are not claimed to be caused by differences in the characteristics of the choices, nor by the numeracy levels of the groups, nor by demographic differences among the numeracy groups. Incorporation of demographic covariates is earmarked as an aspect of future research (see Section 7.4).

The section concludes preliminary observations on the risk attitudes of respondents in the numeracy groups. The thesis has thus answered the first of the sub-questions on the relation between statistical numeracy and risky choice in the absence of risky choce models (Sub-question 2.1). The following sections address the modelling of risky choices referring to the three computational models of expected utility, rank-dependence with the functional forms of Tversky and Kahneman (1992) and Lattimore et al. (1992), and cumulative prospect theory.

6.2.2. Modelling criteria

Before discussing the findings pertaining to the research sub-questions on the modelling of risky choice, it is necessary to assess the approach taken in the thesis to key aspects of modelling.

Sub-questions 2.2 and 2.3 required the estimation of parameters for each of the models and functional forms considered. As previously discussed (Section 2.2), parameters of such models are intended to represent underlying risk preferences (Murphy & ten Brincke, 2017). For the parameter estimates to be meaningful and comparable across the three numeracy groups, the parameters of each of the risky choice models, and thus the models themselves, need to be *identifiable*. Model identifiability is the theoretical possibility of determining a unique set of parameter values that provides the best solution for the relation between the outcomes of the model and the observed data (Moran, 2016; Spektor & Kellen, 2018). This means that there cannot be multiple sets of parameters with different values that provide the

best fit to the data equally well. Identifiability is especially important if the parameter values are to be interpreted, and comparisons of their estimates are to be made between groups (Gershman, 2016). Such was the case in the thesis, as parameter values of each of the three decision models were required to be estimated and compared across the three numeracy groupings, and then assessed for their validity within the groups (Sub-questions 2.2 and 2.3, respectively).

Challenges of identifiability are common in cognitive models based on non-linear functions of their parameters. For example, reinforcement learning models that predict choice behaviour from stimuli and rewards present identifiability problems. Biochemical models, and other complex models in the natural sciences, present with challenges of selecting from tens or hundreds of parameters. Such models may be unidentifiable, or if identifiable, they are termed *sloppy* if their parameters are poorly known and oversimplified, with uncertain interrelationships (Brown & Sethna, 2003). As in the thesis, their modelling is complex, and may require Monte Carlo sampling or other optimization methods (Gershman, 2016).

The results of the thesis, presented in the previous chapter, and discussed in Section 6.2.1, indicated several challenges in estimating the model parameters that may relate to problems of identifiability. The preliminary analyses showed that the patterns of risk preferences in the data, for example risk aversion, were less clear in the case of the responses of the low numeracy group than in those of the high, likely indicating greater random error or noise in the responses of low numerate respondents. Greater random error in the parameter estimates of the low numeracy group was also reflected in the relatively high standard deviations of the maximum likelihood parameter estimates for this group, for all the model parameters estimated (Table 18). Variability in parameter estimates is indicative of low parameter sensitivity. When parameters have low sensitivity, they have limited value, as changes in their values have little impact on the model's outcomes, also making them less identifiable (Li, Lewandowsky, & DeBrunner, 1996). Low parameter sensitivity may partly explain why several of the maximum likelihood estimation attempts failed to provide convergent solutions. The Bayesian estimates also proved challenging as the Gibbs samplers required long chains of sampling iterations to meet the criteria of convergence.

Model identifiability is difficult to prove, and is beyond the scope of this thesis. Proving identifiability would require an exercise involving a process of parameter recovery for each of the models and their functional forms, for each numeracy group. The process would entail generating data sets of simulated choices for each model, per numeracy group, using the sets

of parameters devised in the thesis as input. Thereafter, the simulated data sets would be used to estimate the parameters, the values of the parameters recovered from the simulated data, and then compared to those of the generating data (Murphy & ten Brincke, 2017; Nilsson et al., 2011). Other techniques (Moran, 2016) could be used as additional evidence of model identifiability or non-identifiability. As the identification process would require extensive additional work and hours of modelling, it fell beyond the scope of the thesis.

Instead, the researcher undertook a number of measures, throughout the modelling exercises, that facilitated, or enabled, the conditions for the identifiability of the models and their parameters. These measures included consideration of the error component of the models, the intercorrelations of the parameters of the models, independent parameter estimates and informative prior probabilities for Bayesian hierarchical modelling. Finally, she checked the predictive validity of the models using out-of-sample data (Wilcox, 2008).

Each of these measures is discussed in turn, and they form the context for the discussion of the modelling results presented in Section 6.2.3. However, these measures were not sufficient to prove, or guarantee, identifiability (Moran, 2016).

6.2.2.1. Stochastic response error in models

As discussed in Section 2.3, models of risky choice are probabilistic rather than deterministic, and need to incorporate both random response error and random parameter error in their estimates and predictions. This section relates to the control of random response error, while the next section addresses random parameter error.

In effect, the incorporation of random response error in a computational decision model of risky choice means that there are two models operating jointly in the computational modelling of risky choice outcomes. Stated differently, the decision theory is embedded in a model of stochastic choice (Blavatskyy & Pogrebna, 2010). This situation provides an additional challenge to the identifiability of risky choice models (see Section 2.3.3.1), as there is a complex relation in the various combinations of decision and error models. The complexity involves issues of non-compatibility between the models, differences in model fit, and significant differences in the estimates they produce. Thus, parameter estimates should be interpreted with reference to the stochastic choice model (Blavatskyy & Pogrebna, 2010; Wilcox, 2008). Additionally, some error models, for example the Luce probabilistic error model, are parameter based, and so, require estimation. The additional parameters of error models entail the further complication of correlations between the decision and error

parameters. Krefeld-Schwalb et al. (2019) found a correlation of -.61 between the parameter of the value function of the cumulative probability model, and the noise parameter of the Luce choice model.

The thesis approached response error in the modelling by using the Wilcox contextual error model (Wilcox, 2008, 2011, 2015). This model is parameter free, and so did not add to estimation complexity. There were two contexts within the prospects of the GAINS2 questionnaire, and three within the prospects of the GLM2 questionnaire. However, without conducting parameter recovery exercises, the extent to which stochastic response error was controlled by this error model is unknown. This caveat is identified as a possible limitation of the research (see Section 7.4).

6.2.2.2. Stochastic parameter error

The Bayesian approach is used in complex nonlinear modelling in an attempt to avoid non-identifiability problems (Moran, 2016). It treats estimates probabilistically, and could thus accommodate random parameter error in the modelling tasks of the thesis (see Section 4.8.4). Moreover, hierarchical Bayesian analysis produces more reliable and robust estimates at the individual level as individual differences are partially pooled at the group level (Nilsson et al., 2011; Scheibehenne & Pachur, 2015). Once again, further research is required to check the results of the modelling for these properties (see Section 7.4).

6.2.2.3. Bayesian modelling, informative priors and overfitting

Bayesian modelling experts stress the importance of specifying informative priors for resolving statistical challenges related to model identifiability (Lee & Vanpaemel, 2018). As described in Section 5.3, the researcher used the maximum likelihood estimates based on a subset of the data for the prior probability distributions of the hierarchical Bayesian analysis estimates (computed on a different set of data). This method was assumed to provide exchangeable prior information to the current study (Spiegelhalter, 2004). Selecting prior values based on previously observed distributions of values, is also known as the method of empirical priors (Spektor & Kellen, 2018) (see Section 4.8.4.1). The empirical prior method has been used by researchers with mixed success. For example, while it was used successfully by Gershman (2016), Spektor and Kellen (2018) found that the method did not compensate for low information content in their data.

For the thesis, informative priors were essential for the modelling. Without them, the software encountered computational errors traps. However, maximum likelihood estimates

(computed on a different subset of the data) were not found for the more complex models, such as for the cumulative prospect model, and for the rank-dependent model with the probability weighting function of Lattimore et al. (1992). In these instances (only), the researcher attempted to use a process of iteration for prior specification to avoid error traps in MultiBUGS. (The method is described fully in Sections 6.2.3.3 and 6.2.3.4.) Regrettably, this method made double use of the data for prior specification, also known as 'double dipping'. Typically, double dipping results in the MCMC algorithm sampling from narrow regions of the joint posterior distributions, with poor coverage of the parameter space (S. Van Der Merwe, personal communication, February 17, 2020). Furthermore, it encourages the problematic phenomenon of *statistical overfitting*, or fitting a model that is too complex for the available data. An overfitted model captures random error in the modelling data, and thus fits these observations better than out-of-sample data (Van Calster, McLernon, van Smeden, Wynants, & Steyerberg, 2019). The researcher measured the extent of the overfitting, and double use of the data, by the accuracy of a frequentist analysis of out-of-sample predictions (see Section 5.3.5). Deriving alternative informative priors based on independent data for the situations that lacked the maximum likelihood estimates is earmarked for future research (see Section 7.4). However, even if objective priors are obtained, overfitting is a likely outcome when fitting complex models (S. Van Der Merwe, personal communication, February 17, 2020).

The next section discusses an additional modelling challenge in the form of parameter intercorrelations in the joint posterior distributions of the models.

6.2.2.4. Structural interdependence of the model parameters

Structural parameter interdependencies refer to correlations among parameters *within* the model. These structural interdependencies are distinct from correlations computed between the parameters at the *individual level* that arise from associations between the psychological capacities that are measured by the parameters. Examples of individual-level parameter correlations are associations between noise in the response process and noise in the parameters that characterise an individual's preferences, such as probability sensitivity. These associations can distort observed choice proportions, but do not affect the algorithmic structure of the model (Bhatia & Loomes, 2017; T. Pachur, personal communication, November 24, 2020).

There are several reasons why structural parameter interdependence in modelling is contraindicated. First, high correlations between parameters are symptomatic of non-identifiability (Gershman, 2016). Second, structural interdependence obstructs best fitting parameter estimations, and is more likely to result in poor fitting models compared to models with independent parameters (Li et al., 1996). Third, structural interdependence inhibits chain convergence in MCMC sampling. Fourth, highly interdependent parameters may partially compensate for one another (Krefeld-Schwalb et al., 2019; Spektor & Kellen, 2018). As collinearity increases, the standard errors of the estimates increase, parameter sensitivity decreases, and the parameter estimates become untrustworthy (Field, 2013; Krefeld-Schwalb et al., 2019). In such models, the individual parameter values cannot be meaningfully evaluated, nor meaningfully compared across conditions. Correlations with values of .8 or higher create a situation of multicollinearity (Field, 2013), and lead to multiple solutions of parameter estimates with equally good model fit, and hence non-identifiable models. This means that the values of individual parameter estimates cannot be compared as they predict similarly with exchangeable values. In these situations, the parameter correlations do not reflect associations between the underlying psychological processes (Krefeld-Schwalb et al., 2019).

A common way to test structural parameter interdependencies is by plotting the joint posterior distributions for pairs of parameters against each other at the group or individual level (see Scheibehenne & Pachur, 2015, Figure 1, p. 396). However, the computation of the within-model correlations was considered beyond the scope of the thesis.

6.2.2.5. Scope of the models

The approach for controlling extraneous variables in the model comparisons across the numeracy groups was to hold as many factors constant as possible, other than the numeracy grouping variable. To this end, each model was considered separately. The contextual error model was used throughout so that its effect on the parameter estimates of each model could be assumed equal for all the numeracy groups. Additionally, all respondents were subjected to the same risky choice tasks and experimental conditions, so that these factors were controlled in the between numeracy group comparisons.

The deliberate experimental control of all aspects of the modelling and experimental conditions allowed for more powerful between-groups comparisons. However, its tradeoff was a restriction of the *scope* of the models (Rouder & Lu, 2005). The scope of a model

relates to the extent to which its predictions extend beyond the data on which its estimates are based (Li et al., 1996). In this case, the scope of the model was restricted to the three decision models, the contextual error model, the functional forms considered, the choice tasks used, the two contexts in the gain frame, and the three experimental conditions.

As the three decision models of the thesis are frequently cited in the risky choice literature (Harrison & Ross, 2017), the restriction of scope due to the models selected may not be overly serious. An indication of the scope of the models was provided by the accuracy of the out-of-sample predictions of the models in the validity exercise, which included an additional context of prospect outcomes in the gain frame (see Section 5.3.5).

This section concludes the discussion of the considerations of some of the essential modelling principles, as related to the empirical modelling carried out in the thesis. These principles feature in the discussion of the parameter estimates of each of the models and their associated functional forms applied to each numeracy group (Section 6.2.3). Within these sections, the maximum likelihood estimates are discussed, followed by the results of the Bayesian analyses evaluated for convergence of the underlying chains, model fit, and parameter estimates. For multi-parameter models, the intercorrelations of the estimated parameters of each model are examined in terms of linearity, direction and strength. They are then interpreted jointly for each model, per numeracy group, to gauge the complexity of the risk attitudes of the decision makers (Nilsson et al., 2011; Wakker, 2010). Comparisons of all results are drawn across the numeracy groups, thus addressing Research sub-question 2.2. The discussion incorporates the in-sample estimates and out-of-sample predictions of the models based on the Bayesian-estimated parameters, thus also addressing Research sub-question 2.3.

6.2.3. Comparisons of modelling results across numeracy groups (Research sub-questions 2.2 and 2.3)

This section discusses the findings of the analyses that addressed the second and third sub-questions on risky choice – Research sub-question 2.2: To what extent do the estimated values of the parameters of the normative and descriptive risky choice models considered differ across numeracy groups? Research sub-question 2.3: To what extent do the estimates and predictions of the models hold for the risky choices of decision makers of different numeracy groups?

First, the findings on each of the models are discussed in turn (Sections 6.2.3.1 - 6.2.3.4). Thereafter, the findings from these sections that pertain to Research sub-question 2.2 are integrated in Section 6.2.3.5. Finally, the findings that pertain to Research sub-question 2.3 are integrated in Section 6.2.3.6.

6.2.3.1. Expected utility model (power utility)

The expected utility model was evaluated using gain-framed prospects only, as the model is typically used for gains, and does not accommodate a different function or different parameters for losses. The nonlinear power utility function was used for the utility of the expected utility model (and likewise, for the rank-dependent model and the value function of the cumulative prospect model).

Normative economic theory presumes that most rational people are globally risk averse (Quiggin, 1991). However, meaningfully estimated values of the risk aversion parameter using the power function in the risky choice literature are difficult to find, as studies differ in terms of their samples, prospect sets, elicitation methods, analysis methods and other aspects of methodology (Charness et al., 2013; Harrison & Rutström, 2008; Loomes & Pogrebna, 2014; Pedroni et al., 2017a; Zhou & Hey, 2018). Furthermore, the published estimates of the utility parameter are mostly modelled on cumulative prospect theory, and should not be interpreted separately from the other parameters of the model. Nevertheless, Fox and Poldrack (2009) review the parameters of several such studies, all of which have used the power functional form for utility. Estimates of the power utility risk aversion parameter range from .19 in Stott's (2006) analysis to .91 in the study of Abdellaoui (2005, as cited in Fox & Poldrack, 2009, p. 314).

For the initial maximum likelihood estimates, analysis of responses to gain-framed prospects allowed the researcher to use the more detailed set of gain-framed prospects comprising eight prospect pairs from each of five Marschak-Machina triangles of the GAINS1 questionnaire, and a further eight pairs from each of the two triangles of the GLM1 questionnaire. However, as the sample sizes per numeracy group from the first administration of the questionnaires were small (Table 15), their contributions were limited, and the data was sparse. Still, the maximum likelihood analyses produced a pattern of utility parameter estimates consistent with the evidence of overall risk aversion in the results of the preliminary analyses (Sections 5.3.1 - 5.3.2). The respective maximum likelihood estimates for the risk aversion parameter of the high, middle and low groups were .44, .35 and .23, respectively (Table 18). These

values are far from the linear utility value of 1. They indicate high risk aversion in all groups, with highest risk aversion (most different from 1) for groups with low statistical numeracy. The differences between the pairwise utility parameter values of the numeracy groups were all significant ($p < .05$), although with weak effect sizes.

As described in Section 5.3.4.2, the researcher assumed approximate convergence of the MCMC chains as she had no reason to suspect that the parallel chains had not produced a convergent solution. The Bayesian estimates of the means of the utility parameter had the values of .38, .24 and .21, respectively, for the high, middle and low numeracy groups (Table 19). Although these values are lower than the corresponding maximum likelihood estimates, both sets of estimates indicate high risk aversion. Furthermore, both sets shared the same rank order, and confirmed the pattern observed in the preliminary analyses of the previous section (see Sections 5.3.1 and 5.3.2). The differences between the posterior and prior distributions of the utility parameter for each numeracy group indicated that the data of each numeracy group had provided additional information for the estimation. This is an indication of parameter identifiability (Ntzoufras, 2009).

The fit of the expected utility model to the observed choices was best for the high numeracy group, based on the DIC value for this group (DIC = 2903) that was substantially lower than the DIC values for the other groups (DIC = 3297 and 3294 for the middle and low numeracy groups, respectively) (Table 19). The validity of the model was calculated by the percentage of model predictions that were consistent with the observed choices (safer or riskier) for those prospects. The consistency of the in-sample model estimates with observed choices was similar for the numeracy groups at 67.4%, 66.5% and 63.0% in the R0, R100 and R200 context for the high, middle and low numeracy groups, respectively (see Section 5.3.5.2, Table 25 and Table M 2). These consistency rates are significantly greater than chance prediction (50%), and translate mostly to medium effect sizes relative to chance. The consistency rates of the out-of-sample predictions with observed choices were similar to the in-sample consistency levels when both success rates were based on the R200 context, (65% - 69%) (Table M 2). There was thus no indication of model overfitting.

However, the consistency rate in the R0, R100 and R300 context (58% - 61%) was significantly lower than in the R200 context. This finding suggests that the model may have low scope, i.e. that the predictive success of the model is inferior when evaluated relative to prospects with different contexts (see Section 6.2.2.5).

When the utility parameter means of the posterior joint distributions are compared at the individual level across the numeracy groups using frequentist analysis, the differences are highly significant ($p < .001$), with large Cohen's d effect sizes ($d = 2.7$, 4.0 and 0.9 for the high-medium, high-low and medium-low numeracy group pairwise comparisons, respectively). However, these comparisons may be excessively powerful and exaggerated, as they are based on restricted standard errors within each numeracy group. The restricted standard errors are due to shrinkage, a property of hierarchical Bayesian analysis (see Section 4.8.4.2). It may be fairer to consider the uncertainty in the utility parameter estimates in the form of the 95% credible interval of each group, i.e. the range within which 95% of the utility parameter values of the posterior distribution lie (Boehm, Marsman, Matzke, & Wagenmakers, 2018; Heathcote, et al., 2019). A better approach would have been to use the group-level posterior distributions of the hierarchical estimation as a basis for comparing the numeracy groups (T. Pachur, personal communication, November 24, 2020). However, the finding that these utility intervals do not overlap provides reassurance of meaningful differences between the utility parameter values of the groups. Further, their ordering shows, once again, highest risk aversion for the low numeracy group, and lowest risk aversion for the high numeracy group.

In summary, the expected utility model with Bayesian-estimated parameters for gain-framed prospects showed high levels of risk aversion in the respondents' choices. Overall, the levels of risk aversion for all the numeracy groups are higher than those typically found in the risky choice literature. The fit of the models to the empirical choice data in the domain of gains was best for the high numeracy group, based on DIC difference values. The pattern observed in the results on the estimated utility parameter of the model is that respondents in the low numeracy group have greatest risk aversion, and those in the high numeracy group have the least.

In post-hoc validity checks of the expected utility model, no significant differences were observed between the numeracy groups on the accuracy of its in-sample estimates, nor on the accuracy of its out-of-sample predictions. The model estimated and predicted approximately 65% of the the risky choices of the students correctly, and did not show evidence of overfitting. However, the model did show evidence of restricted scope, as its predictions of prospects out of the context of the modelling data were less consistent.

6.2.3.2. Rank-dependent utility model (power utility, and Tversky and Kahneman, (1992) weighting function)

As previously discussed (see Section 2.1.8), the rank-dependent model is a generalisation of the expected utility model as it adds a nonlinear probability weighting component to the expected utility model. Accordingly, it ascribes less importance to the risk aversion component of risky choices, and less support for the presumption that most rational people are globally risk averse (Quiggin, 1991). It also explains the probabilistic risk attitudes of optimism and pessimism (see Section 2.2.2.2). Diecidue and Wakker (2001) describe several sources of evidence in the psychological and economics literatures for a strong intuitive foundation for the concept of the weighting of outcomes, based on their ranks in decision making (see Section 6.2.3.2).

Like expected utility theory, rank-dependent utility makes no distinction between decisions based on outcomes in terms of gains and losses. However, several researchers in the field extend the theory to incorporate prospects in the loss frame as well as in the gain frame, which has the potential for furthering the explanatory power of the model. For example, ratios of probability weights for losses relative to gains informs the phenomenon of *probabilistic* loss aversion. This concept is the probabilistic equivalent of utility loss aversion, based on the asymmetric utility weightings of losses relative to gains (Harrison & Swarthout, 2016, citing Schmidt & Zank, 2008; Wakker, 2010). The inclusion of losses in the rank-dependence modelling in the thesis also provided a logical progression of modelling results, from the expected utility model applied to gain-framed prospects, to the cumulative prospect model on prospects in the gain, loss and mixed frames.

In the thesis, maximum likelihood estimates were computed for rank-dependence based first on gain-framed prospects, and then on prospects in both gain and loss frames. The same power parameter was used for utility in both frames as advised by researchers in the case of the power function used in cumulative prospect theory (Nilsson et al., 2011; Wakker, 2010). The single-parameter probability weighting function of Tversky and Kahneman (1992) was used. Maximum likelihood solutions were not found in analyses that sought separate estimates for the probability weighting parameter for prospects in the gain and loss frames (see Section 5.3.3). Thus, the researcher assumed the same probability weighting parameter in both frames to arrive at a maximum likelihood solution. For all analyses, the maximum likelihood estimates of the utility parameter for the rank-dependent model were similar to the expected utility model in their means and standard errors (Table 18). As previously, the

estimates for loss and gain-framed prospects decreased from the high to the low numeracy groups, indicating increasingly strong risk aversion across the groups (.35, .26, .10 for the high, middle and low numeracy groups, respectively). By contrast, the estimates for the gamma weighting parameter for each of the numeracy groups did not show differences between the numeracy groups ($p > .05$). The gamma parameter values all approached 1 (values between 0.85 and 0.90), indicative of weighting functions that were near linear, but still followed the typical inverse *S* shape, as theorised (Tversky & Kahneman, 1992).

The picture of the results derived from the MCMC analyses is more complex. In these analyses, the probability weighting function parameter was estimated separately for gains and losses. There was stronger evidence of convergence of the chains for the high numeracy group than for the middle and low numeracy groups. The upper confidence limits of all Gelman-Rubin indicators were below 1.1 for the high numeracy group, but marginally higher at 1.11 for 1 of the 5 estimates for the middle numeracy group, and at 1.12 for 2 of the 5 estimates for the low numeracy group (Table 20 and Table 23, respectively). The comparisons relative to the middle and low numeracy groups are thus made with caution (see Section 5.3.4.2). Furthermore, as for the expected value model, the fit of the model was best for the high numeracy group based on differences in DIC values greater than the guideline of 10.

As previously, the results of the analyses for the three numeracy groups showed the pattern of increasing risk aversion from the high to the low numeracy groups. The inverse *S*-shaped probability weighting curve is more obvious for the high numeracy respondents, and more so for the gain frame (mu.gamma) than for the loss frame (mu.delta). This curve indicates higher probability sensitivity for high and low probabilities, and lower sensitivity for probabilities in-between. It also indicates pessimistic risk attitudes for high probabilities and optimistic risk attitudes for low probabilities (see Section 2.2.2 and Figure 9).

However, two complications underly the interpretation of the results. First, the Tversky and Kahneman (1992) function uses a single parameter, and thus makes no distinction between the elevation and curvature of the function (see Section 2.2.2.2). There may thus be composite processes reflected in the parameter. The second complication relates to the high correlation between the alpha and gamma parameter estimates of the model ($r = .60, .42$ and .47 for the high, middle and low numeracy groups, respectively, for the gain frame) (Table 24). These parameter intercorrelations at the individual level may have resulted in some confounding in the interpretation of the psychological processes represented by the parameter

estimates, particularly within the high numeracy group where the correlation is the highest (see Section 5.3.4.5). Furthermore, the resilience of the model in explaining results at extreme levels of risk aversion is unknown. However, as previously, the utility parameter is ordered from high to low, indicating relatively lowest risk aversion for the high numeracy group and highest risk aversion for the low numeracy group.

Approximately two-thirds of the model estimates are consistent with the observed choices of the in-sample data (67% - 70%) in the gain frame, with no significant differences found between the numeracy groups (Table 25 and Table M 3). The consistency rates of the model estimates with observed choices for the out-of-sample predictions in the gain frame (64% - 69%) were not found to differ between the numeracy groups, and are similar to the in-sample estimates and predictions. There is thus no evidence of model overfitting. However, the consistency levels of the model predictions in the R300 out-of-context out-of-sample data dropped significantly to 60% - 62% (Table M 4). This finding suggests that the model may have low scope, i.e. that the consistency levels of the model predictions are less applicable to different treatment levels than to those on which the modelling is based (see Section 6.2.2.5). However, the model's consistency levels with observed choices are still greater than chance (50%).

As in the case of the expected utility model, frequentist comparisons of the utility parameter means of the posterior joint distributions among the numeracy groups show highly significant differences ($p < .001$), with large Cohen's d effect sizes ($d = 1.8$, 2.9 and 3.0 for the high-middle, high-low and middle-low numeracy group pairwise comparisons, respectively). Furthermore, the 95% credible interval of the high versus low numeracy groups are largely distinct (.23 - .38 versus .20 - .24).

Consistent with the results of the preliminary analyses that showed respondents selecting safer choice options in loss-framed prospects, there was little support for the Tversky and Kahneman (1992) probability weighting function that posits risk seeking in the loss frame. Thus, the model's in-sample estimates in the loss frame faired significantly worse than chance, particularly for the middle and low numeracy groups. The data on losses was too sparse to compute out-of-sample validity estimates for losses. (It should be recalled that there were fewer choices for prospects in the loss and mixed frames than for gains, as there was no equivalent of the GAINS2 questionnaire devoted to losses or mixed prospects.) It is also possible that there was greater error in the risky choices of lower numeracy respondents,

particular in the loss frame. This error may not have been accommodated sufficiently in the modelling, and may have accounted for the model's poorest performance for this group.

In summary, the conclusion drawn for the rank-dependent model, with power utility and Tversky and Kahneman's (1992) probability weighting function, is that of strong risk aversion for all groups, strongest for the low numeracy group. There is also empirical evidence of the theoretical inverse *S*-shaped weighting curve with associated risk attitudes of pessimism for high probabilities and optimism for low probabilities. This evidence is strongest for the high numeracy group. The fit of the model and confidence in the estimates of the analyses are best for the high numeracy group. However, the parameter estimates may be untrustworthy, to some extent, owing to the high parameter intercorrelations. For positively framed prospects, the models' outcomes are consistent with observed choices in approximately two-thirds of choices, both in their in-sample estimates and out-of-sample predictions. This rate of success is significantly better than chance, and equates to medium to small effect sizes. Although the model shows no signs of overfitting, there are indications that it is restricted in scope as its success rate drops in the R300 context. However, the success rate of model estimates in the loss frame was poorer than would be predicted by chance, particularly for the middle and low numeracy groups.

The next model considered both the confounding of elevation and curvature in the probability weighting function, and high parameter intercorrelations. It did so by assuming linear utility, and using the 2-parameter weighting function of Lattimore et al. (1992) (see Section 2.2.2.3).

6.2.3.3. Rank-dependent utility model (linear utility, and Lattimore et al. (1992) weighting function)

The second functional form for probability weighting considered in the thesis for the rank-dependent model, was the inverse *S*-function of Lattimore et al. (1992). As previously described (see Section 2.2.2.3), this function has parameters for curvature (γ) and elevation (δ), estimated separately and independently of each other. The psychological interpretation of the curvature parameter is probability discrimination, or probability sensitivity. As the inverse *S*-shaped function has greatest curvature around probabilities of 0 and 1, sensitivity to probabilities diminishes more rapidly at these extremes, with less sensitivity to probabilities in the middle of the range. A curvature parameter value of 1 represents a linear function of constant sensitivity to changes in probability, i.e. no distortion. The elevation parameter is

interpreted as the attractiveness of the gamble, assessed by the extent to which the curve is underweighted or overweighted relative to a 45° line (Gonzalez & Wu, 1999).

Although 2-parameter models are preferred to single-parameter models in terms of interpretability and fit to data (Cavagnaro et al., 2013; Gonzalez & Wu, 1999; Xing et al., 2019), they present the complexity of estimating an additional parameter, or an extra two if these parameters are estimated separately for gains and losses. Following the advice of P. Wakker (personal communication, June 20, 2019), the researcher estimated the parameters of the Lattimore function with the risk aversion parameter fixed at 1. This strategy addressed the problem of the intercorrelation of the parameters with the utility parameter at the individual level, and reduced the number of parameter estimates. However, despite the restriction, maximum likelihood estimates for the two parameters were not found, due to sparse or missing data in certain regions. Without informative prior specification for the hierarchical Bayesian analyses, the MultiBUGS software gave traps early in the analyses, or non-convergence of chains after long MCMC sampling iterations.

In this untenable situation, the researcher used an initial prior point estimate of 0.8 for the gamma parameter, in both gain and loss frames. This value was approximately in the middle of the range of gamma values found in several studies reviewed by Fox and Poldrack (2009). For the delta parameter, a relatively low value of 0.50 was specified for the initial prior mean, based on the consistent finding of high risk aversion in all the maximum likelihood analyses of the thesis (on the data not used for the Bayesian estimates). Starting with these values, the researcher reduced these prior values successively for the MCMC simulations on the data of each of the three groups, using 31 chains and long sampling periods. Each set of results was inspected for convergence via the CODA plots and the Gelman-Rubin upper confidence limits for all the parameters, and considering differences in DIC values for model fit. For each group, there was a point in the runs at which the DIC values reduced and then increased, and for which the upper limits of the Gelman-Rubin values were less than 1.1. This way, the analyses with the best fitting model estimates for the data were identified. For each numeracy group, the statistics of the posterior distributions of the parameters were used to estimate the model.

Unfortunately, this method made double use of the data for prior specification, with the unfortunate consequences discussed in Section 6.2.2.3. First, the values of the posterior mean parameter values were almost the same as the values specified in the priors. Coinciding posterior and prior distributions indicates redundant parameters, as there is no additional

information in the data to update the prior distributions that result in the posterior distributions, according to Bayes' rule (Heathcote et al., 2019). This redundancy indicated non-identifiable parameters (Ntzoufras, 2009), given the data at hand. Other aspects of the results complemented the bleak picture. Validation of the in-sample estimates in the gain frame revealed that the models correctly predicted 60% and 56% of the choices of the high and middle numeracy respondents, respectively, and only 53% of the choices of the low numeracy respondents (Table 25 and Table M 3). These limited success rates, relative to chance prediction (50%), translated into small effect sizes for the high and middle numeracy groups, and a negligible effect size for the low numeracy group. The predictive validity success based on the out-of-sample data using the same context was worse, and even worse when evaluated against the out-of-sample data using the R300 context (R0, R100, R300), indicative of both model overfitting and restricted model scope (Table M 4). Although convergent MCMC solutions had been obtained, the sampling from the joint posterior distribution had clearly been restricted to a narrow region, and access to the best combination of parameter values had not been achievable.

In view of these results, there is no further discussion of the modeling with the Lattimore et al. function under rank dependence.

6.2.3.4. Cumulative prospect theory model [power utility, and Tversky and Kahneman (1992) weighting function]

Maximum likelihood parameter estimates, based on the cumulative prospect theory model with power utility function and Tversky and Kahneman's (1992) probability weighting function, were derived only for the high numeracy group (Table 18). As found consistently throughout the results of the thesis, the value of the utility power parameter was considerably low, with maximum likelihood estimate of .063 (*SE* = .065), reflecting extreme risk aversion. The estimates are markedly lower than those obtained by Tversky and Kahneman (1992) and in several other studies based on the same functional forms used in the thesis. Scholten and Read (2014) found similar extreme parameter values using maximum likelihood estimates, with values of 0.009 in the gain frame, and 0.002 in the loss frame. Harrison and Swarthout's (2016) estimates were also low (.21 and .06 in the gain and loss frames, respectively).

The maximum likelihood estimate of the loss aversion parameter found for the high numeracy group had a value of 1.262 (*SE* = .501). This value is lower than the Tversky and Kahneman (1992) value of 2.25, and lower than found in several other studies. It was,

however, greater than 1, thus showing the inverse *S*-shaped function consistent with the theoretical literature on probability weighting. The estimated curvature parameter (0.791, *SE* = 0.135) was in the range of published values. However, as discussed in Section 6.2.2, parameter estimates are not necessarily comparable across studies due to different error model specifications and methodologies, and parameter intercorrelations. Furthermore, they should not be considered isolated from the other parameter estimates.

As previously, the maximum likelihood estimates for the high numeracy group served as the priors for the Bayesian analysis of the high numeracy responses (using a different set of choices). For the middle and low numeracy Bayesian priors, the researcher used the same ill-fated approach for specification as used for the rank-dependent model based on the Lattimore weighting function in the previous section. This time however, the maximum likelihood estimates of the high numeracy group were used as the starting values of the priors for the iterative approach to the MCMC analyses of the middle and low numeracy groups. In addition, the mean of the prior probability distribution of the value parameter was adjusted to the alpha value (0.35) of the maximum likelihood estimate found for the previous rank-dependent model. The parameter values were then adjusted iteratively and served as priors in a sequence of MCMC sampling iterations. As previously, each set of Bayesian estimates was inspected for chain convergence and model fit.

The results of the MCMC analyses for the three numeracy groups (Table 22) revealed the usual pattern of high risk aversion, with highest parameter estimate for the high numeracy group, and lowest for the low numeracy group. Furthermore, consistent with the phenomenon of loss aversion, the parameters reflecting the curvatures of the probability weighting functions for losses were more linear than for gains, within each of the numeracy groups (Table 22). This pattern is commonly found in the modelling literature (Abdellaoui, 2000; Harrison & Swarthout, 2016; Scholten & Read, 2014; Tversky & Kahneman, 1992). The parameter values reflected decreasing linearity of the probability weighting functions, and thus increased probability distortions, from the high to the low numeracy groups. Furthermore, for gain-framed lotteries, the consistency of the model predictions with observed choices were satisfactory, considering both the in-sample and out-of-sample data. The consistency rate for the model's estimates and predictions with observed choices ranged from 63% to 70%, and slightly lower for the R0, R100, R300 context (Table M 6). Relative to chance (50%), these success rates translated into small or medium effect sizes (Table 25).

However, as in Section 6.2.3.3, the posterior and prior distributions were similar, with only small differences, mainly on the curvature parameters. The remainder of the findings placed serious doubts on the cumulative prospect theory model with the single parameter probability weighting function. First, the individual-level parameter intercorrelations are extremely high, particularly for the low numeracy group, between the value parameter (alpha) and the curvature parameter for losses (delta) (r = .74) (Table 24). The correlation implies a strong association between the psychological capacities measured by the parameters and are likely compensating. Furthermore, as per the model predictions of the rank-dependent model with the same probability weighting function, the loss-framed prospects were predicted poorly, worse than chance in the loss and mixed frames (Table M 7 and Table M 8, respectively).

Despite the model's reasonably successful outcomes in the gain frame, its dramatically poor estimates and predictions in the loss and mixed frames suggest that it should be disregarded from further discussion. Its key concept of loss aversion is dependent on the asymmetry of the value function for both losses and gains relative to the reference point, and likewise on the asymmetry of the curvature of the weighting functions for losses and gains (see Section 2.1.9). Based on these findings, it appears that the model, using the Tversky and Kahneman utility and probability weighting functions, is not identifiable for the three numeracy groups based on the available data.

In the following sections (6.2.3.5 and 6.2.3.6), the discussions of Section 6.2.3 are synthesised into answering Research question 2.2 on the comparison of each model's characteristics across the numeracy groups, and Research question 2.3 on comparing each model on its in-sample estimates and out-of-sample predictions across the numeracy groups.

6.2.3.5. Summary of comparisons of the models' characteristics across the numeracy groups (Research sub-question 2.2)

Research sub-question 2.2: To what extent do the estimated values of the parameters of the normative and descriptive risky choice models considered differ across numeracy groups?

This question was addressed by the Bayesian-estimated parameters in terms of model fit, convergence, parameter estimates and intercorrelations. Although the thesis intended to answer this question by referring to all the models and functional forms considered at the outset, the rank-dependent model with the Lattimore function and the cumulative prospect model were unidentifaible. Thus, the comments in this section pertain to the expected utility model for gain-framed prospects, and to the rank-dependent model with the single parameter

probability weighting function of Tversky and Kahneman (1992) computed for gain- and loss-framed prospects. Both models used a power transformation for utility. As previously stated, the thesis examines only comparisons across numeracy groups within models; it does not compare the models within any or all of the groups.

Both models were best fitted to the empirical choice data of the high numeracy group, based on lower DIC values compared to those of the lower numeracy groups. The estimated values of the utility parameters of both models were substantially lower than 1 for all groups, indicating considerable curvature of the utility power function and high risk aversion. Utility parameters were lowest for the low numeracy group, indicating highest risk aversion for this group. This finding is consistent with the model-free results of the preliminary analyses, which showed low numeracy respondents the most reluctant to opt for risky choice options. The pairwise group difference between the numeracy groups on the utility parameters were significant with strong effect size, based on frequentist analysis of posterior utilty parameter means.

Of the three numeracy groups, the parameter estimates of the rank-dependent model (utility, curvature in the positive frame, and curvature in the negative frame) for the high numeracy group most closely resembled the theoretical expectation of risk aversion, together with an inverse *S*-shaped probability weighting curve (Gonzalez & Wu, 1999; Lattimore et al., 1992; Tversky & Kahneman, 1992). For all groups, the high individual-level correlations among its parameters indicate substantive associations and possible confounding of the psychological capacities involved in risky decision making, as measured by the model parameters (see Section 6.2.2.4).

Taken collectively, the estimated parameter values differ among the numeracy groups, with parameter interdependencies at the individual level complicating the interpretation of the psychological capacities measured by the parameters.

6.2.3.6. Summary of comparisons of the validity of the models' estimates and predictions across the numeracy groups (Research sub-question 2.3)

Research sub-question 2.3: To what extent do the estimates and predictions of the models hold for the risky choices of decision makers of different numeracy groups?

This question was answered by considering the findings on the validity of the estimates and predictions of the models. Post-hoc validity checks were conducted on the Bayesian-estimated parameters of the expected utility model for gain-framed prospects, and on the

rank-dependent model with the single parameter probability weighting function of Tversky and Kahneman (1992) for gain- and loss-framed prospects. The two models successfully estimated in-sample choices, and predicted out-of-sample choices, about two-thirds of the time for lotteries in the gain frame. These estimates and predictions were based on parameter estimates specific to each numeracy group. For both the expected utility model, and the rank-dependent model for prospects in the gain frame, no significant pairwise numeracy group differences were found on the accuracy of in-sample estimates, nor on the accuracy of out-of-sample predictions. There was no evidence of overfitting. These success levels corresponded to medium-to-low effect size, based on Cohen's *d* statistic. There was, however, evidence of a restriction of scope in both models, as their successful outcomes were lower when the context of the lotteries was changed from a maximum prize of R200 to R300.

The expected utility model was not tested in the loss frame. However, the rank-dependent model was tested for losses, and performed poorly, i.e. significantly below the level of chance in the case of the middle numeracy group, and worse for the low numeracy group. It is speculated that the responses of the low numeracy group may have comprised the highest amount of random error, which may not have been controlled sufficiently by the contextual error model. Inconsistent responses have been linked to lower cognitive ability and lower numeracy (Benjamin et al., 2013; Burks at al., 2009; Dohmen et al., 2018). The data in the loss frame was sparse, and out-of-sample predictive validity checks could not be computed. Considering the preliminary, model-free analyses, there was little-to-no evidence of risk-seeking for losses, as proposed by Tversky and Kahneman (1992), and thus poor support for the weighting functions proposed in the loss frame.

Overall, the findings in the gain frame imply that each risky choice model requires parameters that differ in their values for the numeracy groups. Given these estimates, the models can produce estimations and predictions of risky choices to a similar level of accuracy.

This section concludes the discussion on the findings that addressed Research question 2. The public policy implications of the findings are discussed next.

6.3. Policy implications of the research findings

Behavioural change programmes related to cognitive ability may be linked to Herbert Simon over 40 years ago. As previously discussed, Simon (1977/1992) incorporated the phenomenon of individuals' low numeracy levels in proposing a process theory on *bounded*

rationality. In Simon's theory, the decision maker is limited by her cognition, information and time available (Simon, 1955, 1976) (see Section 1.2). Following Simon, two fundamentally different research programmes emerged on bounded rationality in research on judgment. Both programmes were aimed at changing behaviour, and both are pertinent to individuals' numeracy levels.

6.3.1. Behavioural change programmes: Heuristics and biases versus simple heuristics

The first of the two programmes to emerge was the *heuristics and biases* programme of Kahneman and Tversky, in the early 1970s. The second was the *simple heuristics* programme of Gigerenzer and colleagues, which emerged in the late 1990s (Gigerenzer & Edwards, 2003; Grün-Yanoff & Hertwig, 2016).

6.3.2. Policy initiatives: Nudge versus boost approaches

Each of these research programmes has led to a policy for changing behaviour. The heuristics and biases programme has led to the *nudge* approach (Sunstein & Thaler, 2003; Thaler & Sunstein, 2008) (see Section 3.1.3). The nudge approach attempts to influence the boundedly rational individual to change her behaviour voluntarily in a way that will improve her decisions, and therefore, her welfare. According to Thaler and Sunstein, an individual – for example a boundedly rational decision maker – would be nudged in the direction of improved welfare, as determined by a *benevolent planner* or *choice architect*. This view is known as *libertarian paternalism* (Sunstein & Thaler, 2003). A parent, employer, and government agency are examples of such. Furthermore, Thaler and Sunstein conceptualise the notion of nudging as associated with choice and sovereignty, with a minimum of intrusion and coercion (Dhami, 2016; Thaler & Sunstein, 2008).

By contrast, the simple heuristics programme has led to a policy referred to as *boosting*. The boost approach advocates improving the decision-making competence of the boundedly rational decision maker. The policy interventions are aimed specifically at the individual's skills and knowledge (Grün-Yanoff & Hertwig, 2016). There are two strategies in which boosts may be applied to decision making under risk: the first strategy operates through changing the environment; the second, through extending the individual's decision-making skills or competencies. An example of the first strategy is depicting statistical information on risk as absolute frequencies, instead of as relative probabilities (Gigerenzer et al., 2008). Another example is presenting risk information as an absolute risk reduction, instead of a relative risk reduction. For example, a message about the reduction in prostate cancer

mortality rates through prostate screening could be presented as an absolute risk reduction of "1 in 1410" patients, or as "a reduction from 5 to 4 out of every 1000 men", instead of presenting the relative risk reduction of 20% (Multmeier et al., p. 43). The second strategy is to enhance individual's statistical literacy and their financial literacy skills through initiatives, for example, through adult education programmes and additional programmes in schools – discussed at length below. [There are other policy approaches, for example, the contractarian approach of Sugden (2011), asymmetric paternalism, and light paternalism (Dhami, 2016). However, this discussion is limited to the nudge and boost approaches.]

It is beyond the scope of the thesis to present the differences between the policy approaches fully. However, the main differences between the two approaches can be summarised through their assumptions (Grün-Yanoff & Hertwig, 2016). The nudge approach assumes that the decision maker is unaware of her cognitive errors, and cannot knowingly and successfully control or override them. Thus, her competence and motivation to acquire skills for correcting errors are irrelevant, and the policy designer or benevolent planner who *is* aware of her errors and goals, and is less error-prone, influences her decisions. By contrast, the boost approach assumes that the decision maker *is* aware of her errors and *can* correct them through education and training. The specific goals of the decision maker are irrelevant to the policy planner (Grün-Yanoff & Hertwig, 2016). Thus, the nudge and boost approaches have been dubbed *persuasion* versus *information*, respectively (Multmeier et al., 2014).

Support is divided on the nudge and boosting approaches. Although Thaler is a Nobel laureate for his work on nudge theory, his approach has attracted complex philosophical questions on politics, that is, on the legitimacy of the state to limit the liberty of individuals (Wertheimer, 2002). On the other hand, there are complex issues associated with the boosting approach too. Boosting could be seen as a perpetuation of dominant Western research paradigms and associated with colonialism (Chilisa & Kawulich, 2012). This is because boosting, unlike nudging, involves *educating* the decision maker, and education is central to the colonialist enterprise.

Putting aside these philosophical concerns, the next section discusses how a boosting approach might be implemented in South Africa's education policy.

6.3.3. Policy initiatives related to numeracy in South Africa

As discussed in the introduction of the thesis, the classical expected utility theory framework is often used as a guide for public policy. The framework assumes unlimited rationality and

perfect numeracy. However, the results of the thesis are inconsistent with this assumption. Indeed, the thesis shows that decision makers with compromised levels of numeracy display increased levels of various risk distortions as modelled by risky choice utility theories. Thus, policy that is modelled on the utility framework will be biased against individuals of lower numeracy levels.

In view of the longevity of the use of the classical expected utility theory in framing public policy, it may be unrealistic to expect a change to the theoretical base of policy in the foreseeable future. Instead, lawmakers and politicians could address the inherent numeracy-related bias by focusing on gradually raising the levels of numeracy in the population.

However, the challenge of raising numeracy levels in the South African population is enormous and would take a generation, or several, to effect. Low numeracy levels are one of the consequences of the inferior quality of the Bantu Education system under Apartheid that entrenched racial inequalities and poverty. Twenty-five years after the democratisation of South Africa, high levels of poverty and inequality remain, with three-quarters of public schools impoverished (Graven, 2014; Jojo, 2019). Differences in numeracy competence across socio economic levels present among learners from the start of formal schooling in the Basic Education system (Grades R-12) (Aunio, Mononen, Ragpot, & Törmänen, 2016). Compromised numeracy skills further challenge learners' capacity to learn mathematics. This situation is exacerbated by an overburdened public school system and by mathematics teachers' own compromised mathematics abilities and teaching competencies (Kanyane et al., 2018; Jojo, 2019; Venkat & Spaull, 2015).

Furthermore, stringent national policy on promotion requirements trumps the 40% end-of-year mark criterion for passing mathematics. This policy of promotion dictates that a learner may repeat only one year within each education phase (Grades R-3, 4-6, and 7-9), and then must be promoted to the next grade or phase. This policy results in some learners progressing through Grades R-9, the compulsory General Education and Training band, and exiting the school system with mathematics marks as low as 20%. The non-compulsory Senior Education band (Grades 10-12) offers an alternative and less demanding mathematics curriculum known as Mathematical Literacy, but once again, both the promotion policy and 20% pass threshold apply (Adler & Pillay, 2017; The Conversation, 2020).

To address the problems with this educational system, at least two changes would need to happen. First, the quality of mathematics education that learners receive needs to improve. To

this end, non-governmental organisations, public-private partnerships, and large and small companies are working intensively with rural and underprivileged schools on literacy and mathematics from early grades. Their offerings include outreach programmes, freely downloadable instructional materials, and online teacher resources. Given adequate information and communications technology, and the support at national, provincial and regional levels of the South African government, such initiatives may address poverty and inequality, and ultimately an increase in the numerical competence of the future generation of South Africa (Bridge, 2020; Nelson Mandela Institute, 2020; Spaull, 2020).

Second, the promotion policy, whereby learners with mathematics marks as low as 20% are allowed to pass, needs to be changed. Learners with inadequate mathematical ability should not be allowed to progress to higher grades. However, with an improved mathematical education system (as discussed above), there will be fewer learners with failing marks.

Widespread, long-term empowerment, or *boosting*, of the mathematical ability of the South African population by improving the education system will take years to result in more numerate adults – at least as long as it takes for a generation of learners to progress from Grade R to matric. For those adults who have already progressed through an inadequate education system, boosting and nudging approaches could be used for improved decision making under risk, as discussed above.

Chapter 7: Conclusions

In this last chapter of the thesis, the conclusions of the research are discussed in four main sections. The first section answers the thesis question (Section 7.1); the second section examines the contributions of the thesis (Section 7.2); the third section discusses the strengths and weaknesses of the research (Section 7.3); and the fourth section provides recommendations for future research (Section 7.4). The epilogue concludes the thesis (Section 7.5).

In the first section, the thesis question is answered using the theoretical framework presented in Section 3.8 to reflect on the method and results of the research. This framework is based on the Toulmin structure of a scientific argument (Toulmin, 2003, as cited in Van de Ven, 2007). As previously outlined (Section 3.8), the Toulmin structure comprises six elements: (i) background, context or problem; (ii) the claim, hypothesis or proposition; (iii) reasons, major premise or logic of the claim; (iv) evidence, minor premise or backing data; (v) qualifiers, boundary conditions or assumptions; and (vi) reservations or limitations. All these elements are updated here, with emphasis on the last three, i.e. on the empirical evidence of the thesis, the qualifiers, and the reservations. The last element, i.e. reservations or limitations, is relabelled 'reservations and reflections', as it incorporates the researcher's reflections on the thesis.

7.1. The claim of the thesis: answer to the thesis question

The main claim of the thesis is that statistical numeracy plays a role in computational models of risky choice. In particular, models of risky choice applied to respondents of different numeracy levels are differentiated on their utility parameter that reflects risk aversion. This claim applies to the expected utility model, and the rank-dependent model with the Tversky and Kahneman (1992) single-parameter probability weighting factor in the gain frame. Both models use a power function for the utility function.

The claim also answers the thesis question: *What is the role of statistical numeracy in models of risky choice?* The reasoning behind this claim is based on the Toulmin structure.

7.1.1. The background to the claim: The research questions

The background to the claim was provided in the previous chapters. It is summarised briefly here, with reference to the thesis question and the two research questions.

Addressing the thesis question depended on the use of a valid measure of statistical numeracy for allocating respondents to numeracy groups. As a suitable measurement instrument was not readily available (see Section 3.7.3), its derivation formed the first of two main sections of the thesis, prior to the analysis of the risky choice modelling.

The task of the numeracy analysis was to derive a scale of numeracy which provided psychometrically defensible numeracy groups. The analysis needed to answer the first question of the research: *What is the structure of the respondent groupings based on numeracy?* It answered this question by deriving a 9-item measurement scale that reflected two oblique factors, the first named 'probabilistic operations', and the second named 'probabilistic reasoning' (see Section 6.1.1). Based on these factors, and using a factor mixture analysis model, the analysis revealed three clear latent numeracy groups: a numerically rational group (the high numeracy group); a numerically operational group (the middle numeracy group); and a non-numerical group (the low numeracy group). This realisation prepared the way for fitting the computational models of risky choice to the responses of individuals in the numeracy groups (see Section 6.1.3).

The analysis of the risky choices answered the second research question, i.e. *What risk attitudes and probability distortions are associated with the risky choices of decision makers of different numeracy groups?*

7.1.2. The claim

The original proposition of the thesis was that individuals' statistical numeracy is a meaningful source of heterogeneity in the modelling of risky choices (Section 3.8). This proposition is now stated as a claim as it has been tested by the thesis.

7.1.3. The reasons and logic of the claim

The major reason for making the claim is based on the overall finding of differences in the risky choices of respondents in the statistical numeracy groups (see Section 5.3). The second reason for the claim, and the main reason for the original proposition of the thesis, is that the construct of inductive reasoning underlies both risky decision making and probabilistic reasoning (see Section 3.3). Thus, the constructs are expected to be related.

7.1.4. Evidence for the claim

The main evidence for the claim of the thesis is the consistent finding in the empirical results of the study that the risk aversion levels of the respondents were highest for the low

numeracy respondents and lowest for the high numeracy respondents. This result was found in the preliminary, model-free analyses and in the estimated utility parameters of all the models considered.

The results of the preliminary, model-free analyses indicated a propensity for respondents to choose the safer lottery within the choice pair, indicative of aversion to risk. There was, however, a pattern in the gain-framed choices of all the numeracy groups for riskier options to be selected more frequently as the attractiveness of the risky option increased. In the loss frame, respondents were again averse to risk, as they selected the safer option more frequently as the expected value of the loss increased. All patterns in the choices were stronger for the high numeracy group, and weakest for the low numeracy group (see Sections 5.3.1, 5.3.2 and 6.2.1). Although these patterns were weakest in the low numeracy group, there was, nonetheless, systematic variability in their responses. This is an important finding, as noisy risky choice responses are typical of individuals of lower cognitive ability in the empirical literature on risky choice (see Section 3.5.2).

The results of the maximum likelihood estimates of the utility parameters of the models consistently indicated the pattern of high risk aversion in all the numeracy groups, but highest for the low numeracy respondents. Similarly, the Bayesian estimates confirmed the pattern of decreasing risk aversion with increasing numeracy. The Bayesian parameter estimates of the of the expected utility model, and of the rank-dependent model in the gain frame, indicated satisfactory predictive validity, with medium effect size relative to chance prediction. The loss-framed choices, however, did not reflect the expected theoretical pattern of risk seeking (see Section 5.3).

7.1.5. Qualifiers

There are several qualifiers, or restrictions, to the main claim of the thesis. The main qualifier concerns the theoretical *decision models and functional forms* specified by the thesis. The approach to the modelling of risky choice was purely parametric, using three decision models. These models depend on parametric functional forms for transforming outcome values to subjective values, and objective probabilities to subjective weights (Cavagnaro et al., 2013). Thus, the results of the thesis were limited to the three decision models combined with the selected functional forms: expected utility with power utility; rank-dependence with power utility and Tversky and Kahneman's single parameter probability function; rank-dependence with linear utility and Lattimore et al. probability function; and cumulative

prospect theory with power utility and Tversky and Kahneman's single parameter probability function. Although the functional forms of the thesis used for the modelling are among the most commonly used in the risky choice literature, the models have no power if the functions are inappropriate for the observed choices. This situation arose for loss-framed lotteries. The Tversky and Kahneman single parameter probability function did not describe the observed risky choices in the loss frame.

The second qualifier concerns the *objectives* of the thesis. The thesis was designed to compare the performance of each model across the three numeracy groups. There was no intention to compare the performance of the different models to one another. This restriction avoided the potential complication of developing a measurement instrument for risk that could test the three models with equal power. The construction of such an instrument was not feasible within the constraints of its space and response time.

The third qualifier concerns the *generalisability* of the results of the thesis, focusing on the population accessed (see Section 4.3). Students from the University of the Witwatersrand were recruited to participate in the online experiment on risky choice, with the possibility of winning a monetary prize. This nonprobability sampling method is typical of risky choice research, as the sample is required to provide respondents with variability on a key variable, such as numeracy in the thesis, rather than represent the underlying population. In the case of the thesis, the response was overwhelming, indicating that the students were extremely eager to receive a monetary reward in response for their participation. Their reaction is understandable given that many of them face financial challenges. This factor may explain the generally high level of risk aversion as well. In general, the students preferred safer bets to riskier bets. Thus, the qualification is made that the results of the research may have been different had they been based on respondents of a higher socio-economic status.

The final qualification concerns the *non-causal nature* of the relation between statistical numeracy and risky choice. No claims of causation are made as a myriad of extraneous factors could account for differences in risk attitudes.

7.1.6. Reservations and reflections

The reflections of the researcher are presented in three sections: analytical modelling challenges, model complexity, and testing and modifying economic models (Sections 7.1.6.1 - 7.1.6.3, respectively).

7.1.6.1. Analytical modelling challenges

At the start of the thesis, the researcher identified two main problematic areas in research on risky choice. These areas were a lack of consideration of stochastic errors in the modelling of risky choice, and methodological design flaws in empirical studies. The researcher designed the empirical aspect of the thesis with the intention of avoiding both stochastic error and design flaws. This section is focused on the analytical challenges of accommodating stochastic error.

Stochastic response error was assumed to be addressed by incorporating the Wilcox contextual error model into the parameter estimates (see Section 2.3.3). Without a parameter recovery validation exercise, it is unknown to what extent the error model accounted for the stochastic response error in the data. Stochastic preference error in the parameters was assumed to be accommodated by conducting Bayesian analysis, which treats parameters as distribution-based random variables with random variation in parameter estimates (see Section 4.8.4). The intention was to use maximum likelihood estimates computed on a subset of the data for the prior probability distributions of the Bayesian analyses on the remainder of the data. In practice, maximum likelihood estimation of model parameters was largely unsuccessful when attempts were made to correct the contextual error. However, contextual error was incorporated into the Bayesian analyses.

The hierarchical Bayesian estimation was challenging to implement. The researcher followed the example of the hierarchical Bayesian analysis of Nilsson et al. (2011a), and adapted the accompanying WinBUGS software code. Long MCMC simulations (see Section 5.3.4) were required to reach convergent solutions, and computational traps arose frequently in the sampling iterations.

The hierarchical Bayesian method worked well for estimating the simpler expected utility model, and the rank-dependent model with the single parameter probability weighting function. For these models, the success rates of Bayesian estimates and predictions were consistently above chance relative to in-sample and out-of-sample actual choices. These successful estimations derived from (approximately) convergent solutions, and were based on informative prior distributions (Lee & Vanpaemel, 2018), sourced empirically (Gershman, 2016). However, problems of model non-identifiability arose when attempting to estimate the more complex models with several parameters. The cumulative prospect model for the

middle and low numeracy groups, and the rank-dependent model with the Lattimore probability transformation, were non-identifiable, given the data of the thesis.

The Bayesian modelling clearly demonstrated that if the information content of the modelling data is low relative to the complexity of the decision model that is being estimated, even exhaustive amounts of MCMC modelling may not approximate the parameters of the decision model (Spektor & Kellen, 2018). Good quality data is a necessary condition for an identifiable model, regardless of the sophistication of the statistical techniques and modelling resources used.

It is also very important not to over-interpret the individual parameter estimates separately from one another, or even their individual magnitudes at all, if they are strongly intercorrelated (see Section 5.3.4.5). As intercorrelated parameters may compensate partly or more fully for one another, the researcher had to be careful not to draw conclusions on the individual probability weighting parameters, as they are generally highly, or moderately, correlated with the utility parameter. It may be incorrect even to draw conclusions on risk aversion. *The reason why risk aversion was singled out in the conclusions of the thesis, is the consistent finding throughout the preliminary model-free and the later model-dependent analyses for both single-parameter and multi-parameter functions, that risk aversion is high, and highest for the low numeracy group*. This finding is in accord with literature on risky choice and cognitive ability.

7.1.6.2. Model complexity

Since the days of Herbert Simon (1955), and through the literature of behavioural economics, the descriptive inadequacies of the neoclassical model have been challenged (see Section 2.1.5). The result is a rich array of non-expected utility models designed to be descriptively accurate. In place of the simplicity of expected utility theory, these non-expected utility models are more complex with more parameters. The model parameters are quantified, and assumed to reflect underlying psychological processes involved in risky decision making (Scheibehenne & Pachur, 2015). These processes may be *as if* processes that have explanatory value for risky choice, but are not necessarily accurate in an ontological or metaphysical sense. This issue is not a concern for the researcher as the thesis is concerned with testing the model predictions based on a variance account rather than a process account. Nevertheless, the parameters of each of the three models tested in the thesis are theoretically founded and constructed to represent a psychological process.

The objective of the thesis was to model the risky choices of individuals in numeracy groups using three core decision models in the risky choice literature. The thesis has shown that decision models that are moderately successful in their predictions differ in their parameters based on the statistical numeracy of the decision maker. Statistical numeracy explains significant heterogeneity in the behavioural choices of individuals under risk. The finding is in line with the objectives of behavioural economists who seek descriptive models of risky choice that explain variability in the observed choices of individuals under risk (Brañas-Garza & Smith, 2016).

There are, of course, many variables other than statistical numeracy that may also explain heterogeneity in risky choices. Yet every variable that is added to a predictive model also adds to its complexity, its potential non-identifiability, and the potential of misinterpreting parameter estimates that are highly intercorrelated. Statistical numeracy adds to the complexity of descriptive models of risk choice, but this knowledge can nevertheless lessen the inaccuracies of model predictions if used with care. For instance, the thesis has shown that under the same risk model, different parameter values can be used to model the choices of individuals, dependent on their numeracy levels. The findings of the thesis further the general finding in the behavioural economics literature of a link between numeracy and risky choice.

There is one further question that still needs answering. That is, whether the proposal to modify models of risky choice, as indicated by the results of this thesis, satisfies the criteria of Friedman (1953) and Rabin (2013).

7.1.6.3. Testing and modifying economic models

Economists advocate the testing of economic models. For example, Friedman (1953/2007) declared that the importance of testing theories is: "…to specify the circumstances under which the formula works or, more precisely, the general magnitude of the error in its predictions under various circumstances" (p. 155). More recently, Rabin (2013a) acknowledged the importance of modifications to economic models. He stipulates, however, that these modifications be (i) "broadly applicable", (ii) "predictively sharp", and (iii) "mathematically precise" (p. 617). On this view, (iv) modifications must incorporate mathematical theory and measurable variables to provide (v) useful new input for economic theory. Resultant models should have explanatory power that is (vi) "…general in their applicability across contexts but *specific* in their predictions within each context" (p. 618).

The researcher believes that this thesis offers a modification to descriptive models of risky choice in the gain frame. For instance, considering the rank-dependent model in the gain frame, using the same formula but with different parameter values dependent on numeracy, satisfies the various requirements cited. The modification the researcher proposes is incorporating statistical numeracy as a context, or circumstance, that moderates the predictions of the expected utility model, and the rank-dependent model in the gain frame, for three broad groups of individuals. Rabin's criteria are examined in turn, in the context of the proposed modification.

First, statistical numeracy in modelling risky choice could be applied to many situations involving risky choice, for example, in a medical situation or a financial situation, and so it is broadly applicable. Second, the formulae proposed for predicting risky choices in high, middle and low numeracy conditions are, to some extent, predictively sharp, as they predict correctly at least two out of three choices, i.e. a prediction of medium effect size (Cohen, 1992). Third, the modelling provides parameter estimates with 95% credible intervals and standard deviations, which quantify uncertainty around the predictions of risky choices at each numeracy level, and so is mathematically precise. Fourth, the modification incorporates mathematical theory (hierarchical Bayesian analysis), underlying decision theory and stochastic error theory, dependent on numeracy levels for its predictions. It also provides measurement, as statistical numeracy is quantified using a psychometrically sound measurement scale across a wide range of numerical ability levels. Fifth, the modification provides useful new input for economic theory, as the aversion to risk in the choices of individuals is seen as inversely related to their numeracy levels (general in their applicability across contexts). Comparing the model predictions for high versus low numeracy groups, for example, reveals the gap between individuals' evaluations of risk, dependent on their numeracy levels. As these comparisons are quantifiable through the model predictions, they could provide input to the analysis of costs and benefits of policy implementations, for example of a nudge policy, or of a boost policy. Sixth, the modification is potentially applicable across all the risky choice models considered, and therefore satisfies Rabin's requirement for broad applicability. The formulae for the models span the range of numerical abilities (across contexts), and provide specific predictions within them. Thus, the proposed modification satisfies Rabin's requirements for a modification to a descriptive model of risky choice.

The thesis would also be considered useful on Friedman's condition for testing a theory, as it specifies the statistical numeracy circumstances, or context, under which the risky choice formulae work best, and quantifies the magnitude of the error in the predictions.

7.2. Contributions of the thesis

The discussions of the previous sections are consolidated into four contributions that the thesis makes to the existing literature. The thesis: (i) provides an improved measurement instrument of statistical numeracy with an application to establishing three clearly defined numeracy groups; (ii) evaluates three prominent models of risky choice, each compared under three numeracy conditions, and establishes a link between statistical numeracy and risk aversion in the model parameters; (iii) modifies some of the existing risky decision models; and (iv) furthers the unification of two disparate fields of research.

(i) The thesis has produced a measurement instrument of statistical numeracy that amalgamates frequently used instruments of statistical numeracy and cognitive reflection. Based on the empirical data of the thesis, the instrument is psychometrically sound and discriminates across a broad range of statistical numeracy. It provides a clear factor structure that is invariant to known covariates in the numeracy literature. Based on these factors, the thesis finds three latent numeracy groups: the numerically rational group (the high numeracy group); the numerically operational group (the middle numeracy group); and the non-numerical group (the low numeracy group). The presence of three numeracy groups provides the literatures on statistical numeracy and decision making with a richer categorisation than a median split or other convenient numeracy split. Although the numeracy instrument and the categorisation of the thesis are yet to be tested in future research, they provide the potential for policy evaluation or theory development, based on statistical numeracy (Cokely et al., 2014).

(ii) The thesis has modelled and compared the risky choices of three numeracy groups on three core decision models – expected utility, rank-dependent and cumulative prospect theory. It has used both frequentist and Bayesian analysis approaches to estimate the model parameters of the separate numeracy groups, and incorporated stochastic error in the analyses. It has examined the Bayesian modelled results for convergence, for parameter intercorrelations and for model identifiability, and assessed the models for their in-sample estimates and out-of-sample predictions. In so doing, the thesis has answered the empirically-based research question on the role of statistical numeracy in modelling risky choice. To the

knowledge of the researcher, this is the most thorough parametric modelling study of risky choice conducted on well-defined numeracy groups.

(iii) The thesis has proposed a modification of two of the risky choice models, by claiming that the models need to be applied taking cognisance of the statistical numeracy of decision makers. The modification applies to the expected utility theory model, and to the rank-dependent model with the Tversky and Kahneman probability weighting function, both in the gain frame. To limit adding complexity to the models with additional parameters, the thesis provides a separate set of parameter estimates for each model and numeracy group.

(iv) Finally, the thesis has contributed to a limited number of studies (e.g. Lopes, 1995; Millroth et al., 2019; Pachur et al., 2017c) that have attempted to unify two largely independent research areas: the one being the decision theory modelling literature in economics, and in behavioural economics that considers the individual's limited cognitive ability; the other being the policy literature based on the judgment field concerned with cognitive biases and simple heuristics. The thesis provides empirical findings that can be used by policy makers in either nudging decision makers to make better choices, or in boosting their decision-making skills.

7.3. Strengths and limitations of the research

The strengths of the research lie in its scientific approach, its creation of a psychometrically sound measure of statistical numeracy and clear numeracy groups, its attempts to control stochastic errors in the modelling of risky choice, and its attempts at rigorous analysis. Other worthy aspects were discussed as contributions of the thesis in the previous section (7.2).

The following limitations of the research are acknowledged:

The first limitation concerns the parametric approach used for the modelling. The approach requires the specification of a utility function, and of a probability weighting function in the case of multi-parameter models. However, the observed choices in the loss frame and the predictions of the rank-dependent and cumulative prospect models were mismatched, which was particularly serious for the thesis. The mismatch typified the inherent vulnerability of parametric modelling approaches to misspecification of utility or probability weighting functions. As outlined by Booij et al. (2010), a misspecified probability weighting function will bias the estimated utility function, and *vice versa*. Booij et al. provide examples of parametric studies that may have fallen prey to the same unfortunate phenomenon (Donkers, Melenberg, & van Soest, 2001, as cited in Booij et al., 2010; Harrison & Rutström, 2009). As

in the thesis, these studies found high utility curvature and low loss aversion, which may indicate misspecification of the probability weighting functions. It is likely that this effect occurred in the modelling of the thesis, as the cumulative prospect model was nonidentifiable, and discarded (see Section 6.2.3.4). It is also likely to have obscured the clear identification of the probability weighting parameters in the modelling. There are modelling approaches other than parametric that invite different challenges. Non-parametric modelling is earmarked as an approach to be considered in future research (see Section 7.4).

The second limitation concerns the measurement of risky choice. Contrary to the measurement of numeracy which derives from the strong psychometric principles of psychology, such as reliability, factor structure and measurement invariance, the investigation of these principles is largely absent in measures of risky choice in economics and behavioural economics (Beauchamp, Cesarini, & Johannesson, 2017). The risky choice measure of the thesis was restricted in the quality of the data it elicited, due partly to practical restrictions of time and content.

The third limitation may be the extent to which the contextual error model selected to control stochastic response error achieved this purpose. An alternative error model may have been preferable. A future parameter recovery exercise would inform this issue.

A fourth limitation is that the researcher did not investigate the models for potential evidence of problematic structural parameter interdependencies (Scheibehenne & Pachur, 2015; see Section 6.2.2.4). The investigation would have involved plotting the posterior distributions of parameter pairs against one other, either at the individual or group level, but was considered beyond the scope of the thesis. The group-level posterior distributions of the hierarchical estimations could also have been used as a basis for comparing the numeracy groups (Section 6.2.3.1).

The final limitation concerns the control of covariates in the study. As repeatedly stated throughout the thesis, the research was not designed to examine a causal relation between numeracy and risky choice in computational decision models. Nevertheless, the numeracy groups differed on demographic variables, such as gender and income level. These variables would have introduced additional parameters for estimation and were not controlled in the modelling.

7.4. Further research

The thesis highlights several potential areas of research.

First, a quantitative meta-analysis summarising the previous analyses that examined an association between risky choice and numeracy could highlight whether the methodological limitations of previous research cited in the thesis are in fact practically substantive and make a difference for the conclusions obtained (T. Pachur, personal communication, November 24, 2020).

Second, follow-up and extensions to the data of the thesis may provide valuable insight on the role of numeracy in risky choice. For the numeracy part of the thesis, validation exercises need to be carried out on the factor structure of the numeracy scale and the latent numeracy groups of the research, using independent data. It would be valuable to link the numeracy factors of the thesis to other measurement scales, for example, the rational thinking scale of Stanovich et al. (2016) and the subjective numeracy scale of (see Section 3.2.2).

Third, for the risky choice modelling part, parameter recovery exercises should be conducted on the two most successful models and their functional forms in the gain frame, i.e. on the expected utility model, and on the rank-dependent model with power utility and the Tversky and Kahneman single parameter probability weighting function. Furthermore, a different weighting function should be considered as the weighting function used did not describe risky choices in the loss frame. An additional stochastic error model may need to be used, with the Wilcox contextual error, to account for stochastic response error. Additionally, alternative sources of informative prior probabilities, based on independent data, need to be obtained for future hierarchical Bayesian analyses. Control of demographic covariates may be considered in future modelling.

Fourth, a richer set of risky choice data is necessary for further modelling of the role of statistical numeracy. A battery of prospect pairs, framed in the gain, loss and mixed frames, and comprising at least three outcomes, is necessary to test the models, including cumulative prospect theory. An example of such a battery is provided by Harrison and Swarthout (2016). These researchers designed their battery of 100 pairs for “stress-testing” (Harrison & Swarthout, 2016, p. 7) five core decision models, including those used in the thesis. Such data would enable comparisons of models, and the creation of mixtures of models, for application to numeracy groups (Conte, Hey, & Moffatt, 2011; Harrison & Rutström, 2009).

There are, however, logistical limitations that need to be accommodated to use such a battery, or equivalent, aside from author permissions. In the thesis, the questionnaires could accommodate only about 48 prospect pairs in a single administration. This was the feasible limit, given the rest of the content of the questionnaires – descriptions of ethical considerations, instructions, explanations, as well as the other sections on statistical numeracy and demographic items (see Section 4.5). Generally, empirical studies that use long batteries, or repeated measures experiments, test individuals over multiple sessions. Securing a rich set of data online would require multiple online administrations too.

Thus, a fifth aspect of future research on the modelling topic is innovative online experiments. This methodology offers a promising avenue for designing online research with multiple sessions and large samples. Results from a large replication laboratory study show that online interactive experiments yield adequate and reliable data (Arechar, Gächter, & Molleman, 2018). The method is convenient and efficient for selecting targeted respondents, instead of relying on community and student samples. It could be used in the context of risky choice experiments to test the validity of previous laboratory experiment results, to test the stability of choices over time of respondents of different statistical numeracy levels, and to test the effects of online numeracy education programmes over time. There are other platforms that offer online labour markets, including Amazon Mechanical Turk online (Gureckis et al., 2016).

Online experimental platforms also offer the potential of developing elicitation instruments that cover additional constructs (Dohmen et al., 2018). For instance, measures of subjective numeracy could be included , as could novel measures of inductive statistical numeracy that require responses to scenarios with no specific correct response. A broader measurement perspective on numeracy may lead to a more comprehensive factor structure of numeracy and possible subclusters (Fagerlin, et al., 2007). Furthermore, while minding the necessity of avoiding complex models, rich sets of risky choices, together with psychometrically sound measures of other constructs, for example, affect, offer interesting prospects for modelling research. Additionally, the modelling of risky choice could be expanded to include cognitive decision algorithms using eye tracking technology (Fiedler & Glöckner, 2012). Although each of these initiatives has been carried out previously, a careful experimental design whereby groups of individuals are measured on theoretically-based variables, including their risky choices, may offer a method of modelling risky choice by identifying key correlates, in addition to, or in place of numeracy. Thus, a sixth suggestion for future modelling research is

testing large online samples on multiple constructs, using sophisticated modelling methods, to arrive at simple, powerful models. A non-parametric modelling approach to the risk elicitation and modelling could also be considered instead of a parametric approach.

The last suggestion is to conduct these future research endeavours in the spirit of open science collaboration (Nosek et al., 2020). Data collected could be made available to researchers who collaborate to innovate, check and repeat each other's research. The result could be advanced Bayesian decision making modelling and hypothesis testing, with the increased potential to inform policy making.

7.5. Epilogue

I end my thesis in a world that is far different from when I started reading about risk. More accurately, it is far different from two weeks ago, even from a week ago, or since yesterday, and from tomorrow. Global economic, social and psychological norms and standards are changing dramatically in response to the ever increasing COVID-19 crisis. The threats of the virus are constantly in conversations of all types. Terms like 'risk', 'probability' and 'flattening the curve' are commonplace, bandied about in everyday social interactions. They never were, at least to this extent, as far as I can remember. The social media is now filled with chats in which people passionately debate or refute statistics like the risks of mortality from the WHO. Numerical and logical misinterpretations are common, as are misinterpretations of cognitive biases.

By contrast, there are many responsible and informative academic articles shared across social media (e.g. Battegay et al., 2020; Baud et al., 2020), and helpful websites displaying graphics with explanations and simulations of flattening the curve of infections (Stevens, 2020). However, to what extent individuals with low numeracy skills are receptive to online, statistically-based information, is not known. There are calls for global leaders to arrange for scientists to address the hysteria, by educating citizens on assessing risk assessment and structuring responsible behaviour in a disaster scenario (Whiteside, 2020). Interpreting probabilities and risk are central to these endeavours.

Finally, there is an interesting correlation between the extent of countries' success rates in containing the pandemic thus far, and the average mathematical literacy levels of their citizens. While the relationship is correlational, not causal, it demonstrates that countries with the best containment records also score highly in mathematics, as measured by their students' mathematics proficiency, based on PISA scores produced by the OECD. Specifically, China,

Singapore, Japan and South Korea are ranked in the top seven of the participating countries on the mathematics proficiency of their students, while Italy, Spain and the United States are ranked between 31 and 37, out of 77 countries (Koujianou Goldberg, 2020; Schleicher, 2019).

Statistical numeracy matters in the risky choices that people make, and can contribute to the modelling of risky choice.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science, 46*(11), 1497-1512. doi:https://doi.org/10.1287/mnsc.46.11.1497.12080

Adler, J., & Pillay, V. (2017). Mathematics education in South Africa. In J. Adler, & A. Sfard (Eds.), *Research for Educational Change: Transforming researchers' insights into improvement in mathematics teaching and learning.* NY, New York: Routledge.

Alipour, P. (2019). Optimal income taxation with correlated multi-dimensional preference heterogeneity (Working Paper). Retrieved from https://www.researchgate.net/publication/337447410

Amador Hidalgo, L., Brañas-Garza, P., Espín, A. M., Garcia, T., & Hernández Román, A. (2019). Consistent and inconsistent choices under uncertainty: The role of cognitive abilities. *MPRA Paper No. 95178*. Retrieved from https://mpra.ub.uni-muenchen.de/95178/

Andersen, S., Cox, J. C., Harrison, G. W., Lau, M. I., Rutström, E. E., & Sadiraj, V. (2018). Asset integration and attitudes toward risk: Theory and evidence. *Review of Economics and Statistics*, 816-830. doi:10.1162/rest_a_00719

Andersen, S., Harrison, G. W., & Rutström, E. E. (2006). Choice behavior, asset integration and natural reference points. (Vol. 4). Working Paper 06.

Anderson, B. L., & Schulkin, J. (2014). Introduction. In B. L. Anderson, & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 1-10). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139644358

Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2016). Risk aversion relates to cognitive ability: Preferences or noise? *Journal of the European Economic Association, 14*(5), 1129-1154. doi:10.1111/jeea.12179

Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2018). Robust inference in risk elicitation tasks. Discussion Papers 18-09, University of Copenhagen. Department of Economics. Retrieved from https://ideas.repec.org/p/kud/kuiedp/1809.html

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics, 21*, 99–131. doi:10.1007/s10683-017-9527-2

Ashby, N. J. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making, 12*(2), 128–139.

Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives, 19*(3), 131–145.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438. doi:10.1080/10705510903008204

Aunio, P., Mononen, R., Ragpot, L., & Törmänen, M. (2016). Early numeracy performance of South African school beginners. *South African Journal of Childhood Education, 6*(1), 8 pages. Retrieved from https://doi.org/10.4102/sajce.v6i1.496

Azrieli, Y., Chambers, C. P., & Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy, 126*(4), 1472-1503. doi:10.1086/698136

Baker, M. (2016). Is there a reproducibility crisis? *Nature, 533,* 452-454.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2018). Estimating risk preferences in the field. *Journal of Economic Literature, 56*(2), 501–564. Retrieved from https://doi.org/10.1257/jel.20161148

Basel, J. S., & Brühl, R. (2013). Rationality and dual process models of reasoning in managerial cognition and decision making. *European Management Journal, 31*(6), 745–754. doi:10.1016/j.emj.2013.07.004

Battegay, M., Kuehl, R., Tschudin-Sutter, S., Hirsch, H. H., Widmer, A. F., & Neher, R. A. (2020). 2019-Novel Coronavirus (2019-nCoV): estimating the case fatality rate – a word of caution. *Swiss Medical Weekly, 150: w20203*. Retrieved from https://doi.org/10.4414/smw.2020.20203

Baud, D., Qi, X., N-S, K., Musso, D., Pomar, L., & Favre, G. (2020). Real estimates of mortality following COVID-19 infection. Retrieved from https://doi.org/10.1016/S1473-3099(20)30195-X (preprint)

Beauchamp, J. P., Cesarini, D., & Johannesson, M. (2017). The psychometric and empirical properties of measures of risk preferences. *Journal of Risk and Uncertainty, 54*, 203–237. doi:10.1007/s11166-017-9261-3

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science, 8*(1), 41-55. doi:https://doi.org/10.1002/bs.3830080106

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is 'behavioral'? Cognitive ability and anomalous preferences. *Journal of the European Economic Association, 11*(16), 1231–1255. doi:10.1111/jeea.12055

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk (Trans. L. Sommer). *Econometrica, 22*(1), 23-36. (Original work published 1738). doi:10.1.1.436.7181

Bhargava, S., & Loewenstein, G. (2015). Behavioral economics and public policy 102: Beyond nudging. *American Economic Review: Papers & Proceedings, 105*(5), 396–401. Retrieved from http://dx.doi.org/10.1257/aer.p20151049

Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review, 124*(5), 678–687. doi:http://dx.doi.org/10.1037/rev0000073

Bichi, A. A., & Talib, R. (2018). Item response theory: An introduction to latent trait models to test and item development. *International Journal of Evaluation and Research in Education, 7*(2), 142-151. doi:10.11591/ ijere.v7.i2.pp142-151

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*(2), 463–501. doi:10.1037/0033-295X.115.2.463

Birnbaum, M. H. (2011). Testing theories of risky decision making via critical tests. *Frontiers in Psychology, 315*(2), 1-3. doi:10.3389/fpsyg.2011.00315

Birnbaum, M. H., Coffey, G., Mellers, B. A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology, 18*(2), 331-346.

Blavatskyy, P. R. (2007). Stochastic expected utility theory. *Journal of Risk and Uncertainty, 34*, 259–286. doi:10.1007/s11166-007-9009-6

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25*(6), 963-986. doi:10.1002/jae.1116

Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E.-J. (2018). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior Research Methods, 50*, 1614–1631. Retrieved from https://doi.org/10.3758/s13428-018-1054-3

Booij, A. S., & Van de Kuilen, G. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic psychology, 30*(4), 651-666. doi:https://doi.org/10.1016/j.joep.2009.05.004

Booij, A. S., van Praag, B. M., & van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision, 68*, 115-148. doi:10.1007/s11238-009-9144-4

Booysen, S. (Ed.). (2016). *Fees must fall: Student revolt, decolonisation and governance in South Africa.* Johannesburg, South Africa: Wits University Press.

Brañas-Garza, P., & Smith, J. (2016). Cognitive abilities and economic behavior. *Journal of Behavioral and Experimental Economics, 64(C)*, 1-4. doi:10.1016/j.socec.2016.06.005

Brand, M., Labudda, K., & Markowitsch, H. J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. *Neural Networks, 19*(8), 1266-1276. doi:10.1016/j.neunet.2006.03.001

Brand, M., Schiebener, J., Pertl, M.-T., & Delazer, M. (2014). Know the risk, take the win: How executive functions and probability processing influence advantageous decision making under risk conditions. *Journal of Clinical and Experimental Neuropsychology, 36*(9), 914–929. Retrieved from http://dx.doi.org/10.1080/13803395.2014.955783

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113*(2), 409–432. doi:10.1037/0033-295X.113.2.409

Bridge. (2020). Retrieved from Bridge.org: https://www.bridge.org.za/tackling-the-challenge-of-improving-south-african-learners-maths-results/

Bristow, W. (2017). Enlightenment. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). Retrieved from https://plato.stanford.edu/archives/fall2017/entries/enlightenment/

Brooks, M. E., & Pui, S. Y. (2010). Are individual differences in numeracy unique from general mental ability? A closer look at a common measure of numeracy. *Individual Differences Research, 8*(4), 257-265.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*(4), 434-455.

Brooks, S., & Gelman, A. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472.

Brown, K. S., & Sethna, J. P. (2003). Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E, 68*, 021904.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13*(2), 149-154. doi:10.1177/1745691617706516

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1). doi:https://doi.org/10.1177/1745691610393980

Burke, C. J., & Tobler, P. N. (2011). Coding of reward probability and risk by single neurons in animals. *Frontiers in Neuroscience, 5*(121), 1-9. doi:10.3389/fnins.2011.00121

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences of the United States of America, 106*(19), 7745–7750. doi:10.1073pnas.0812360106

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review, 100*(3), 432-459. doi:10.1037/0033-295X.100.3.432

Caldwell, B. J. (1980). A critique of Friedman's methodological instrumentalism. *Southern Economic Journal, 47*(2), 366-374. doi:10.2307/1057529

Camerer, C. F. (2000). Prospect theory in the wild: Evidence from the field. In D. Kahneman, & A. Tversky (Eds.), *Choices, values, and frames* (pp. 288-300). Cambridge, UK: Cambridge University Press.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7-72.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory and Cognition, 42*(3), 434–447. doi:10.3758/s13421-013-0367-9

Campitelli, G., & Gobet, F. (2010). Herbert Simon's decision-making approach: Investigation of cognitive processes in experts. *Review of General Psychology, 14*(4), 354-364. doi:http://dx.doi.org/10.1037/a0021256

Carr, P. G. (2016). NAEP 2015 twelfth-grade mathematics and reading results. 2016 April 27. In: National Center for Education Statistics, IES. Washington. Retrieved from https://nces.ed.gov/WhatsNew/commissioner/remarks2016/04_27_2016.asp

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., & Myung, J. I. (2013). Discriminating among probability weighting functions using adaptive design optimization. *Journal of Risk and Uncertainty, 47*(3), 255-289. doi:10.1007/s11166-013-9179-3

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization, 87,* 43– 51. doi:10.1016/j.jebo.2012.12.023

Chilisa, B., & Kawulich, B. B. (2012). Selecting a research approach: Paradigm, methodology and methods. In C. Wagner, B. B. Kawulich, & M. Garner (Eds.), *Doing social research: A global context* (pp. 51-61). London, UK: McGraw-Hill Higher Education.

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 681-703. doi:10.1080/10705511.2013.824786

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. Retrieved from https://doi.org/10.1037/0033-2909.112.1.155

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4*(1), 20-33.

Cokely, E. T., Feltz, A., Ghazal, S., & Allan, J. N. (2016). Decision making skill: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *Cambridge handbook of expertise and expert performance* (2nd ed., pp. 476-505). Cambridge, UK: Cambridge University Press.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*(1), 25–47.

Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Measuring numeracy. In B. L. Anderson, & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 11-38). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139644358.002

Conte, A., Hey, J. D., & Moffatt, P. G. (2011). Mixture models of choice under risk. *Journal of Econometrics, 162*, 79–88. doi:10.1016/j.jeconom.2009.10.011

Cowles, M. K., & Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*(434), 883-904. Retrieved from http://www.jstor.org/stable/2291683

Cox, J. C., Sadiraj, V., & Schmidt, U. (2015). Paradoxes and mechanisms for choice under risk. *Experimental Economics, 18*, 215–250. doi:10.1007/s10683-014-9398-8

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics:The role of executive function in thedevelopment of mathematics proficiency. *Trends in Neuroscience and Education, 3*, 63-68. Retrieved from http://dx.doi.org/10.1016/j.tine.2013.12.001

Dave, C., Eckel, C. C., Johnson, C. A., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty, 41*, 219–243. doi:10.1007/s11166-010-9103-z

Davidson, D., Siegel, S., & Suppes, P. (1955). *Some experiments and related theory on the measurement of utility and subjective probability.* Office of Naval Research, Technical Report No. 1. Stanford University, CA.

Dhami, S. (2016). *The foundations of behavioral economic analysis.* Oxford, UK: Oxford University Press.

Di Girolamo, A., Harrison, G. W., Lau, M. I., & Swarthout, T. (2015). Subjective belief distributions and the characterization of economic literacy. *Journal of Behavioral and Experimental Economics, 59*(1), 1-12. Retrieved from http://dx.doi.org/10.1016/j.socec.2015.08.004

Diecidue, E., & Wakker, P. P. (2001). On the intuition of rank-dependent utility. *The Journal of Risk and Uncertainty,* 281–298.

Dieckmann, N. F., Peters, E., Leon, J., Benavides, M., D, P., & Norris, A. (2015). The role of objective numeracy and fluid intelligence in sex-related protective behaviors. *Current HIV Research, 13*(5), 337-346.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238-1260. doi:10.1257/aer.100.3.1238

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2018). On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives, 32*(2), 115-134. doi:10.1257/jep.32.2.115

Edwards, W. (1953). Probability-preferences in gambling. *The American Journal of Psychology, 66*(3), 349-364. doi:10.2307/1418231

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*(4), 380-417.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*(3), 193-242. doi:10.1037/h0044139

Elstein, A. S. (2004). On the origins and development of evidence-based medicine and medical decision making. *Inflammation Research, 53*(2), S184-S189.

Emergencies. (2020, March 6). Retrieved from World Health Organization: https://www.who.int

English Oxford Living Dictionaries. (n.d.). Retrieved 1 18, 2018, from https://en.oxforddictionaries.com/definition/parsimony

Ert, E., & Erev, I. (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. *Judgment and Decision Making, 8*(3), 214–235.

Fagerlin, A., Zikmund-Fisher, B., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. *Medical Decision Making, 27*(5), 672-680. doi:10.1177/0272989X07304449

Fiedler, K., & von Sydow, M. (2015). Heuristics and biases: beyond Tversky and Kahneman's (1974) judgment under uncertainty. In M. W. Eysenck, & D. Groome, *Cognitive psychology: revisiting the classical studies* (pp. 146-161). Thousand Oaks, CA: Sage Publications.

Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology, 3*(335), 1-18. doi:10.3389/fpsyg.2012.00335

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage.

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Encyclopedia of special education* (pp. 368-382). West Sussex, UK: John Wiley & Sons, Inc. doi:10.1002/9781118660584.ese0431

Fox, C. R., & Poldrack, R. A. (2009). Prospect theory and the brain. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics – Decision making and the brain* (pp. 145-173). London, UK: Elsevier Academic Press.

Fox, C. R., Erner, C., & Walters, D. J. (2015). Decision under risk: From the field to the laboratory and back. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 43-88). West Sussex, UK: John Wiley & Sons, Ltd.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.

Friedman, M. (1953/2007). The methodology of positive economics. In D. M. Hausman, *The philosophy of economics: An anthology* (3rd ed., pp. 145-178). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511819025.010

Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology, 28*(3), 497–516. doi:10.1002/jcpy.1047

Galesic, M., & Garcia-Retamero, R. (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine, 170*(5), 462-468. Retrieved from http://jamanetwork.com/

Garcia-Retamero, R., & Cokely, E. T. (2014). Using visual aids to help people with low numeracy make better decisions. In B. L. Anderson, & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 153-174). Cambridge, UK: Cambridge University Press. Retrieved June 6, 2017

Garcia-Retamero, R., Sobkow, A., Petrova, D., Garrido, D., & Traczyk, J. (2019). Numeracy and risk literacy: What have we learned so far? *The Spanish Journal of Psychology, 22*(e10). doi:10.1017/sjp.2019.16

Geiser, C. (2013). *Data analysis with Mplus.* New York, NY: The Guilford Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, Fla: Chapman and Hall/CRC.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences, 110*(28), 11267–11271. doi:10.1073/pnas.1220568110

Gerber, A. S., & Green, D. P. (2011). Field Experiments and Natural Experiments. In *The Oxford Handbook of Political Science.* Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199604456.013.0050

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology, 71*, 1-6. Retrieved from http://dx.doi.org/10.1016/j.jmp.2016.01.006

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science, 7*(4), 473-511. Retrieved from https://www.jstor.org/stable/2246094

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences (Psychology series).* San Fransisco: W. H. Freeman and Company.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2*, 83–115.

Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer, & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 37-50). Cambridge, MA: The MIT Press.

Gigerenzer, G. (2015). *Simply rational: Decision making in the real world.* Oxford, UK: Oxford University Press.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107-143. doi:10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal, 327*, 741-744. doi:https://doi.org/10.1136/bmj.327.7417.741

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53-96. doi:10.1111/j.1539-6053.2008.00033.x.

Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (2009). Introduction: A brief history of neuroeconomics. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain* (1st ed., pp. 1-12). London, UK: Elsevier Academic Press.

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition, 123*(1), 21-32. doi:10.1016/j.cognition.2011.12.002

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review, 94*(2), 236-254.

Goldstein, W. M., & Hogarth, R. M. (1997). Judgment and decision research: Some historical context. In W. M. Goldstein, & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections and controversies* (pp. 3-65). Cambridge, UK: Cambridge University Press.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*, 129–166. doi:0010-0285/99

Gorman, J. (2020). *The Coronavirus, by the numbers*. Retrieved from The New York Times: https://www.nytimes.com/2020/03/05/health/coronavirus-deaths-rates.html

Goudie, R. J., Turner, R. M., & De Angelis, D. (2017). MultiBUGS: A parallel implementation of the BUGS modelling framework for faster Bayesian inference. Retrieved from https://arxiv.org/abs/1704.03216

Graven, M. H. (2014). Poverty, inequality and mathematics performance: the case of South Africa's post-apartheid context. *ZDM Mathematics Education, 46*(7), 1039-1049. doi:10.1007/s11858-013-0566-7

Gravetter, F. J., & Forzano, L.-A. B. (2012). *Research methods for the behavioral sciences* (4th ed.). Belmont, CA: Wadsworth Cengage Learning.

Grün-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds & Machines, 26*, 149-183. doi:10.1007/s11023-015-9367-9

Gujarati, D. N. (2003). *Basic econometrics.* New York, NY: Mc-Graw Hill.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., . . . Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods, 48*(3), 829–842. Retrieved from https://doi.org/10.3758/s13428-015-0642-8

Hacking, I. (2006). *Probability: A philosophical study of early ideas about probability, induction and statistical inference* (2nd ed.). Cambridge, UK: Cambridge University Press.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Upper Saddle River, NJ: Pearson.

Hajimoladarvish, N. (2018). How do people reduce compound lotteries? *Journal of Behavioral and Experimental Economics, 75*, 126-133. doi:10.1016/j.socec.2018.06.002

Hammond, K. R. (2001). The organism and the causal texture of the environment [1935]. In K. R. Hammond, & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications.* Oxford, UK: Oxford University Press.

Harrison, G. W. (2008). Maximum likelihood estimation of utility functions using Stata. In *CEAR Working Paper 06-12.* College of Business Administration, University of Central Florida, 2006: Department of Economics.

Harrison, G. W. (2013). Field experiments and methodological intolerance. *Journal of Economic Methodology, 20*(2), 103-117. doi:10.1080/1350178X.2013.804678

Harrison, G. W. (2014). Real choices and hypothetical choices. In S. Hess, & A. Daly (Eds.), *Handbook of choice modelling* (pp. 236–254). Elgar original reference. Retrieved from https://doi.org/10.4337/9781781003152

Harrison, G. W. (2019). The methodologies of behavioral econometrics. In M. Nagatsu, & A. Ruzzene (Eds.), *Contemporary philosophy and social science: An interdisciplinary dialogue* (pp. 107-137). London, UK: Bloomsbury Academic.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.

Harrison, G. W., & Ross, D. (2017). The empirical adequacy of cumulative prospect theory and its implications for normative assessment. *Journal of Economic Methodology, 24*(2), 150–165. doi:https://doi.org/10.1080/1350178X.2017.1309753

Harrison, G. W., & Rutström, E. E. (2008). Risk Aversion in Experiments. In J. C. Cox, & G. W. Harrison (Eds.), *Research in Experimental Economics* (Vol. 12, pp. 41–196). Bingley, UK: Emerald Group Publishing Limited. doi:10.1016/S0193-2306(08)00003-3

Harrison, G. W., & Rutström, E. E. (2009). Expected utility theory and Prospect theory: One wedding and a decent funeral. *Experimental Economics, 12*(2), 133-158. doi:10.1007/s10683-008-9203-7

Harrison, G. W., & Swarthout, J. T. (2014). Experimental payment protocols and the bipolar behaviorist. *Theory and Decision, 77*(3), 423–438. doi:10.1007/s11238-014-9447-y

Harrison, G. W., & Swarthout, J. T. (2016). Cumulative Prospect Theory in the laboratory: A reconsideration. *CEAR Working Paper 2016-05,* Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016.

Harrison, G. W., Johnson, E., McInnes, M. M., & Rutström, E. E. (2005). Risk Aversion and Incentive Effects: Comment. *The American Economic Review, 95*(3), 896-901.

Harrison, G. W., Lau, M. I., & Rutström, E. E. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics, 109*(2), 341–368. doi:10.1111/j.1467-9442.2007.00496.x

Harstad, R. M., & Selten, R. (2013). Bounded-rationality models: Tasks to become intellectually competitive. *Journal of Economic Literature, 51*(2), 496–511. Retrieved from http://dx.doi.org/10.1257/jel.51.2.496

Hausman, D. M. (2007). Introduction. In D. M. Hausman (Ed.), *The philosophy of economics: An anthology* (3rd ed., pp. 1-38). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511819025.010

Hawthorne, J. (2017). *Inductive Logic.* (E. N. Zalta, Ed.) Retrieved October 3, 2017, from The Stanford Encyclopedia of Philosophy (Spring 2017 Edition): https://plato.stanford.edu/archives/spr2017/entries/logic-inductive/

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods, 51*(2), 961–985. Retrieved from https://doi.org/10.3758/s13428-018-1067-y

Hens, T., & Rieger, M. O. (2016). Decision theory. In T. Hens, & M. O. Rieger, *Financial economics: A concise introduction to classical and behavioral finance* (pp. 15-89). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-662-49688-6_2

Hertwig, R., & Ortmann, A. (2003). Economists' and psychologists' experimental practices: How they differ, why they differ, and how they could converge. In I. Brocas, & J. D. Carrillo (Eds.), *The Psychology of economic decisions.* (pp. 253-272). Oxford, UK: Oxford University Press.

Heukelom, F. (2007). *Kahneman and Tversky and the origin of behavioral economics.* Retrieved from Tinbergen Institute Discussion Paper No. 07-003/1 (January 2007) website: http://dx.doi.org/10.2139/ssrn.956887

Hey, J. D. (2001). Does repetition improve consistency? *Experimental Economics, 4*, 5-54. doi:https://doi.org/10.1023/A:1011486405114

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62*(6), 1291-1326. doi:10.2307/2951750

Hibbard, J. H., Peters, E., Dixon, A., & Tusler, M. (2007). Consumer competencies and the use of comparative quality information: It isn't just about literacy. *Medical Care Research and Review, 64*(4), 379-394.

Hofweber, T. (2018). Logic and Ontology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 ed.). Retrieved from https://plato.stanford.edu/archives/sum2018/entries/logic-ontology/

Holden, S. T., & Tilahun, M. (2019). The devil is in the details: Risk preferences, choice list design, and measurement error. Norwegian University of Life Sciences. Preprint: Centre for Land Tenure Studies Working Papers. *3/19*. Retrieved from https://www.nmbu.no/download/file/fid/40386

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655. doi:10.1257/000282802762024700

Holt, C. A., & Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *The American Economic Review, 95*(3), 902-904.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53-60. Retrieved from www.ejbrm.com

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics, 14*(3), 399–425. doi:https://doi.org/10.1007/s10683-011-9273-9

Jasper, J. D., Bhattacharya, C., & Corser, R. (2017). Numeracy predicts more effortful and elaborative search strategies in a complex risky choice context: A process-tracing approach. *Journal of Behavioral Decision Making, 30*(2), 224-235. doi:10.1002/bdm.1934

Johns Hopkins University. (2020). Retrieved from Coronavirus COVID-19: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

Jojo, Z. M. (2019). Mathematics Continuous Professional Development and its relevance to the new era in South Africa. In I. Hussein Amzat, & N. P. Valdez (Eds.), *Teacher empowerment: Professional development and practices* (pp. 103-119). Singapore: Springer Nature. Retrieved from https://doi.org/10.1007/978-981-10-4151-8

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes, 51*, 296-312.

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-292.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341-350.

Kamatera. (2020). *Cloud servers*. Retrieved from Kamatera Performance Cloud: https://www.kamatera.com

Kanyane, M. H., Pophiwa, N., Masiya, T., Wentzel, M., Viljoen, J., Mdlongwa, T., & Zikhali, T. (2018). Adult education and training: Challenges in local municipalities. *HSRC Review, 16*(1), 15-16.

Keren, G., & Wu, G. (2015). A bird's-eye view of the history of judgment and decision making. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 1-39). West Sussex, UK: John Wiley & Sons, Ltd.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences.* Washington, DC: American Psychological Association. doi:10.1037/14136-010

Knight, F. H. (1921/2006). *Risk, uncertainty and profit.* New York, NY: Cosimo. (Original work published 1921).

Kőszegi, B., & Rabin , M. (2007). Reference-dependent risk attitudes. *The American Economic Review*, *97*(4), 1047-1073.

Koujianou Goldberg, P. (2020). *A silent hero of the coronavirus crisis*. Retrieved from Daily Maverick: https://www.dailymaverick.co.za/opinionista/2020-03-18-a-silent-hero-of-the-coronavirus-crisis/

Krefeld-Schwalb, A., Scheibehenne, B., & Pachur, T. (2019, February 1). Structural parameter interdependencies in computational models of cognition. doi:10.31234/osf.io/pxmnw

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. Busemeyer, J. Townsend, Z. J. Wang, & A. Eidels (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279-299). Oxford, UK: Oxford University Press.

L'Haridon, O., & Vieider, F. M. (2019a). All over the map: A worldwide comparison of risk preferences. *Quantitative Economics, 10*, 185–215. Retrieved from http://qeconomics.org. https://doi.org/10.3982/QE898

L'Haridon, O., & Vieider, F. M. (2019b). Supplement to "All over the map: A worldwide comparison of risk preferences": Appendix A. *Quantitative Economics*, 185–215. Retrieved from http://qeconomics.org. https://doi.org/10.3982/QE898

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior and Organization, 17*(3), 377-400. Retrieved from https://ssrn.com/abstract=994514

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55*, 1-7. doi:10.1016/j.jmp.2010.08.013

Lee, M. D., & Vanpaemel, W. (2018). Determining informative priors for cognitive models. *Psychonomic Bulletin and Review, 25*, 114-127. doi:10.3758/s13423-017-1238-3

Li, S.-C., Lewandowsky, S., & DeBrunner, V. E. (1996). Using parameter sensitivity and interdependence to predict model scope and falsifiability. *Journal of Experimental Psychology: General, 125*(4), 360–369. Retrieved from https://doi.org/10.1037/0096-3445.125.4.360

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25*, 361–381. doi:10.1002/bdm.752

Lilleholt, L. (2019). Cognitive ability and risk aversion: A systematic review and meta analysis. *Judgment and Decision Making, 14*(3), 234-279.

Lindbeck, A. (1985). The prize in economic science in memory of Alfred Nobel. *Journal of Economic Literature, 23*(March), 37-56.

Lipkus, I. M., & Peters, E. (2009). Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. *Health Education and Behavior, 36*(6), 1065-1081. doi:10.1177/1090198109341533

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*(1), 37-44. doi:10.1177/0272989X0102100105

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as Feelings. *Psychological Bulletin, 127*(2), 267-286. doi:10.1037//0033-2909.127.2.267

Loomes, G., & Pogrebna, G. (2014). Measuring individual risk attitudes when preferences are imprecise. *The Economic Journal, 124*(576), 569–593. doi:https://doi.org/10.1111/ecoj.12143

Loomes, G., & Pogrebna, G. (2017). Do preference reversals disappear when we allow for probabilistic choice? *Management Science, 63*(1), 166-184. doi:https://doi.org/10.1287/mnsc.2015.2333

Loomes, G., & Sugden, R. (1982). Regret Theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92*(368), 805-824. Retrieved from https://www.jstor.org/stable/2232669

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica, 65*(260), 581-598. doi:10.1111/1468-0335.00147

Lopes, L. L. (1995). Algebra and process in the modeling of risky choice. *The Psychology of Learning and Motivation*, 177-220.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing, 10*, 325–337.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28*(25), 3049–3067. Retrieved from https://doi.org/10.1002/sim.3680

Lusardi, A. (2012). Numeracy, financial literacy, and financial decision-making. *Numeracy, 5*(1), Article 2. Retrieved from http://dx.doi.org/10.5038/1936-4660.5.1.2

Machina, M. J. (1987). Decision-making in the presence of risk. *Science, 236*, 537-543.

Marinescu, A. (2016). Axiomatical examination of the neoclassical economic model. Logical assessment of the assumptions of neoclassical economic model. *Theoretical and Applied Economics, 2*(607), 47-64.

McDowell, M., Galesic, M., & Gigerenzer, G. (2018). Natural frequencies do foster public understanding of medical tests: Comment on Pighin, Gonzalez, Savadori, and Girotto (2016). *Medical Decision Making, 38*(3), 390–399. doi:10.1177/0272989X18754508

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. *Judgment and Decision Making, 13*(3), 246–259.

Millroth, P., & Juslin, P. (2015). Prospect evaluation as a function of numeracy and probability denominator. *Cognition, 138*, 1-9. Retrieved from http://dx.doi.org/10.1016/j.cognition.2015.01.014

Millroth, P., Nilsson, H., & Juslin, P. (2019). The decision paradoxes motivating Prospect Theory: The prevalence of the paradoxes increases with numerical ability. *Judgment and Decision Making, 14*(4), 513-533. Retrieved from https://ideas.repec.org/a/jdm/journl/v14y2019i4p513-533.html

Mishra, S. (2014). Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Personality and Social Psychology Review, 18*(3), 280-307. doi:10.1177/1088868314530517

Mitchell, O., & Lusardi, A. (2015). Financial literacy and economic outcomes: Evidence and policy implications. *The Journal of Retirement, 3*(1), 107-114. Retrieved from https://doi.org/10.3905/jor.2015.3.1.107

Moran, R. (2016). Thou shalt identify! The identifiability of two high-threshold models in confidence-rating recognition (and super-recognition) paradigms. *Journal of Mathematical Psychology, 73*, 1-11. Retrieved from http://dx.doi.org/10.1016/j.jmp.2016.03.002

Morsanyi, K., & Szucs, D. (2015). Intuition in mathematical and probabilistic reasoning. In R. Cohen Kadosh, & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 180-200). Oxford, UK: Oxford University Press.

Moscati, I. (2018). *Measuring utility: From the marginal revolution to behavioral economics.* Oxford, UK: Oxford Scholarship Online. doi:10.1093/oso/9780199372768.001.0001

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy, 59*(5), 371–404.

Mouton, J. (2002). *Understanding social research* (3rd ed.). Pretoria, South Africa: Van Schaik.

MRC Biostatistics Unit. (n.d.). Retrieved from University of Cambridge: https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/

Mueller, S. M., Schiebener, J., Delazer, M., & Brand, M. (2018). Risk approximation in decision making: approximative numeric abilities predict advantageous decisions under objective risk. *Cognitive Processing, 19*, 297–315. Retrieved from https://doi.org/10.1007/s10339-018-0854-9

Multmeier, J., Gaissmaier, W., & Wegwarth, O. (2014). Collective statistical illiteracy in health. In B. L. Anderson, & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 39-58). Cambridge, UK: Cambridge University Press.

Murphy, R. O., & ten Brincke, R. H. (2017). Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates. *Management Science, Articles in Advance, 64*(1), 308-326. Retrieved from https://doi.org/10.1287/mnsc.2016.2591

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods and Research, 47*(4), 637-664. doi:10.1177/0049124117701488

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L., & Muthén, B. (2010). *Mplus: Statistical analysis with latent variables user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47,* 90–100. doi:10.1016/S0022-2496(02)00028-7

Navarro-Martinez, D., Loomes, G., Isoni, A., Butler, D., & Alaoui, L. (2017). Boundedly rational expected utility theory. University of Warwick working paper. Retrieved from https://mpra.ub.uni-muenchen.de/79893/1/MPRA_paper_79893.pdf

Nelson Mandela Institute. (2020). Retrieved from Nelson Mandela Institute for Education and Rural Development: www.mandelainstitute.org.za

Neth, H., & Gigerenzer, G. (2015). Heuristics: Tools for an uncertain world. In R. Scott, & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource resource* (pp. 1-18). New York, NY: Wiley Online Library. doi:10.1002/9781118900772.etrds0394

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools, 47*(7), 621-634. doi:10.1002/pits.20495

Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011a). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology, 55,* 84-93. doi:10.1016/j.jmp.2010.08.006

Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011b). Hierarchical Bayesian parameter estimation for cumulative prospect theory. Supplement: Data & modeling code. *Journal of Mathematical Psychology, 55*(1), 84-93. doi:10.1016/j.jmp.2010.08.006

Nobelprize.org. (2002, October 9). *The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2002: Press Release.* Retrieved September 13, 2017, from The Official Web Site of the Nobel Prize: https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2002/press.html

Nosek, B. A., Spies, J. R., Cohn, M., Bartmess, E., Lakens, D., Holman, D., . . . Lewis, M. (2020). *OSF Home*. Retrieved from Open Science Collaboration: https://osf.io/vmrgu/

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS: An introduction.* Hoboken, NJ: John Wiley & Sons, Inc., Publication.

OECD. (2016). *The Survey of Adult Skills: Reader's companion.* Paris, France: OECD Skills Studies, OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264258075-en

OECD. (2019). *Skills matter: Additional results from the survey of adult skills.* Paris, France: OECD Skills Studies, OECD Publishing. Retrieved from https://doi.org/10.1787/1f029d8f-en

Omori, Y. (2007). Efficient Gibbs sampler for Bayesian analysis of a sample selection model. *Statistics and Probability Letters, 77*(12), 1300-1311. doi:10.1016/j.spl.2007.03.015

Open Science Collaboration. (2015). Psychology. Estimating the reproducibility of psychological science. *Science, 349*(6251 aac4716). doi:10.1126/science.aac4716

Pachur, T., & Galesic, M. (2013). Strategy selection in risky choice: The impact of numeracy, affect, and cross-cultural differences. *Journal of Behavioral Decision Making, 26*(3), 260-271.

Pachur, T., Hertwig, R., & Wolkewitz, R. (2014). The affect gap in risky choice: Affect-rich outcomes attenuate attention to probability information. *Decision, 1*(1), 64–78. doi:10.1037/dec0000006

Pachur, T., Hertwig, R., Gigerenzer, G., & Brandstätter, E. (2013). Testing process predictions of models of risky choice: a quantitative model comparison approach. *Frontiers in Psychology, 4*(646). doi:10.3389/fpsyg.2013.00646

Pachur, T., Mata, R., & Hertwig, R. (2017a). Who dares, who errs? Disentangling cognitive and motivational roots of age differences in decisions under risk. *Psychological Science, 28*(4), 504-518. doi:10.1177/0956797616687729

Pachur, T., Mata, R., & Hertwig, R. (2017b). Supplemental Material for: Pachur, Mata, & Hertwig: Who dares, who errs? Disentangling cognitive and motivational roots of age differences in decisions under risk. Retrieved from http://journals.sagepub.com/doi/suppl/10.1177/0956797616687729

Pachur, T., Suter, R. S., & Hertwig, H. (2017c). How the twain can meet: Prospect theory and models of heuristics in risky choice. *Cognitive Psychology, 93*, 44-73. doi:http://dx.doi.org/10.1016/j.cogpsych.2017.01.001

Patalano, A. L., Saltiel, J. R., Machlin, L., & Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychonomic Bulletin and Review, 22*, 1820–1829. doi:10.3758/s13423-015-0849-9

Patel, N., Baker, S. G., & Scherer, L. D. (2019). Evaluating the Cognitive Reflection Test as a measure of intuition/reflection, numeracy, and insight problem solving, and the implications for understanding real-world judgments and beliefs. *Journal of Experimental Psychology: General, 148*(12), 2129-2153. doi:http://dx.doi.org/10.1037/xge0000592

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017a). The risk elicitation puzzle. *Nature Human Behaviour, 1*(11). doi:10.1038/s41562-017-0219-x

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017b). The risk elicitation puzzle: Supplementary information. *Nature Human Behaviour*. doi:10.1038/s41562-017-0219-x

Pendlebury, M. (2013). Elementary logic: A brief introduction. (4th ed.). Johannesburg, South Africa: Department of Philosophy, University of the Witwatersrand.

Peng, P., Wang, T., Wang, C. C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin, 145*(2), 189-236. Retrieved from http://dx.doi.org/10.1037/bul0000182

Peter, F. (2017). Political Legitimacy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). Retrieved from https://plato.stanford.edu/archives/sum2017/entries/legitimacy/

Peters, E., Baker, D. P., Dieckmann, N. F., Leon, J., & Collins, J. (2010). Explaining the effect of education on health: A field study in Ghana. *Psychological Science, 21*(10), 1369-1376. doi:10.1177/0956797610381506

Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., . . . Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PLoS ONE, 12*(7), e0180674. Retrieved from https://doi.org/10.1371/journal.pone.0180674

Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*(5), 407-413.

Petrova, D. G., van der Pligt, J., & Garcia-Retamero, R. (2013). Feeling the numbers: On the interplay between risk, affect, and numeracy. *Behavioral decision making, 27*(3), 191–199. doi:10.1002/bdm.1803

Petrova, D., Garcia-Retamero, R., & Cokely, E. T. (2015). Understanding the harms and benefits of cancer screening: A model of factors that shape informed decision making. *Medical Decision Making, 35*(7), 847–858. doi:10.1177/0272989X15587676

Pighin, S., Gonzalez, M., Savadori, L., & Girotto, V. (2016). Natural frequencies do not foster public understanding of medical test results. *Medical Decision Making, 36*(6), 686-91. doi:10.1177/0272989X16640785

Pleskac, T. J. (2015). Decision and choice: Luce's choice axiom. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 895–900). doi:10.1016/b978-0-08-097086-8.43031-x

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News, 6*(1), 7-11. Retrieved from https://journal.r-project.org/archive/ and https://CRAN.R-project.org/package=coda

Preston, M. G., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *The American Journal of Psychology, 61*(2), 183-193.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41,* 71-90. doi:10.1016/j.dr.2016.06.004

Quiggin, J. (1981). Risk perception and the analysis of risk attitudes. *Australian Journal of Agricultural Economics, 25*(2), 160-169.

Quiggin, J. (1991). Comparative statics for rank-dependent expected utility theory. *Journal of Risk and Uncertainty, 4,* 339-350.

R Development Core Team. (2018). *R: A language and environment for statistical computing [Computer software].* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.Rproject.org

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica, 68*(5), 1281-1292.

Rabin, M. (2013a). An approach to incorporating psychology into economics. *American Economic Review, 83*(5), 617–622. doi:http://dx.doi.org/10.1257/aer.103.3.617

Rabin, M. (2013b). Incorporating limited rationality into economics. *Journal of Economic Literature, 51*(2), 528–543. Retrieved from http://dx.doi.org/10.1257/jel.51.2.528

Rammstedt, B., Lechner, C. M., & Danner, D. (2018). Relationships between personality and cognitive ability: A facet-level analysis. *Journal of Intelligence, 6*(28), 1-13. doi:10.3390/jintelligence6020028

Reyna, V. F., & Brust-Renck, P. G. (2014). A review of theories of numeracy: Psychological mechanisms and implications for medical decision making. In B. L. Anderson, & J. Schulkin, *Numerical reasoning in judgments and decision making about health* (pp. 215-251). Cambridge, UK: Cambridge University Press.

Reyna, V. F., Nelson, W., Han, P., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135*(6), 943-973.

Rieger, M. O., Wang, M., & Hens, T. (2017). Estimating cumulative prospect theory parameters from an international survey. *82*, 567–596. doi:10.1007/s11238-016-9582-8

Rieskamp, J. (2008). The Probabilistic Nature of Preferential Choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(doi), 1446–1465. doi:10.1037/a0013646

RiskLiteracy.org. (2020). Retrieved from Science for Informed Decision Making: http://www.riskliteracy.org/

Robert, C. P., & Casella, G. (2010). *Introducing Monte Carlo Methods with R.* New York, NY: Springer Science+Business Media. doi:10.1007/978-1-4419-1576-4

Rodríguez, V., Andrade, A. D., García-Retamero, R., Anam, R., Rodríguez, R., Lisigurski, M., . . . Ruiz, J. G. (2013). Health literacy, numeracy, and graphical literacy among veterans in primary care and their effect on shared decision making and trust in physicians. *Journal of Health Communication, 18: sup1*, 273-289. doi:https://doi.org/10.1080/10810730.2013.829137

Roos, R. (2003). *News and perspective.* Retrieved from Center for Infectious Disease Research and Policy: http://www.cidrap.umn.edu/news-perspective/2003/05/estimates-sars-death-rates-revised-upward

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin and Review, 12*(4), 573-604.

Rouder, J. N., Morey, R. D., & Pratte, M. S. (2016). Bayesian hierarchical models of cognition. In W. Batchelder, H. Colonius, E. Dzhafarov, & J. Myung (Eds.), *New handbook of mathematical psychology* (pp. 504-551). Cambridge, UK: Cambridge University Press. doi:10.1017/9781139245913.010

SAS Institute Inc. (2014). *SAS/STAT® 13.2 User's Guide.* Cary, NC: SAS Institute Inc.

Schapira, M. M., Walker, C. M., Cappaert, K. J., Ganschow, P. S., Fletcher, K. E., McGinley, E. L., & Jacobs, E. A. (2012). The Numeracy Understanding in Medicine instrument: A measure of health numeracy developed using item response theory. *Medical Decision Making, 32*(6), 851–865. Retrieved from https://doi.org/10.1177/0272989X12447239

Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review, 22*, 391–407. doi:10.3758/s13423-014-0684-4

Schiebener, J., Zamarian, L., Delazer, M., & Brand, M. (2011). Executive functions, categorization of probabilities, and learning from feedback: What does really matter for decision making under explicit risk conditions? *Journal of Clinical and Experimental Neuropsychology, 33*(9), 1025-1039. doi:10.1080/13803395.2011.595702

Schleicher, A. (2019). *PISA 2019: Insights and interpretations.* Retrieved from OECD.org: https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf

Schley, D. R., & Peters, E. (2014). Assessing "economic value": Symbolic-number mappings predict risky and riskless valuations. *Psychological Science, 25*(3), 753–761. doi:10.1177/0956797613515485

Schoemaker, P. J. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature, 20*(2), 529-563.

Scholten, M., & Read, D. (2014). Prospect theory and the "forgotten" fourfold pattern of risk preferences. *Journal of Risk and Uncertainty, 48*, 67-83. doi:10.1007/s11166-014-9183-2

Schuessler, R. (2016). Probability in Medieval and Renaissance Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Retrieved from https://plato.stanford.edu/archives/win2016/entries/probability-medieval-renaissance/

Schwartz, A. (2011). Medical decision making and medical education: challenges and opportunities. *Perspectives in Biology and Medicine, 54*(1), 68-74. doi:10.1353/pbm.2011.0010

Schwartz, L. M., Woloshin, S., & Welch, H. G. (2005). Can patients interpret health information? An assessment of the Medical Data Interpretation Test. *Medical Decision Making, 25*, 290-300. doi:10.1177/0272989X05276860

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*(11), 966-972.

Science Media Centre. (2020, February 12). Retrieved from sciencemediacentre.org: https://www.sciencemediacentre.org/smc-novel-coronavirus-factsheet/

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*(2), 207–222. doi:10.1037/0033-2909.134.2.207

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32*, 1248–1284. doi:10.1080/03640210802414826

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*, 487-510. doi:10.1146/annurev-psych-122216-011845

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129–138.

Simon, H. A. (1976). From substantive to procedural rationality. In T. J. Kastelein, S. K. Kuipers, W. A. Nijenhuis, & G. R. Wagenaar (Eds.), *25 Years of Economic Theory* (pp. 65-86). Boston, MA: Springer. doi:https://doi.org/10.1007/978-1-4613-4367-7_6

Simon, H. A. (1977/1992). Rational decision-making in business organizations: Nobel memorial lecture, 8 December,1977. In A. Lindbeck (Ed.), *Nobel lectures in economic sciences (1969-1980)* (pp. 343-371). Farrer Road, Singapore: World Scientific.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1-19.

Simonson, I., & Kivetz, R. (2018). Bringing (contingent) loss aversion down to earth — A comment on Gal & Rucker's rejection of "Losses loom larger than gains". *Journal of Consumer Psychology, 28*(3), 517-522. doi:10.1002/jcpy.1046

Singh, K. (2007). *Quantitative social research methods.* Thousand Oaks, CA: Sage Publications.

Skagerlund, K., Lind, T., Strömbäck, C., Tinghög, G., & Västfjäll, D. (2018). Financial literacy and the role of numeracy–How individuals' attitude and affinity with numbers influence financial literacy. *Journal of Behavioral and Experimental Economics, 74*, 18-25. Retrieved from https://doi.org/10.1016/j.socec.2018.03.004

SMC. (2020). *Novel Coronavirus (COVID-19) factsheet*. Retrieved from Science Centre Media: www.sciencemediacentre.org/smc-novel-coronavirus-factsheet/

Spaull, N. (2020). *Nic Spaull*. Retrieved from Launching "Bala Wande: Calculating with Confidence"!: https://nicspaull.com/2020/01/27/launching-bala-wande-calculating-with-confidence/

Spektor, M. S., & Kellen, D. (2018). The relative merit of empirical priors in non-identifiable and sloppy models: Applications to models of learning and decision-making. *Psychonomic Bulletin and Review, 25*, 2047–2068. Retrieved from https://doi.org/10.3758/s13423-018-1446-5

Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statistical Science, 19*(1), 156-173. doi:10.1214/088342304000000080

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, statistical methodology., 64*, 583-639.

Stanovich, K. E. (2016). The comprehensive assessment of rational thinking. *Educational Psychologist, 51*(1), 23-34. doi:10.1080/00461520.2015.1125787

Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In D. Preiss, & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching and human development* (pp. 195-237). New York, NY: Springer. doi:10.1038/scientificamericangenius0115-12

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645–726. doi:10.1017/S0140525X00623439

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality Quotient: Towards a test of rational thinking.* Cambridge, MA: The MIT Press.

StataCorp. (2017). Introduction to Bayesian analysis. In *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC. Retrieved from https://www.stata.com

StataCorp. (2017). *Stata Statistical Software: Release 15.* College Station, TX: StataCorp LLC.

Steup, M. (2018). Epistemology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 ed.). Retrieved from https://plato.stanford.edu/archives/win2018/entries/epistemology/

Stevens, H. (2020, March 14). *Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve".* Retrieved from The Washington Post: https://www.washingtonpost.com/graphics/2020/world/corona-simulator/

Stieger, S., & Reips, U. (2016). A limitation of the Cognitive Reflection Test: Familiarity. *PeerJ 4:e2395*. doi:10.7717/peerj.2395

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32*(2), 101–130. doi:10.1007/sl 1166-006-8289-6

Sugden, R. (2011). The behavioural economist and the social planner: To whom should behavioural welfare economics be addressed? *Papers on economics and evolution*(1121). Retrieved from http://hdl.handle.net/10419/57559

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism. *American Economic Review, Papers and Proceedings, 93*(2), 175-179.

Survey Anyplace. (2020). Retrieved from surveyanyplace.com: https://surveyanyplace.com/

Suter, R. S., Pachur, T., Hertwig, R., Endestad, T., & Biele, G. (2015). The neural basis of risky choice with affective outcomes. *PLoS ONE, 10*(4), e0122475. doi:10.1371/journal.pone.0122475

Taylor, M. P. (2013). Bias and brains: Risk aversion and cognitive ability across real and hypothetical settings. *Journal of Risk and Uncertainty, 46*, 299–320. doi:10.1007/s11166-013-9166-8

Teovanović, P., Knežević, G., & Stankov, L. (2015). Individual differences in cognitive biases: Evidence against a one-factor theory of rationality. *Intelligence, 50*, 75-86. doi:10.1016/j.intell.2015.02.008

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness.* New Haven, CT: Yale University Press.

The Conversation. (2020). Retrieved from businesstech.org: https://businesstech.co.za/news/government/147669/why-south-african-schools-had-no-choice-but-to-drop-maths-pass-mark/

TIBCO. (2019). Statistica software. *TIBCO Statistica(R)*. TIBCO Software Inc. Retrieved from https://docs.tibco.com.

Toet, A., Brouwer, A.-M., van den Bosch, K., & Korteling, J. E. (2016). Effects of personal characteristics on susceptibility to decision bias. *International Journal of Humanities and Social Sciences, 8*(5), 1-17. Retrieved from http://aajhss.org/index.php/ijhss

Traczyk, J., & Fulawka, K. (2016). Numeracy moderates the influence of task-irrelevant affect on probability weighting. *Cognition, 151*, 37-41. doi:10.1016/j.cognition.2016.03.002

Traczyk, J., Sobkow, A., Fulawka, K., Kus, J., Petrova, D., & Garcia-Retamero, R. (2018). Numerate decision makers don't use more effortful strategies unless it pays: A process tracing investigation of skilled and adaptive strategy selection in risky decision making. *Judgment and Decision Making, 13*(4), 372–381.

Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis, 9*(2), 163-173. Retrieved from https://www.jstor.org/stable/20010465

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297-323.

Vallverdú, J. (2003). The false dilemma: Bayesian vs. Frequentist. *XIIth International Congress of Logic, Methodology and Philosophy of Science.* Oviedo.

van Aert, R. C., & van Assen, M. A. (2018). Examining reproducibility in psychology: A hybrid method for combining a statistically significant original study and a replication. *Behavior Research Methods*, 1515–1539. doi:10.3758/s13428-017-0967-6

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: the Achilles heel of predictive analytics. *BMC Medicine, 17*, 1-7. Retrieved from https://doi.org/10.1186/s12916-019-1466-7

Van de Ven, A. H. (2007). *Engaged Scholarship.* Oxford, UK: Oxford University Press.

van Inwagen, P., & Sullivan, M. (2018). Metaphysics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 ed.). Retrieved from https://plato.stanford.edu/archives/spr2018/entries/metaphysics/

Van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic Bulletin and Review, 25*, 143–154. doi:10.3758/s13423-016-1015-8

VanderWeele, T. J. (2016). Mediation analysis: A practitioner's guide. *Annual Review of Public Health, 37*, 17-32. doi:10.1146/annurev-publhealth-032315-021402

Venkat, H., & Spaull, N. (2015). What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. *International Journal of Educational Development, 41*, 121–130. Retrieved from https://doi.org/10.1016/j.ijedudev.2015.02.002

Von Gaudecker, H.-M., van Soest, A., & Wengstrom, E. (2011). Heterogeneity in risky choice behavior in a broad population. *The American Economic Review, 101*(2), 664-694. doi:10.1257/aer.101.2.664

Von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior* (3rd ed.). Princeton, NJ: Princeton University Press.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics, 17*, 1329–1344. doi:10.1002/hec.1331

Wakker, P. P. (2010). *Prospect Theory: For risk and ambiguity.* Cambridge, UK: Cambridge University Press.

Weber, P., Binder, K., & Krauss, S. (2018). Why can only 24% solve Bayesian reasoning problems in natural frequencies: Frequency phobia in spite of probability blindness. *Frontiers in Psychology, 9*(1833). doi:10.3389/fpsyg.2018.01833

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making, 26*, 198–212. doi:10.1002/bdm.1751

Welsh, M., Burns, N., & Delfabbro, P. (2013). The Cognitive Reflection Test: How much more than numerical ability? In M. Knauff, N. Sebanz, M. Pauen, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1587–1592). Berlin, Germany: Cognitive Science Society.

Wertheimer, A. (2002). Liberty, coercion, and the limits of the state. In R. L. Simon (Ed.), *The Blackwell guide to social and political philosophy* (pp. 38-59). Oxford, UK: Blackwell Publishers.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide using statistical software.* Boca Raton, Fla: Chapman and Hall/CRC.

Whiteside, A. (2020). *Alan Whiteside*. Retrieved from Covid-19 Watch: The numbers climb to a mixed response: https://alan-whiteside.com/

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J. C. Cox, & G. W. Harrison (Eds.), *Research in experimental economics: Risk aversion in experiments* (Vol. 12, pp. 197-292). Bingley, UK: Emerald Group Publishing. doi:10.1016/S0193-2306(08)00004-5

Wilcox, N. T. (2010). A comparison of three probabilistic models of binary discrete choice under risk. In *Working Paper.* Economic Science Institute: Chapman University.

Wilcox, N. T. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics, 162*(1), 89-104. doi:10.1016/j.jeconom.2009.10.012

Wilcox, N. T. (2015). Error and generalization in discrete choice under risk. ESI Working Paper 15-11. Retrieved from http://digitalcommons.chapman.edu/esi_working_papers/160

Wotela, K. (2017). Conceptualising conceptual frameworks in public and business management research. Kidmore End: Academic Conferences International Limited. Retrieved from https://search.proquest.com/docview/1967314155?accountid=15083

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*(12), 1676-1690. doi:https://doi.org/10.1287/mnsc.42.12.1676

Wu, G., Zhang, J., & Gonzalez, R. (2004). Decision under risk. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 399-423). Ma, USA: Blackwell Publishing Ltd.

Xiaoxue, G. (2017). Three Essays on Experimental Economics and Individual Decision Making Under Risk. *Dissertation, Georgia State University*. Retrieved from https://scholarworks.gsu.edu/econ_diss/136

Xing, C., Paul, J., Zax, A., Cordes, S., Barth, H., & Patalano, A. L. (2019). Probability range and probability distortion in a gambling task. *Acta Psychologica, 197*(June), 39-51. doi:10.1016/j.actpsy.2019.03.003

Yechiam, E. (2019). Acceptable losses: the debatable origins of loss aversion. *Psychological Research, 83*, 1327–1339. doi:https://doi.org/10.1007/s00426-018-1013-8

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin, 139*(2), 497–518. doi:10.1037/a0029383

Zabell, S. L. (1990). Daniel Bernoulli. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave: Utility and probability* (pp. 12-14). London, UK: The Macmillan Press Limited. doi:10.1007/978-1-349-20568-4

Zhou, W., & Hey, J. (2018). Context matters. *Experimental Economics, 21*, 723–756. doi:10.1007/s10683-017-9546-z

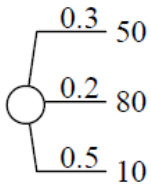
Appendix A: Calculation of rank-dependence

The steps required are described as follows (Wakker, 2010, pp. 165-166):

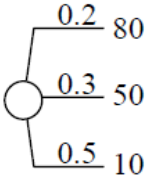
i. Rank the outcomes from best to worst.
ii. Calculate the rank *r* for each outcome.
iii. Calculate the *w* value for each rank.
iv. Calculate the decision weight i.e. the marginal *w* contribution of the probability *p* of each outcome x to its rank, i.e. calculate $w(p+r) - w(r)$ where $w(p+r)$ is the rank of the outcome in the prospect next-worse to x.
v. Determine the utility of each outcome.
vi. Multiply the utility of each outcome by its decision weight and sum the results.

Example:

Calculate the rank-dependent utility for the following prospect: (.3, 50; .2, 80; .5, 10). Assume $w(p) = p^2$ and U(x) = x. Using a decision tree to depict the prospect (Wakker, 2010, pp. 165-166):



i. Ranking the outcomes 50, 80, 10 from best to worst gives 80, 50, 10, i.e.



ii. Calculate the rank r for each outcome, i.e.

r(80) = 0

r(50) = 0.2

r(10) = 0.5

iii. Calculate the *w* value for each rank, i.e.

80: $w(0) = 0$

50: $w(0.2) = (0.2)^2 = .04$

10: $w(0.5) = (0.5)^2 = .25$

iv. Calculate the decision weight i.e. the marginal w contribution of the probability p of each outcome x to its rank, i.e.

- The decision weight of 80 is $w(0.2) - w(0) = 0.04 - 0 = 0.04$
- The decision weight of 50 is $w(0.5) - w(0.2) = 0.25 - 0.04 = 0.21$
- The decision weight of 10 is $w(1) - w(0.5) = 1 - 0.25 = 0.75$

v. Determine the utility of each outcome, i.e.

- $Y(80) = 80$.
- $Y(50) = 50$.
- $Y(10) = 10$.

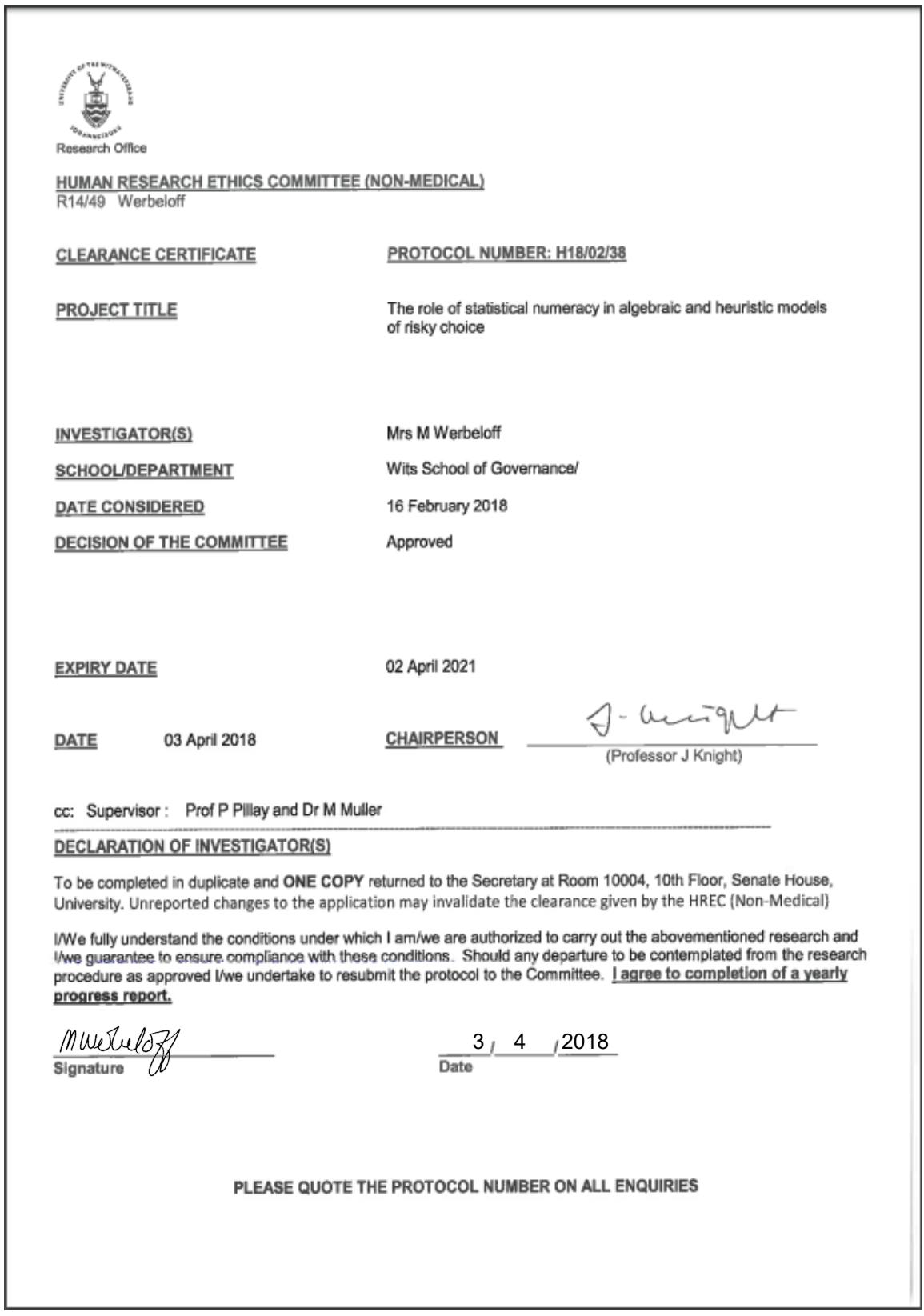
vi. Multiply the utility of each outcome by its decision weight and sum the results, i.e.

$$RDU = w(0.2)U(80) + [w(0.5) - w(0.2)]U(50) + [1 - w(0.5)]U(10)$$

$$= 0.04 \times 80 + 0.21 \times 50 + 0.75 \times 10 = \mathbf{21.2}.$$

Note: Reprinted from Wakker (2010, pp. 165-166)

Appendix B: Ethical clearance



Research Office

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)
R14/49 Werbeloff

CLEARANCE CERTIFICATE	PROTOCOL NUMBER: H18/02/38
PROJECT TITLE	The role of statistical numeracy in algebraic and heuristic models of risky choice
INVESTIGATOR(S)	Mrs M Werbeloff
SCHOOL/DEPARTMENT	Wits School of Governance/
DATE CONSIDERED	16 February 2018
DECISION OF THE COMMITTEE	Approved
EXPIRY DATE	02 April 2021

DATE 03 April 2018 CHAIRPERSON (Professor J Knight)

cc: Supervisor : Prof P Pillay and Dr M Muller

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University. Unreported changes to the application may invalidate the clearance given by the HREC (Non-Medical)

I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to completion of a yearly progress report.**

Signature 3 / 4 / 2018 Date

PLEASE QUOTE THE PROTOCOL NUMBER ON ALL ENQUIRIES

Figure B 1: Ethical clearance certificate

Appendix C: Selected screenshots of GLM2 questionnaire

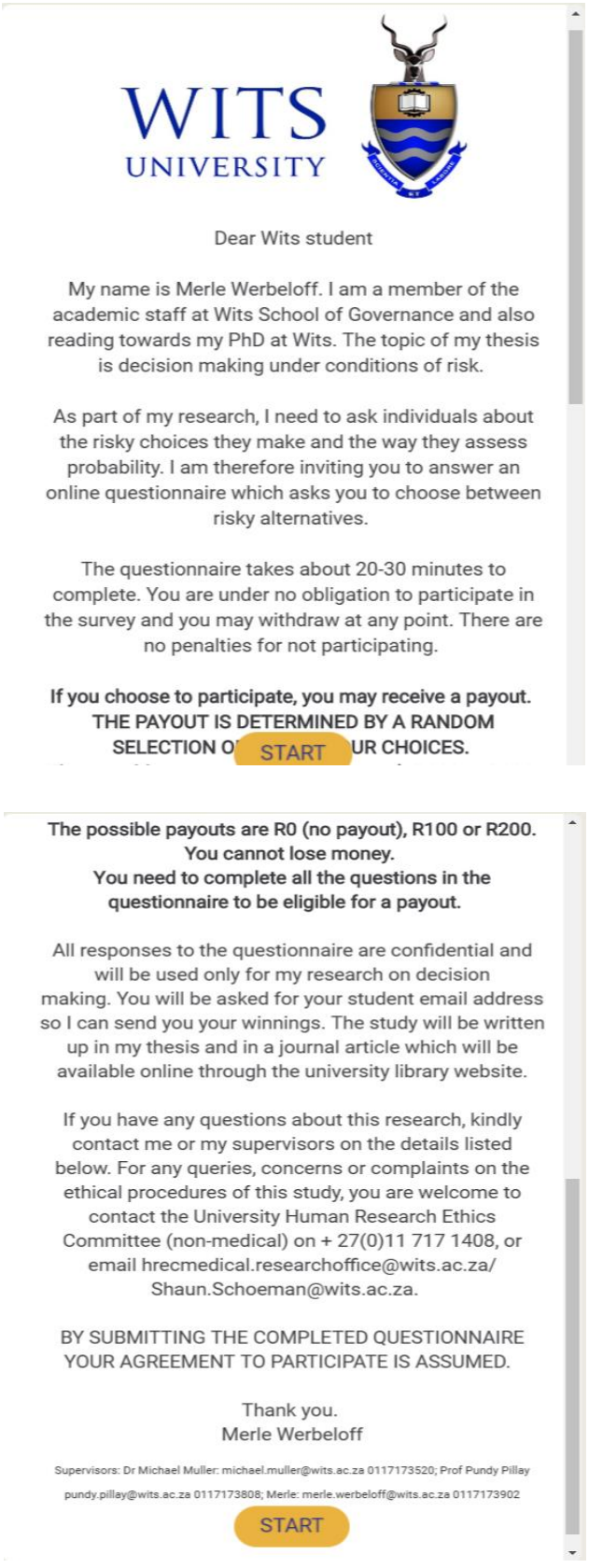


Figure C 1: Screenshots of introduction to the GLM2 questionnaire

Appendix D: Screenshots of lottery pairs

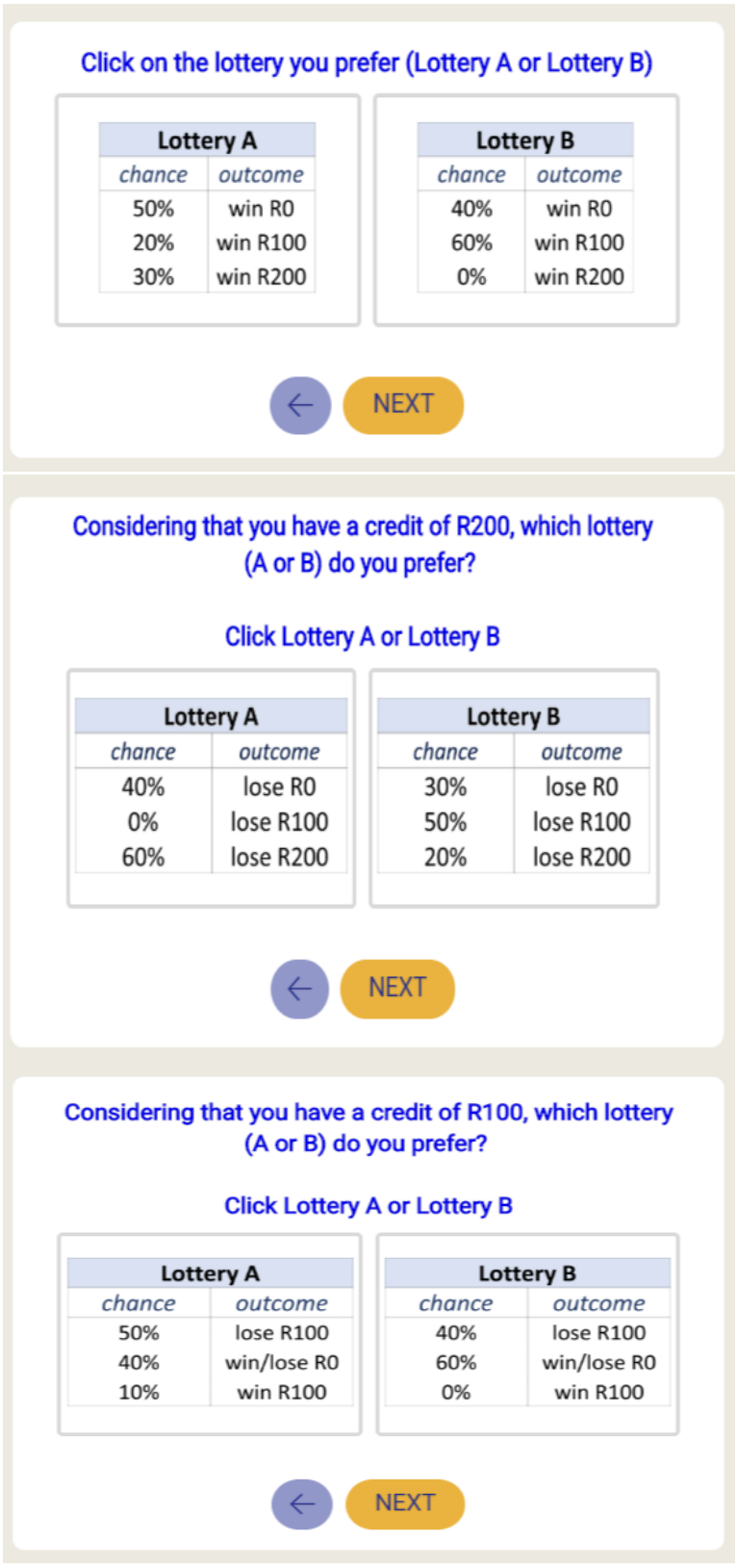


Figure D 1: Screenshots of risky choices framed as gains, losses and mixed prospect pairs

Appendix E: Prospect details

Table E 1: Prospect details for GAINS1

Loomes Sugden triangle	Gradient	chord	Riskier A Left *p*1	Left *p*2	Left *p*3	Safer B Right *p*1	Right *p*2	Right *p*3	Left/ Right *x1*	Left/ Right *x*2	Left/ Right *x*3	Endowment	Type
1	0.67	18.03	15	0	85	0	25	75	0	100	200	0	Gains
1	0.67	18.03	30	0	70	15	25	60	0	100	200	0	Gains
1	0.67	36.06	30	0	70	0	50	50	0	100	200	0	Gains
1	0.67	18.03	15	25	60	0	50	50	0	100	200	0	Gains
1	0.67	18.03	15	75	10	0	100	0	0	100	200	0	Gains
1	0.67	72.11	60	0	40	0	100	0	0	100	200	0	Gains
1	0.67	54.08	60	0	40	15	75	10	0	100	200	0	Gains
1	0.67	18.03	90	0	10	75	25	0	0	100	200	0	Gains
2	1.00	14.14	10	0	90	0	20	80	0	100	200	0	Gains
2	1.00	56.57	50	0	50	10	80	10	0	100	200	0	Gains
2	1.00	70.71	50	0	50	0	100	0	0	100	200	0	Gains
2	1.00	14.14	10	80	10	0	100	0	0	100	200	0	Gains
2	1.00	28.28	70	0	30	50	40	10	0	100	200	0	Gains
2	1.00	42.43	70	0	30	40	60	0	0	100	200	0	Gains
2	1.00	14.14	50	40	10	40	60	0	0	100	200	0	Gains
2	1.00	14.14	90	0	10	80	20	0	0	100	200	0	Gains
3	1.50	18.03	10	0	90	0	25	75	0	100	200	0	Gains
3	1.50	54.08	40	0	60	10	75	15	0	100	200	0	Gains
3	1.50	72.11	40	0	60	0	100	0	0	100	200	0	Gains
3	1.50	18.03	10	75	15	0	100	0	0	100	200	0	Gains
3	1.50	18.03	70	0	30	60	25	15	0	100	200	0	Gains
3	1.50	36.06	70	0	30	50	50	0	0	100	200	0	Gains
3	1.00	18.03	60	25	15	50	50	0	0	100	200	0	Gains
3	1.50	18.03	85	0	15	75	25	0	0	100	200	0	Gains
4	2.00	22.36	10	0	90	0	30	70	0	100	200	0	Gains
4	2.00	44.72	40	0	60	20	60	20	0	100	200	0	Gains
4	2.00	67.08	40	0	60	10	90	0	0	100	200	0	Gains
4	2.00	22.36	20	60	20	10	90	0	0	100	200	0	Gains
4	2.00	22.36	60	0	40	50	30	20	0	100	200	0	Gains
4	2.00	44.72	60	0	40	40	60	0	0	100	200	0	Gains
4	2.00	22.36	50	30	20	40	60	0	0	100	200	0	Gains
4	2.00	22.36	80	0	20	70	30	0	0	100	200	0	Gains
5	3.00	31.62	10	0	90	0	40	60	0	100	200	0	Gains
5	3.00	47.43	25	0	75	10	60	30	0	100	200	0	Gains
5	3.00	79.06	25	0	75	0	100	0	0	100	200	0	Gains
5	3.00	31.62	10	60	30	0	100	0	0	100	200	0	Gains
5	3.00	31.62	50	20	30	40	60	0	0	100	200	0	Gains
5	3.00	47.43	55	0	45	40	60	0	0	100	200	0	Gains
5	3.00	15.81	55	0	45	50	20	30	0	100	200	0	Gains
5	3.00	31.62	70	0	30	60	40	0	0	100	200	0	Gains

Note for tables E 1 to E 4: * indicates a prospect compiled by the researcher, i.e. not derived from Loomes and Sugden (1998) or Harrison and Swarthout (2016).

Table E 2: Prospect details for GAINS2

Loomes Sugden triangle	Gradient	chord	Riskier A Left *p*1	Left *p*2	Left *p*3	Safer B Right *p*1	Right *p*2	Right *p*3	Left/ Right *x1*	Left/ Right *x*2	Left/ Right *x*3	Endowment	Type
1	0.67	18.03	15	0	85	0	25	75	0	100	200	0	Gains
1*	0.67	18.03	60	0	40	45	25	30	0	100	200	0	Gains
1	0.67	18.03	15	75	10	0	100	0	0	100	200	0	Gains
1	0.67	18.03	90	0	10	75	25	0	0	100	200	0	Gains
2	1.00	14.14	10	0	90	0	20	80	0	100	200	0	Gains
2*	1.00	14.14	50	0	50	40	20	40	0	100	200	0	Gains
2	1.00	14.14	10	80	10	0	100	0	0	100	200	0	Gains
2	1.00	14.14	90	0	10	80	20	0	0	100	200	0	Gains
3	1.50	18.03	10	0	90	0	25	75	0	100	200	0	Gains
3	1.50	18.03	70	0	30	60	25	15	0	100	200	0	Gains
3	1.50	18.03	10	75	15	0	100	0	0	100	200	0	Gains
3	1.50	18.03	85	0	15	75	25	0	0	100	200	0	Gains
4	2.00	22.36	10	0	90	0	30	70	0	100	200	0	Gains
4	2.00	22.36	60	0	40	50	30	20	0	100	200	0	Gains
4*	2.00	22.36	10	70	20	0	100	0	0	100	200	0	Gains
4	2.00	22.36	80	0	20	70	30	0	0	100	200	0	Gains
5	3.00	31.62	10	0	90	0	40	60	0	100	200	0	Gains
5*	3.00	31.62	55	0	45	45	40	15	0	100	200	0	Gains
5	3.00	31.62	10	60	30	0	100	0	0	100	200	0	Gains
5	3.00	31.62	70	0	30	60	40	0	0	100	200	0	Gains
6	4.00	41.23	10	0	90	0	50	50	0	100	200	0	Gains
6	4.00	41.23	40	0	60	30	50	20	0	100	200	0	Gains
6	4.00	41.23	10	50	40	0	100	0	0	100	200	0	Gains
6	4.00	41.23	60	0	40	50	50	0	0	100	200	0	Gains
1	0.67	18.03	15	0	85	0	25	75	0	100	300	0	Gains
1*	0.67	18.03	60	0	40	45	25	30	0	100	300	0	Gains
1	0.67	18.03	15	75	10	0	100	0	0	100	300	0	Gains
1	0.67	18.03	90	0	10	75	25	0	0	100	300	0	Gains
2	1.00	14.14	10	0	90	0	20	80	0	100	300	0	Gains
2*	1.00	14.14	50	0	50	40	20	40	0	100	300	0	Gains
2	1.00	14.14	10	80	10	0	100	0	0	100	300	0	Gains
2	1.00	14.14	90	0	10	80	20	0	0	100	300	0	Gains
3	1.50	18.03	10	0	90	0	25	75	0	100	300	0	Gains
3	1.50	18.03	70	0	30	60	25	15	0	100	300	0	Gains
3	1.50	18.03	10	75	15	0	100	0	0	100	300	0	Gains
3	1.50	18.03	85	0	15	75	25	0	0	100	300	0	Gains
4	2.00	22.36	10	0	90	0	30	70	0	100	300	0	Gains
4	2.00	22.36	60	0	40	50	30	20	0	100	300	0	Gains
4*	2.00	22.36	10	70	20	0	100	0	0	100	300	0	Gains
4	2.00	22.36	80	0	20	70	30	0	0	100	300	0	Gains
5	3.00	31.62	10	0	90	0	40	60	0	100	300	0	Gains
5*	3.00	31.62	55	0	45	45	40	15	0	100	300	0	Gains
5	3.00	31.62	10	60	30	0	100	0	0	100	300	0	Gains
5	3.00	31.62	70	0	30	60	40	0	0	100	300	0	Gains
6	4.00	41.23	10	0	90	0	50	50	0	100	300	0	Gains
6	4.00	41.23	40	0	60	30	50	20	0	100	300	0	Gains
6	4.00	41.23	10	50	40	0	100	0	0	100	300	0	Gains
6	4.00	41.23	60	0	40	50	50	0	0	100	300	0	Gains

Table E 3: Prospect details for GLM1

Loomes Sugden triangle	Gradient	chord	Riskier A Left *p*1	Left *p*2	Left *p*3	Safer B Right *p*1	Right *p*2	Right *p*3	Left/ Right *x1*	Left/ Right *x*2	Left/ Right *x*3	Endowment	Type
2	1.00	14.14	10	0	90	0	20	80	0	100	200	0	Gains
2	1.00	56.57	50	0	50	10	80	10	0	100	200	0	Gains
2	1.00	70.71	50	0	50	0	100	0	0	100	200	0	Gains
2	1.00	14.14	10	80	10	0	100	0	0	100	200	0	Gains
2	1.00	28.28	70	0	30	50	40	10	0	100	200	0	Gains
2	1.00	42.43	70	0	30	40	60	0	0	100	200	0	Gains
2	1.00	14.14	50	40	10	40	60	0	0	100	200	0	Gains
2	1.00	14.14	90	0	10	80	20	0	0	100	200	0	Gains
5	3.00	31.62	10	0	90	0	40	60	0	100	200	0	Gains
5	3.00	47.43	25	0	75	10	60	30	0	100	200	0	Gains
5	3.00	79.06	25	0	75	0	100	0	0	100	200	0	Gains
5	3.00	31.62	10	60	30	0	100	0	0	100	200	0	Gains
5	3.00	31.62	50	20	30	40	60	0	0	100	200	0	Gains
5	3.00	47.43	55	0	45	40	60	0	0	100	200	0	Gains
5	3.00	15.81	55	0	45	50	20	30	0	100	200	0	Gains
5	3.00	31.62	70	0	30	60	40	0	0	100	200	0	Gains
2	1.00	14.14	10	0	90	0	20	80	0	-100	-200	200	Losses
2	1.00	56.57	50	0	50	10	80	10	0	-100	-200	200	Losses
2	1.00	70.71	50	0	50	0	100	0	0	-100	-200	200	Losses
2	1.00	14.14	10	80	10	0	100	0	0	-100	-200	200	Losses
2	1.00	28.28	70	0	30	50	40	10	0	-100	-200	200	Losses
2	1.00	42.43	70	0	30	40	60	0	0	-100	-200	200	Losses
2	1.00	14.14	50	40	10	40	60	0	0	-100	-200	200	Losses
2	1.00	14.14	90	0	10	80	20	0	0	-100	-200	200	Losses
5	3.00	31.62	10	0	90	0	40	60	0	-100	-200	200	Losses
5	3.00	47.43	25	0	75	10	60	30	0	-100	-200	200	Losses
5	3.00	79.06	25	0	75	0	100	0	0	-100	-200	200	Losses
5	3.00	31.62	10	60	30	0	100	0	0	-100	-200	200	Losses
5	3.00	31.62	50	20	30	40	60	0	0	-100	-200	200	Losses
5	3.00	47.43	55	0	45	40	60	0	0	-100	-200	200	Losses
5	3.00	15.81	55	0	45	50	20	30	0	-100	-200	200	Losses
5	3.00	31.62	70	0	30	60	40	0	0	-100	-200	200	Losses
2	1.00	14.14	10	0	90	0	20	80	-100	0	100	100	Mixed
2	1.00	56.57	50	0	50	10	80	10	-100	0	100	100	Mixed
2	1.00	70.71	50	0	50	0	100	0	-100	0	100	100	Mixed
2	1.00	14.14	10	80	10	0	100	0	-100	0	100	100	Mixed
2	1.00	28.28	70	0	30	50	40	10	-100	0	100	100	Mixed
2	1.00	42.43	70	0	30	40	60	0	-100	0	100	100	Mixed
2	1.00	14.14	50	40	10	40	60	0	-100	0	100	100	Mixed
2	1.00	14.14	90	0	10	80	20	0	-100	0	100	100	Mixed
5	3.00	31.62	10	0	90	0	40	60	-100	0	100	100	Mixed
5	3.00	47.43	25	0	75	10	60	30	-100	0	100	100	Mixed
5	3.00	79.06	25	0	75	0	100	0	-100	0	100	100	Mixed
5	3.00	31.62	10	60	30	0	100	0	-100	0	100	100	Mixed
5	3.00	31.62	50	20	30	40	60	0	-100	0	100	100	Mixed
5	3.00	47.43	55	0	45	40	60	0	-100	0	100	100	Mixed
5	3.00	15.81	55	0	45	50	20	30	-100	0	100	100	Mixed
5	3.00	31.62	70	0	30	60	40	0	-100	0	100	100	Mixed

Table E 4: Prospect details for GLM2

Loomes Sugden triangle	Gradient	chord	Riskier A Left *p*1	Left *p*2	Left *p*3	Safer B Right *p*1	Right *p*2	Right *p*3	Left/ Right *x1*	Left/ Right *x*2	Left/ Right *x*3	Endowment	Type
1	0.67	18.03	15	0	85	0	25	75	0	100	200	0	Gains
1*	0.67	18.03	60	0	40	45	25	30	0	100	200	0	Gains
1	0.67	18.03	15	75	10	0	100	0	0	100	200	0	Gains
1	0.67	18.03	90	0	10	75	25	0	0	100	200	0	Gains
2	1	14.14	10	0	90	0	20	80	0	100	200	0	Gains
2*	1	14.14	50	0	50	40	20	40	0	100	200	0	Gains
2	1	14.14	10	80	10	0	100	0	0	100	200	0	Gains
2	1	14.14	90	0	10	80	20	0	0	100	200	0	Gains
5	3	31.62	10	0	90	0	40	60	0	100	200	0	Gains
5*	3	31.62	55	0	45	45	40	15	0	100	200	0	Gains
5	3	31.62	10	60	30	0	100	0	0	100	200	0	Gains
5	3	31.62	70	0	30	60	40	0	0	100	200	0	Gains
6	4	41.23	10	0	90	0	50	50	0	100	200	0	Gains
6	4	41.23	40	0	60	30	50	20	0	100	200	0	Gains
6	4	41.23	10	50	40	0	100	0	0	100	200	0	Gains
6	4	41.23	60	0	40	50	50	0	0	100	200	0	Gains
1	0.67	18.03	15	0	85	0	25	75	0	-100	-200	200	Losses
1*	0.67	18.03	60	0	40	45	25	30	0	-100	-200	200	Losses
1	0.67	18.03	15	75	10	0	100	0	0	-100	-200	200	Losses
1	0.67	18.03	90	0	10	75	25	0	0	-100	-200	200	Losses
2	1	14.14	10	0	90	0	20	80	0	-100	-200	200	Losses
2*	1	14.14	50	0	50	40	20	40	0	-100	-200	200	Losses
2	1	14.14	10	80	10	0	100	0	0	-100	-200	200	Losses
2	1	14.14	90	0	10	80	20	0	0	-100	-200	200	Losses
5	3	31.62	10	0	90	0	40	60	0	-100	-200	200	Losses
5*	3	31.62	55	0	45	45	40	15	0	-100	-200	200	Losses
5	3	31.62	10	60	30	0	100	0	0	-100	-200	200	Losses
5	3	31.62	70	0	30	60	40	0	0	-100	-200	200	Losses
6	4	41.23	10	0	90	0	50	50	0	-100	-200	200	Losses
6	4	41.23	40	0	60	30	50	20	0	-100	-200	200	Losses
6	4	41.23	10	50	40	0	100	0	0	-100	-200	200	Losses
6	4	41.23	60	0	40	50	50	0	0	-100	-200	200	Losses
1	0.67	18.03	15	0	85	0	25	75	-100	0	100	100	Mixed
1*	0.67	18.03	60	0	40	45	25	30	-100	0	100	100	Mixed
1	0.67	18.03	15	75	10	0	100	0	-100	0	100	100	Mixed
1	0.67	18.03	90	0	10	75	25	0	-100	0	100	100	Mixed
2	1	14.14	10	0	90	0	20	80	-100	0	100	100	Mixed
2*	1	14.14	50	0	50	40	20	40	-100	0	100	100	Mixed
2	1	14.14	10	80	10	0	100	0	-100	0	100	100	Mixed
2	1	14.14	90	0	10	80	20	0	-100	0	100	100	Mixed
5	3	31.62	10	0	90	0	40	60	-100	0	100	100	Mixed
5*	3	31.62	55	0	45	45	40	15	-100	0	100	100	Mixed
5	3	31.62	10	60	30	0	100	0	-100	0	100	100	Mixed
5	3	31.62	70	0	30	60	40	0	-100	0	100	100	Mixed
6	4	41.23	10	0	90	0	50	50	-100	0	100	100	Mixed
6	4	41.23	40	0	60	30	50	20	-100	0	100	100	Mixed
6	4	41.23	10	50	40	0	100	0	-100	0	100	100	Mixed
6	4	41.23	60	0	40	50	50	0	-100	0	100	100	Mixed

Appendix F: Numeracy tests – original and revised

Table F 1: Numeracy tests – original and revised items used in the thesis

Original	Thesis
Schwartz/Lipkus (Lipkus, Samsa, & Rimer, 2001; Schwartz, Woloshin, Black, & Welch, 1997)	
Imagine that we rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (2, 4, or 6)?	Imagine that we rolled a fair, six-sided die 1,000 times. (The word 'die' is the singular for dice.) Out of 1,000 rolls, about how many times do you think the die would come up even (2, 4, or 6)?
In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS?	In a lottery, the chance of winning a R10 prize is 1%. About how many people would win a R10 prize if 1,000 people each buy a single lottery ticket?
In ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?	In a competition, the chance of winning a car is 1 in 1,000. What percent of tickets win the car?
Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10	Which of the following numbers represents the biggest risk of getting a disease? 1 in 1,000; 1 in 100; 1 in 10
Berlin Numeracy Test (Cokely et al., 2012)	
Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?	Imagine we are throwing a five-sided die 50 times. (The word 'die' is the singular for dice.) On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3, or 5)?
Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in a choir 100 are men. Out of the 500 inhabitants that are not in a choir 300 are men. What is the probability that a randomly drawn man is a member of the choir?	Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir, 100 are men. Out of the 500 inhabitants that are not in the choir, 300 are men. What is the probability that a randomly drawn man is a member of the choir?
Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?	Imagine we are throwing a loaded (unbalanced) die with six sides. (The word 'die' is the singular for dice.) The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?
In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?	In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?
The Cognitive Reflection Test (Frederick, 2005)	
A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	A bat and a ball cost R1.10 in total. The bat costs R1.00 more than the ball. How much does the ball cost?
If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	If it takes 5 machines 5 minutes to make 5 items, how long would it take 100 machines to make 100 items?
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Appendix G: Distributions of numeracy tests

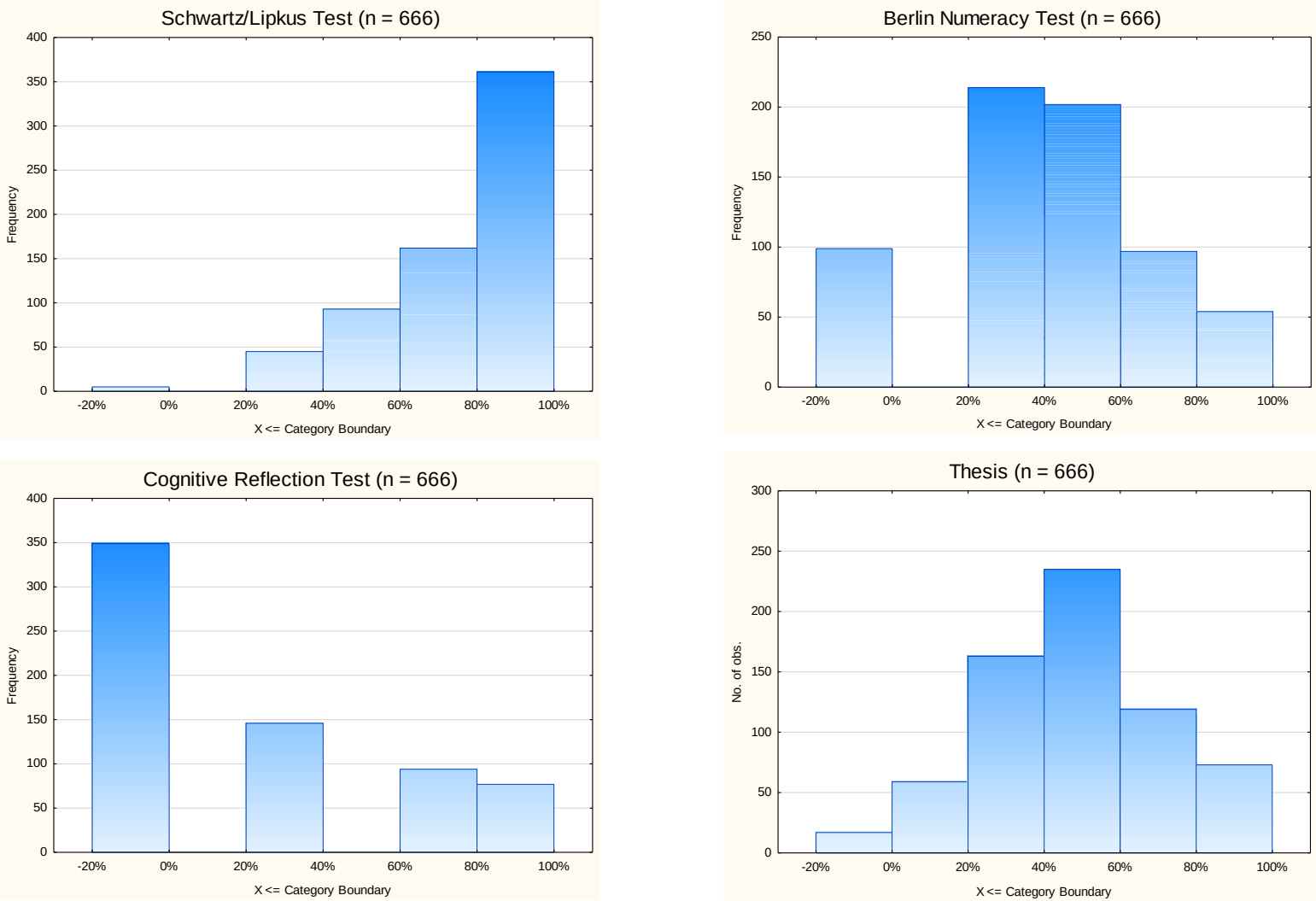


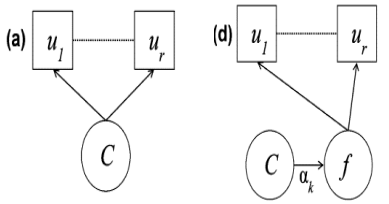
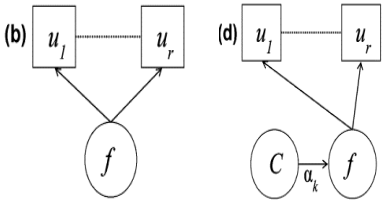
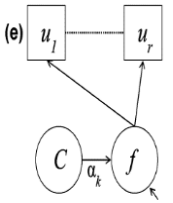
Figure G 1: Distributions of Schwartz/Lipkus, Berlin Numeracy, Cognitive Reflection tests, and thesis test

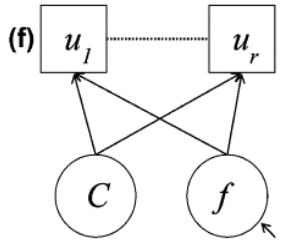
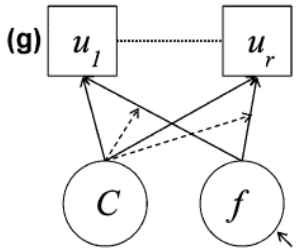
Table G 1: Descriptive statistics of numeracy tests

	Mean	Median	Q1	Q3	*SD*	Skewness	SE Skewness	Kurtosis	SE Kurtosis
Schwartz/ Lipkus	81.1%	100.0%	75.0%	100.0%	24.5%	-1.13	0.09	0.33	0.19
BNT	42.2%	50.0%	25.0%	50.0%	28.4%	0.35	0.09	-0.57	0.19
CRT	28.3%	0.0%	0.0%	66.7%	35.1%	0.90	0.09	-0.54	0.19
Thesis	**48.3%**	**44.4%**	**33.3%**	**66.7%**	**25.3%**	**0.25**	**0.09**	**-0.59**	**0.19**

Appendix H: Variations of the factor mixture model

Table H 1: Variations of the factor mixture model
Note: Source: (Clark et al., 2013)

Input	Article	Model	Classes	Factors	Estimating	Thresholds (tau)	Factor loadings (lambda)	Factor covariance matrix	Factor means	Diagram in Clark et al 2013	Diagram
1a	Clark et al 2013 p.700	FMM-1; 1a	2	2	Estimating both factor means in each class	Class-invariant	Class-invariant	ψ equals zero. Fixing covariance between f1 and f2 at 0	**Class varying** factor means.	1a	
1b	Clark et al 2013 p.701	FMM-1; 1b	3	1	Estimating all factor loadings and factor means	Class-invariant	Class-invariant	equals zero	Class varying factor means	1b	
2	Clark et al 2013 p.701	FMM-2	2	2	Estimating all factor loadings, factor means, variances & covariances	Class-invariant	Class-invariant	Class varying factor covariance matrix (§k)	Class varying factor means	1e	

Input	Article	Model	Classes	Factors	Estimating	Thresholds (tau)	Factor loadings (lambda)	Factor covariance matrix	Factor means	Diagram in Clark et al 2013	Diagram
3	Clark et al., 2013 p.702	FMM-3	2	2	Estimating all factor loadings, factor means, variances & covariances	Class varying	Class-invariant	Class varying factor covariance matrix (§k)	Factor means fixed at zero	1f	
4	Clark et al 2013 p.702	FMM-4	2	2	Estimating all factor loadings, factor means, variances & covariances	Class varying	Class varying	Class varying	Factor means fixed at zero	1g	

Note: Reprinted and adapted from Clark et al. (2013, p. 683)

Appendix I: MultiBUGS model code for the expected utility model (GAINS2 responses of high numeracy group, n = 44)

Extract from Table 18: Maximum likelihood estimates per numeracy group

Model	functional form	numeracy group	n	clusters	parameter	value	robust std. error adjusted for clusters	[95% conf.	interval]
EUT Gains	Power	high	1556	77	alpha	0.439	0.012	0.415	0.462

The extract shows the value of the empirical prior used in the MultiBUGS model code for the MCMC estimation of the expected utility model with power functional form for the high numeracy group. The MultiBUGS code was computed on the responses of 44 high numeracy individuals to the 48 gain-framed prospect pairs of GAINS2.

The value of the empirical prior was derived from the maximum likelihood estimates of the parameter of the expected utility model with the same functional form. The estimate is based on responses to gain-framed prospects of GAINS1, GLM1 and GLM2 of a different set of 77 high numeracy individuals (3, 10 and 64 respondents, respectively; see Table 15).

The parameter value is used in transformed form in the following MultiBUGS model code (6^{th} line highlighted) adapted from Nilsson, Rieskamp, & Wagenmakers (2011b).

```
model{
	for( j in 1 : 44 ) {
		alpha[j] <- phi(alpha.phi[j])
		alpha.phi[j] ~ dnorm(mu.phi.alpha, tau.phi.alpha)
	}
	mu.phi.alpha ~ dnorm(-0.15, tau.phi.alpha)
	tau.phi.alpha ~ dgamma(10, 0.1)
	sigma.phi.alpha <- 1 / sqrt(tau.phi.alpha)
	mu.alpha <- phi(mu.phi.alpha)
	for( j in 1 : 44 ) {
		for( i in 1 : 48 ) {
			v.x.a[i , j] <- pow(prospects.a[i , 1], alpha[j])
			v.y.a[i , j] <- pow(prospects.a[i , 3], alpha[j])
			v.u.a[i , j] <- pow(prospects.a[i , 5], alpha[j])
			r.x.a[i , j] <- prospects.a[i , 2]
			r.y.a[i , j] <- prospects.a[i , 4]
			r.u.a[i , j] <- prospects.a[i , 6]
			Vf.a[i , j] <- r.x.a[i , j] * v.x.a[i , j] + r.y.a[i , j] * v.y.a[i , j] + r.u.a[i , j] * v.u.a[i , j]
		}
		for( i in 1 : 48 ) {
			v.x.b[i , j] <- pow(prospects.b[i , 1], alpha[j])
			v.y.b[i , j] <- pow(prospects.b[i , 3], alpha[j])
			v.u.b[i , j] <- pow(prospects.b[i , 5], alpha[j])
			r.x.b[i , j] <- prospects.b[i , 2]
			r.y.b[i , j] <- prospects.b[i , 4]
			r.u.b[i , j] <- prospects.b[i , 6]
			Vf.b[i , j] <- r.x.b[i , j] * v.x.b[i , j] + r.y.b[i , j] * v.y.b[i , j] + r.u.b[i , j] * v.u.b[i , j]
		}
		for( i in 1 : 48 ) {
			binval[i , j] <- 1 / (1 + exp((( -1) / (v.x.b[i , j] - v.u.b[i , j])) * (Vf.b[i , j] - Vf.a[i , j])))
			rawdata[i , j] ~ dbern(binval[i , j])
		}
	}
}
```

Appendix J: MultiBUGS model code for the rank-dependent utility model (GLM2 responses of high numeracy group, n = 64)

Extract from Table 18: Maximum likelihood estimates per numeracy group

Model	functional form	numeracy group	n	clusters	parameter	value	robust std. error adjusted for clusters	[95% conf.	interval]
RDU Gains and losses	(Tversky & Kahneman, 1992)	high	2634	57	alpha	0.347	0.018	0.312	0.383
					gamma	0.851	0.042	0.768	0.934

The extract shows the values of the empirical priors used in the MultiBUGS model code for the MCMC estimation of the rank-dependent utility model with Tversky and Kahneman (1992) functional form for the high numeracy group. The MultiBUGS code was computed on the responses of 64 high numeracy individuals to the prospect pairs of GLM2.

The values of the empirical priors were derived from the maximum likelihood estimates of the parameters of the rank-dependent model with the same functional form. These estimates are based on responses to GAINS1, GLM1 and GAINS2 of a different set of 57 high numeracy individuals (3, 10 and 44 respondents, respectively; see Table 15).

These parameter values are used in transformed form in the following MultiBUGS model code (10th, 13th and 16th lines highlighted) adapted from Nilsson, Rieskamp, & Wagenmakers (2011b).

```
model{
    for( j in 1 : 64 ) {
        alpha[j] <- phi(alpha.phi[j])
        gamma[j] <- exp(lgamma[j])
        delta[j] <- exp(ldelta[j])
        alpha.phi[j] ~ dnorm(mu.phi.alpha, tau.phi.alpha)
        lgamma[j] ~ dnorm(lmu.gamma, ltau.gamma)
        ldelta[j] ~ dnorm(lmu.delta, ltau.delta)
    }
    mu.phi.alpha ~ dnorm(-0.385, tau.phi.alpha)
    tau.phi.alpha ~ dgamma(10, 0.1)
    sigma.phi.alpha <- 1 / sqrt(tau.phi.alpha)
    lmu.gamma ~ dnorm(-0.2, ltau.gamma)
    ltau.gamma ~ dgamma(10, 0.1)
    lsigma.gamma <- 1 / sqrt(ltau.gamma)
    lmu.delta ~ dnorm(-0.2, ltau.delta)
    ltau.delta ~ dgamma(10, 0.1)
    lsigma.delta <- 1 / sqrt(ltau.delta)
    mu.alpha <- phi(mu.phi.alpha)
    mu.gamma <- exp(lmu.gamma)
    mu.delta <- exp(lmu.delta)
    for( j in 1 : 64 ) {
        for( i in 1 : 16 ) {
            v.x.a[i , j] <- pow(prospects.a[i , 1], alpha[j])
            v.y.a[i , j] <- pow(prospects.a[i , 3], alpha[j])
            v.u.a[i , j] <- pow(prospects.a[i , 5], alpha[j])
            r.x.a[i , j] <- 0
            r.y.a[i , j] <- prospects.a[i , 2]
            r.u.a[i , j] <- prospects.a[i , 2] + prospects.a[i , 4]
            w.x.a[i , j] <- pow(r.x.a[i , j], gamma[j]) / pow(pow(r.x.a[i , j], gamma[j]) + pow(1 - r.x.a[i , j], gamma[j]), 1 / gamma[j])
            w.y.a[i , j] <- pow(r.y.a[i , j], gamma[j]) / pow(pow(r.y.a[i , j], gamma[j]) + pow(1 - r.y.a[i , j], gamma[j]), 1 / gamma[j])
            w.u.a[i , j] <- pow(r.u.a[i , j], gamma[j]) / pow(pow(r.u.a[i , j], gamma[j]) + pow(1 - r.u.a[i , j], gamma[j]), 1 / gamma[j])
```

```
        d.x.a[i , j] <- w.y.a[i , j] - w.x.a[i , j]
        d.y.a[i , j] <- w.u.a[i , j] - w.y.a[i , j]
        d.u.a[i , j] <- 1 - w.u.a[i , j]
        Vf.a[i , j] <- d.x.a[i , j] * v.x.a[i , j] + d.y.a[i , j] * v.y.a[i , j] + d.u.a[i , j] * v.u.a[i , j]
}
for( i in 1 : 16 ) {
        v.x.b[i , j] <- pow(prospects.b[i , 1], alpha[j])
        v.y.b[i , j] <- pow(prospects.b[i , 3], alpha[j])
        v.u.b[i , j] <- pow(prospects.b[i , 5], alpha[j])
        r.x.b[i , j] <- 0
        r.y.b[i , j] <- prospects.b[i , 2]
        r.u.b[i , j] <- prospects.b[i , 2] + prospects.b[i , 4]
        w.x.b[i , j] <- pow(r.x.b[i , j], gamma[j]) / pow(pow(r.x.b[i , j], gamma[j]) + pow(1 - r.x.b[i , j], gamma[j]), 1 / gamma[j])
        w.y.b[i , j] <- pow(r.y.b[i , j], gamma[j]) / pow(pow(r.y.b[i , j], gamma[j]) + pow(1 - r.y.b[i , j], gamma[j]), 1 / gamma[j])
        w.u.b[i , j] <- pow(r.u.b[i , j], gamma[j]) / pow(pow(r.u.b[i , j], gamma[j]) + pow(1 - r.u.b[i , j], gamma[j]), 1 / gamma[j])
        d.x.b[i , j] <- w.y.b[i , j] - w.x.b[i , j]
        d.y.b[i , j] <- w.u.b[i , j] - w.y.b[i , j]
        d.u.b[i , j] <- 1 - w.u.b[i , j]
        Vf.b[i , j] <- d.x.b[i , j] * v.x.b[i , j] + d.y.b[i , j] * v.y.b[i , j] + d.u.b[i , j] * v.u.b[i , j]
}
for( i in 1 : 16 ) {
        binval[i , j] <- 1 / (1 + exp((( -1) / (v.x.b[i , j] - v.u.b[i , j])) * (Vf.b[i , j] - Vf.a[i , j])))
        rawdata[i , j] ~ dbern(binval[i , j])
}
for( i in 17 : 32 ) {
        v.x.a[i , j] <- ( -1) * pow(( -1) * prospects.a[i , 1], alpha[j])
        v.y.a[i , j] <- ( -1) * pow(( -1) * prospects.a[i , 3], alpha[j])
        v.u.a[i , j] <- ( -1) * pow(( -1) * prospects.a[i , 5], alpha[j])
        r.x.a[i , j] <- 0
        r.y.a[i , j] <- prospects.a[i , 2]
        r.u.a[i , j] <- prospects.a[i , 2] + prospects.a[i , 4]
        w.x.a[i , j] <- pow(r.x.a[i , j], delta[j]) / pow(pow(r.x.a[i , j], delta[j]) + pow(1 - r.x.a[i , j], delta[j]), 1 / delta[j])
```

```
                w.y.a[i , j] <- pow(r.y.a[i , j], delta[j]) / pow(pow(r.y.a[i , j], delta[j]) + pow(1 - r.y.a[i , j], delta[j]), 1 / delta[j])
                w.u.a[i , j] <- pow(r.u.a[i , j], delta[j]) / pow(pow(r.u.a[i , j], delta[j]) + pow(1 - r.u.a[i , j], delta[j]), 1 / delta[j])
                d.x.a[i , j] <- w.y.a[i , j] - w.x.a[i , j]
                d.y.a[i , j] <- w.u.a[i , j] - w.y.a[i , j]
                d.u.a[i , j] <- 1 - w.u.a[i , j]
                Vf.a[i , j] <- d.x.a[i , j] * v.x.a[i , j] + d.y.a[i , j] * v.y.a[i , j] + d.u.a[i , j] * v.u.a[i , j]
        }
        for( i in 17 : 32 ) {
                v.x.b[i , j] <- ( -1) * pow(( -1) * prospects.b[i , 1], alpha[j])
                v.y.b[i , j] <- ( -1) * pow(( -1) * prospects.b[i , 3], alpha[j])
                v.u.b[i , j] <- ( -1) * pow(( -1) * prospects.b[i , 5], alpha[j])
                r.x.b[i , j] <- 0
                r.y.b[i , j] <- prospects.b[i , 2]
                r.u.b[i , j] <- prospects.b[i , 2] + prospects.b[i , 4]
                w.x.b[i , j] <- pow(r.x.b[i , j], delta[j]) / pow(pow(r.x.b[i , j], delta[j]) + pow(1 - r.x.b[i , j], delta[j]), 1 / delta[j])
                w.y.b[i , j] <- pow(r.y.b[i , j], delta[j]) / pow(pow(r.y.b[i , j], delta[j]) + pow(1 - r.y.b[i , j], delta[j]), 1 / delta[j])
                w.u.b[i , j] <- pow(r.u.b[i , j], delta[j]) / pow(pow(r.u.b[i , j], delta[j]) + pow(1 - r.u.b[i , j], delta[j]), 1 / delta[j])
                d.x.b[i , j] <- w.y.b[i , j] - w.x.b[i , j]
                d.y.b[i , j] <- w.u.b[i , j] - w.y.b[i , j]
                d.u.b[i , j] <- 1 - w.u.b[i , j]
                Vf.b[i , j] <- d.x.b[i , j] * v.x.b[i , j] + d.y.b[i , j] * v.y.b[i , j] + d.u.b[i , j] * v.u.b[i , j]
        }
        for( i in 17 : 32 ) {
                binval[i , j] <- 1 / (1 + exp((( -1) / (v.x.b[i , j] - v.u.b[i , j])) * (Vf.b[i , j] - Vf.a[i , j])))
                rawdata[i , j] ~ dbern(binval[i , j])
        }
    }
}
```

Appendix K: MultiBUGS model code for cumulative prospect theory (GLM2 responses of high numeracy group, n = 64)

Extract from Table 18: Maximum likelihood estimates per numeracy group

Model	functional form	numeracy group	n	clusters	parameter	value	robust std. error adjusted for clusters	[95% conf.	interval]
CPT Gains, losses and mixed	(Tversky & Kahneman, 1992)	high	2795	57	alpha	0.063	0.065	-0.065	0.191
					lambda	1.262	0.501	.280	2.244
					gamma	0.791	0.135	0.526	1.056

The extract shows the values of the empirical priors used in the MultiBUGS model code for the MCMC estimation of the cumulative prospect model with Tversky and Kahneman (1992) functional form for the high numeracy group. The MultiBUGS code was computed on the responses of 64 high numeracy individuals to the prospect pairs of GLM2.

The values of the empirical priors were derived from the maximum likelihood estimates of the parameters of the cumulative prospect model with the same functional form. These estimates are based on responses to GAINS1, GLM1 and GAINS2 of a different set of 57 high numeracy individuals (3, 10 and 44 respondents, respectively; see Table 15).

These parameter values are used in transformed form in the following MultiBUGS model code (12th, 15th and 18th lines highlighted) adapted from Nilsson, Rieskamp, & Wagenmakers (2011b).

```
model{
    for( j in 1 : 64 ) {
        alpha[j] <- phi(alpha.phi[j])
        gamma[j] <- exp(lgamma[j])
        delta[j] <- exp(ldelta[j])
        lambda[j] <- exp(llambda[j])
        alpha.phi[j] ~ dnorm(mu.phi.alpha, tau.phi.alpha)
        lgamma[j] ~ dnorm(lmu.gamma, ltau.gamma)
        ldelta[j] ~ dnorm(lmu.delta, ltau.delta)
        llambda[j] ~ dnorm(lmu.lambda, ltau.lambda)
    }
    mu.phi.alpha ~ dnorm(-1.3, tau.phi.alpha)
    tau.phi.alpha ~ dgamma(10, 0.1)
    sigma.phi.alpha <- 1 / sqrt(tau.phi.alpha)
    lmu.gamma ~ dnorm(-0.22, ltau.gamma)
    ltau.gamma ~ dgamma(10, 0.1)
    lsigma.gamma <- 1 / sqrt(ltau.gamma)
    lmu.delta ~ dnorm(-0.22, ltau.delta)
    ltau.delta ~ dgamma(10, 0.1)
    lsigma.delta <- 1 / sqrt(ltau.delta)
    lmu.lambda ~ dnorm(0.18, ltau.lambda)
    ltau.lambda ~ dgamma(10, 0.1)
    lsigma.lambda <- 1 / sqrt(ltau.lambda)
    mu.alpha <- phi(mu.phi.alpha)
    mu.gamma <- exp(lmu.gamma)
    mu.delta <- exp(lmu.delta)
    mu.lambda <- exp(lmu.lambda)
    for( j in 1 : 64 ) {
        for( i in 1 : 16 ) {
            v.x.a[i , j] <- pow(prospects.a[i , 1], alpha[j])
            v.y.a[i , j] <- pow(prospects.a[i , 3], alpha[j])
            v.u.a[i , j] <- pow(prospects.a[i , 5], alpha[j])
```

```
        r.x.a[i , j] <- 0
        r.y.a[i , j] <- prospects.a[i , 2]
        r.u.a[i , j] <- prospects.a[i , 2] + prospects.a[i , 4]
        w.x.a[i , j] <- pow(r.x.a[i , j], gamma[j]) / pow(pow(r.x.a[i , j], gamma[j]) + pow(1 - r.x.a[i , j], gamma[j]), 1 / gamma[j])
        w.y.a[i , j] <- pow(r.y.a[i , j], gamma[j]) / pow(pow(r.y.a[i , j], gamma[j]) + pow(1 - r.y.a[i , j], gamma[j]), 1 / gamma[j])
        w.u.a[i , j] <- pow(r.u.a[i , j], gamma[j]) / pow(pow(r.u.a[i , j], gamma[j]) + pow(1 - r.u.a[i , j], gamma[j]), 1 / gamma[j])
        d.x.a[i , j] <- w.y.a[i , j] - w.x.a[i , j]
        d.y.a[i , j] <- w.u.a[i , j] - w.y.a[i , j]
        d.u.a[i , j] <- 1 - w.u.a[i , j]
        Vf.a[i , j] <- d.x.a[i , j] * v.x.a[i , j] + d.y.a[i , j] * v.y.a[i , j] + d.u.a[i , j] * v.u.a[i , j]
}
for( i in 1 : 16 ) {
        v.x.b[i , j] <- pow(prospects.b[i , 1], alpha[j])
        v.y.b[i , j] <- pow(prospects.b[i , 3], alpha[j])
        v.u.b[i , j] <- pow(prospects.b[i , 5], alpha[j])
        r.x.b[i , j] <- 0
        r.y.b[i , j] <- prospects.b[i , 2]
        r.u.b[i , j] <- prospects.b[i , 2] + prospects.b[i , 4]
        w.x.b[i , j] <- pow(r.x.b[i , j], gamma[j]) / pow(pow(r.x.b[i , j], gamma[j]) + pow(1 - r.x.b[i , j], gamma[j]), 1 / gamma[j])
        w.y.b[i , j] <- pow(r.y.b[i , j], gamma[j]) / pow(pow(r.y.b[i , j], gamma[j]) + pow(1 - r.y.b[i , j], gamma[j]), 1 / gamma[j])
        w.u.b[i , j] <- pow(r.u.b[i , j], gamma[j]) / pow(pow(r.u.b[i , j], gamma[j]) + pow(1 - r.u.b[i , j], gamma[j]), 1 / gamma[j])
        d.x.b[i , j] <- w.y.b[i , j] - w.x.b[i , j]
        d.y.b[i , j] <- w.u.b[i , j] - w.y.b[i , j]
        d.u.b[i , j] <- 1 - w.u.b[i , j]
        Vf.b[i , j] <- d.x.b[i , j] * v.x.b[i , j] + d.y.b[i , j] * v.y.b[i , j] + d.u.b[i , j] * v.u.b[i , j]
}
for( i in 1 : 16 ) {
        binval[i , j] <- 1 / (1 + exp((( -1) / (v.x.b[i , j] - v.u.b[i , j])) * (Vf.b[i , j] - Vf.a[i , j])))
        rawdata[i , j] ~ dbern(binval[i , j])
}
for( i in 17 : 32 ) {
        v.x.a[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.a[i , 1], alpha[j])
```

```
        v.y.a[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.a[i , 3], alpha[j])
        v.u.a[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.a[i , 5], alpha[j])
        r.x.a[i , j] <- 0
        r.y.a[i , j] <- prospects.a[i , 2]
        r.u.a[i , j] <- prospects.a[i , 2] + prospects.a[i , 4]
        w.x.a[i , j] <- pow(r.x.a[i , j], delta[j]) / pow(pow(r.x.a[i , j], delta[j]) + pow(1 - r.x.a[i , j], delta[j]), 1 / delta[j])
        w.y.a[i , j] <- pow(r.y.a[i , j], delta[j]) / pow(pow(r.y.a[i , j], delta[j]) + pow(1 - r.y.a[i , j], delta[j]), 1 / delta[j])
        w.u.a[i , j] <- pow(r.u.a[i , j], delta[j]) / pow(pow(r.u.a[i , j], delta[j]) + pow(1 - r.u.a[i , j], delta[j]), 1 / delta[j])
        d.x.a[i , j] <- w.y.a[i , j] - w.x.a[i , j]
        d.y.a[i , j] <- w.u.a[i , j] - w.y.a[i , j]
        d.u.a[i , j] <- 1 - w.u.a[i , j]
        Vf.a[i , j] <- d.x.a[i , j] * v.x.a[i , j] + d.y.a[i , j] * v.y.a[i , j] + d.u.a[i , j] * v.u.a[i , j]
}
for( i in 17 : 32 ) {
        v.x.b[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.b[i , 1], alpha[j])
        v.y.b[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.b[i , 3], alpha[j])
        v.u.b[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.b[i , 5], alpha[j])
        r.x.b[i , j] <- 0
        r.y.b[i , j] <- prospects.b[i , 2]
        r.u.b[i , j] <- prospects.b[i , 2] + prospects.b[i , 4]
        w.x.b[i , j] <- pow(r.x.b[i , j], delta[j]) / pow(pow(r.x.b[i , j], delta[j]) + pow(1 - r.x.b[i , j], delta[j]), 1 / delta[j])
        w.y.b[i , j] <- pow(r.y.b[i , j], delta[j]) / pow(pow(r.y.b[i , j], delta[j]) + pow(1 - r.y.b[i , j], delta[j]), 1 / delta[j])
        w.u.b[i , j] <- pow(r.u.b[i , j], delta[j]) / pow(pow(r.u.b[i , j], delta[j]) + pow(1 - r.u.b[i , j], delta[j]), 1 / delta[j])
        d.x.b[i , j] <- w.y.b[i , j] - w.x.b[i , j]
        d.y.b[i , j] <- w.u.b[i , j] - w.y.b[i , j]
        d.u.b[i , j] <- 1 - w.u.b[i , j]
        Vf.b[i , j] <- d.x.b[i , j] * v.x.b[i , j] + d.y.b[i , j] * v.y.b[i , j] + d.u.b[i , j] * v.u.b[i , j]
}
for( i in 17 : 32 ) {
        binval[i , j] <- 1 / (1 + exp((( -1) / (v.x.b[i , j] - v.u.b[i , j])) * (Vf.b[i , j] - Vf.a[i , j])))
        rawdata[i , j] ~ dbern(binval[i , j])
}
```

```
for( i in 33 : 48 ) {
        v.x.a[i , j] <- pow(prospects.a[i , 1], alpha[j])
        v.y.a[i , j] <- pow(prospects.a[i , 3], alpha[j])
        v.u.a[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.a[i , 5], alpha[j])
        r.x.a[i , j] <- 0
        r.y.a[i , j] <- prospects.a[i , 2]
        r.u.a[i , j] <- prospects.a[i , 2] + prospects.a[i , 4]
        w.x.a[i , j] <- pow(r.x.a[i , j], gamma[j]) / pow(pow(r.x.a[i , j], gamma[j]) + pow(1 - r.x.a[i , j], gamma[j]), 1 / gamma[j])
        w.y.a[i , j] <- pow(r.y.a[i , j], gamma[j]) / pow(pow(r.y.a[i , j], gamma[j]) + pow(1 - r.y.a[i , j], gamma[j]), 1 / gamma[j])
        w.u.a[i , j] <- pow(prospects.a[i , 6], delta[j]) / pow(pow(prospects.a[i , 6], delta[j]) + pow(1 - prospects.a[i , 6], delta[j]), 1 / delta[j])
        d.x.a[i , j] <- w.y.a[i , j] - w.x.a[i , j]
        d.y.a[i , j] <- w.u.a[i , j] - w.y.a[i , j]
        d.u.a[i , j] <- w.u.a[i , j]
        Vf.a[i , j] <- d.x.a[i , j] * v.x.a[i , j] + d.y.a[i , j] * v.y.a[i , j] + d.u.a[i , j] * v.u.a[i , j]
}
for( i in 33 : 48 ) {
        v.x.b[i , j] <- pow(prospects.b[i , 1], alpha[j])
        v.y.b[i , j] <- pow(prospects.b[i , 3], alpha[j])
        v.u.b[i , j] <- (( -1) * lambda[j]) * pow(( -1) * prospects.b[i , 5], alpha[j])
        r.x.b[i , j] <- 0
        r.y.b[i , j] <- prospects.b[i , 2]
        r.u.b[i , j] <- prospects.b[i , 2] + prospects.b[i , 4]
        w.x.b[i , j] <- pow(r.x.b[i , j], gamma[j]) / pow(pow(r.x.b[i , j], gamma[j]) + pow(1 - r.x.b[i , j], gamma[j]), 1 / gamma[j])
        w.y.b[i , j] <- pow(r.y.b[i , j], gamma[j]) / pow(pow(r.y.b[i , j], gamma[j]) + pow(1 - r.y.b[i , j], gamma[j]), 1 / gamma[j])
        w.u.b[i , j] <- pow(prospects.b[i , 6], delta[j]) / pow(pow(prospects.b[i , 6], delta[j]) + pow(1 - prospects.b[i , 6], delta[j]), 1 / delta[j])
        d.x.b[i , j] <- w.y.b[i , j] - w.x.b[i , j]
        d.y.b[i , j] <- w.u.b[i , j] - w.y.b[i , j]
        d.u.b[i , j] <- w.u.b[i , j]
        Vf.b[i , j] <- d.x.b[i , j] * v.x.b[i , j] + d.y.b[i , j] * v.y.b[i , j] + d.u.b[i , j] * v.u.b[i , j]
}
for( i in 33 : 48 ) {
        binval[i , j] <- 1 / (1 + exp((( -1) / (v.x.b[i , j] - v.u.b[i , j])) * (Vf.b[i , j] - Vf.a[i , j])))
```

```
            rawdata[i , j] ~ dbern(binval[i , j])
        }
}
```

Appendix L: MCMC diagnostic plots

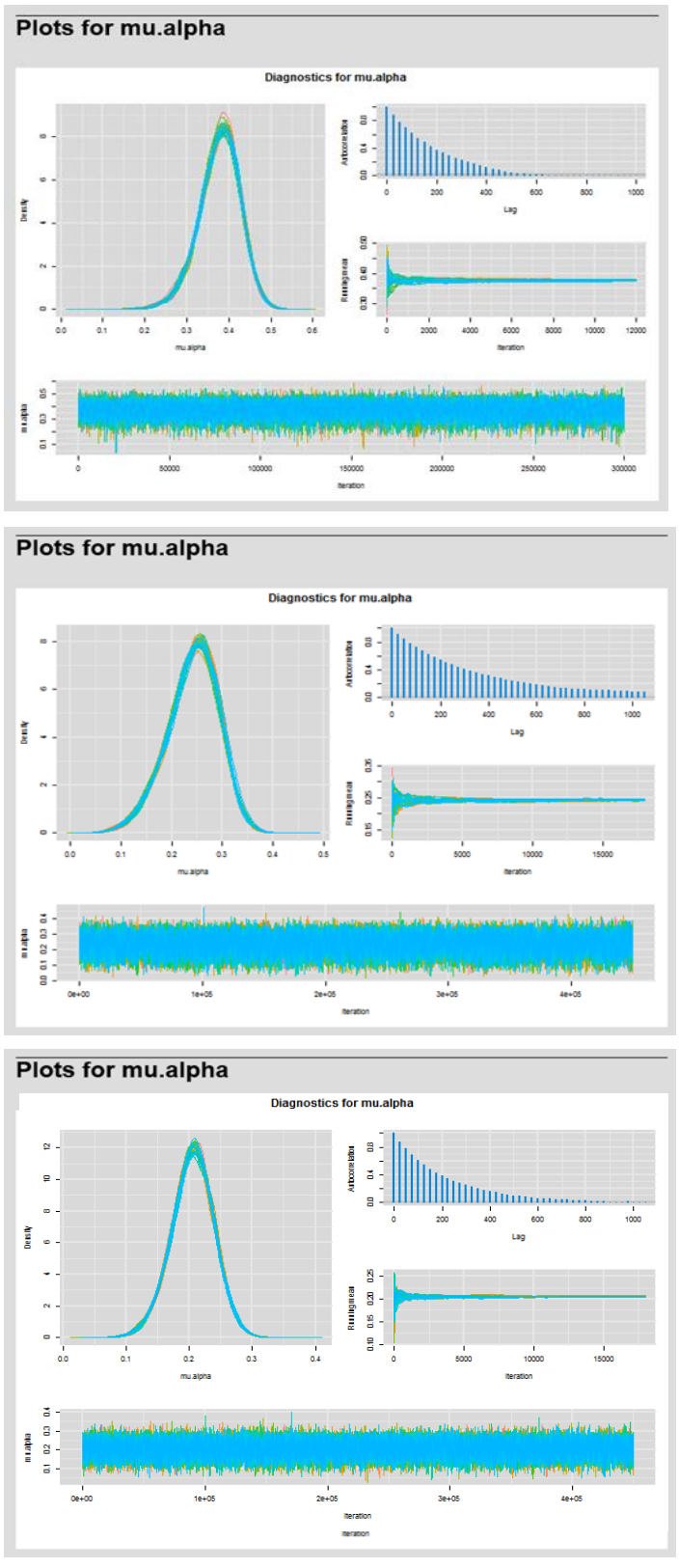


Figure L 1: MCMC diagnostic plots for expected utility theory with power function, for high, middle and low numeracy groups, respectively

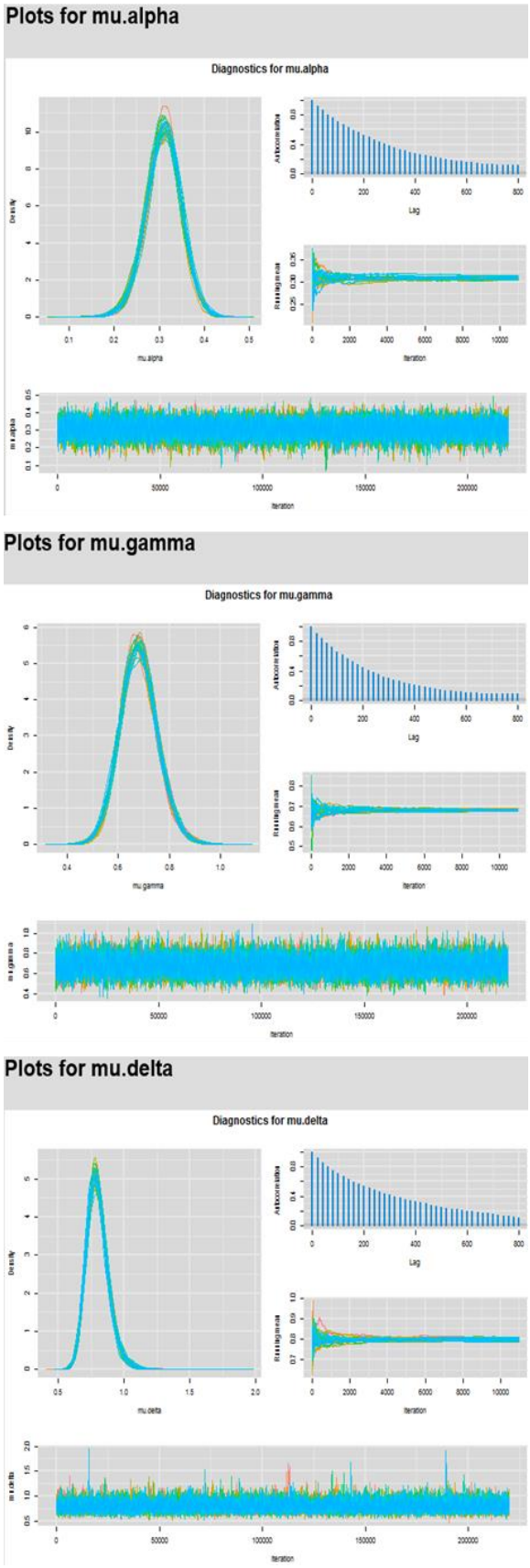


Figure L 2: High numeracy group: MCMC diagnostic plots for rank-dependence with Tversky and Kahneman probability weighting function

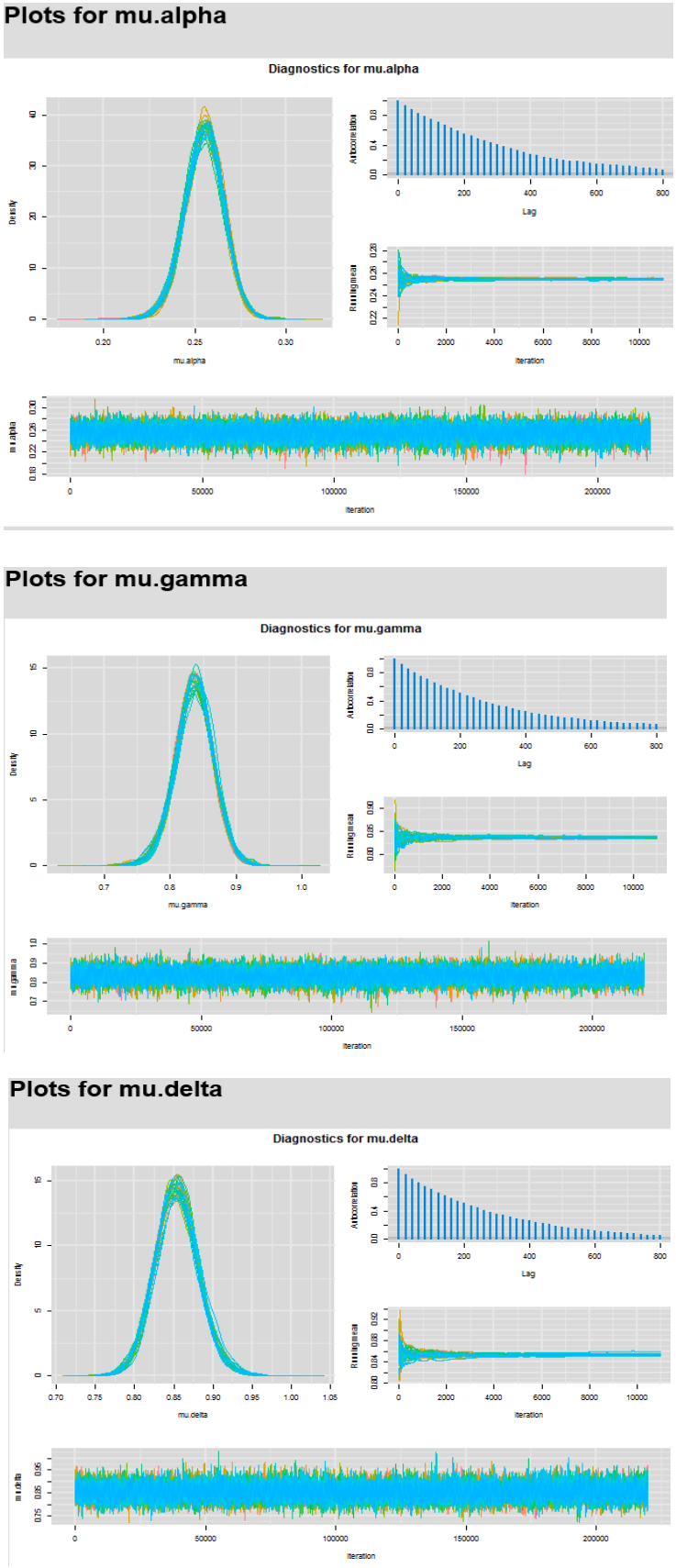


Figure L 3: Middle numeracy group: MCMC diagnostic plots for rank-dependence with Tversky and Kahneman probability weighting function

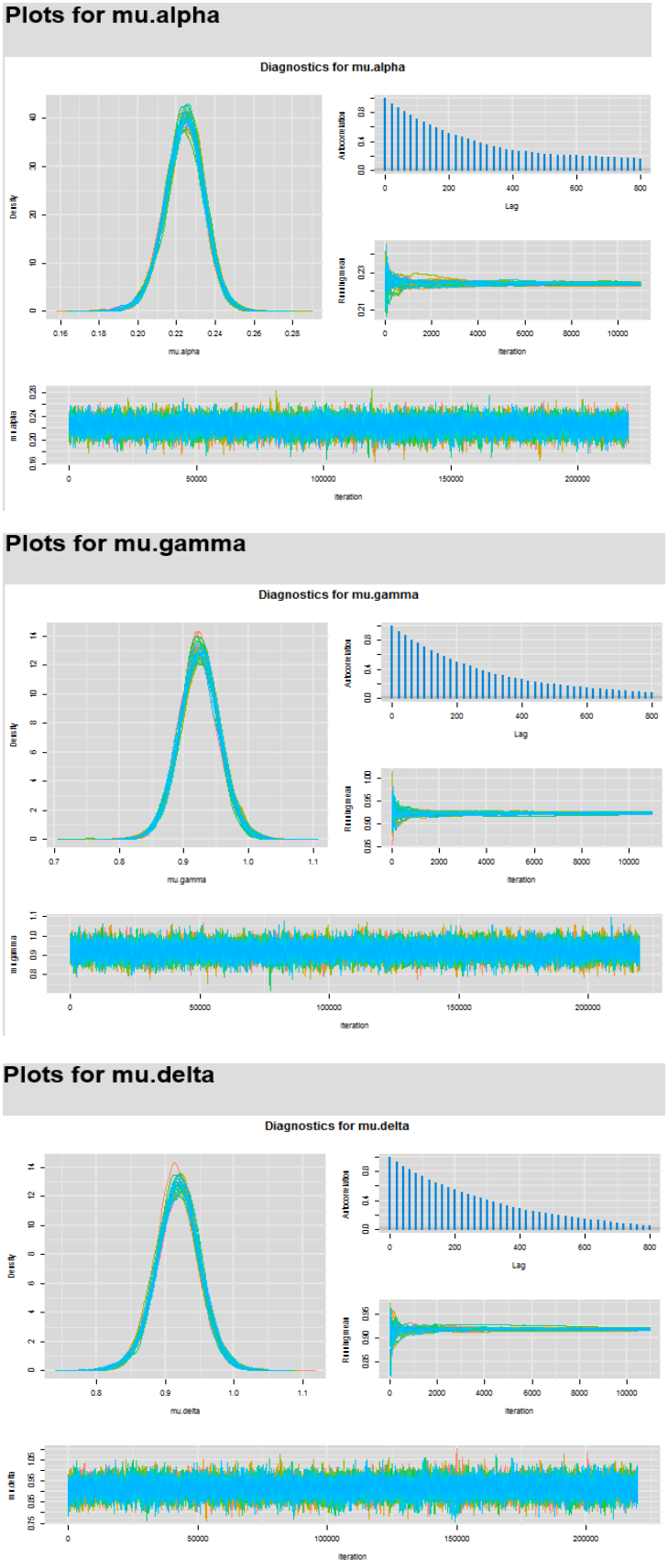


Figure L 4: Low numeracy group: MCMC diagnostic plots for rank-dependence with Tversky and Kahneman probability weighting function

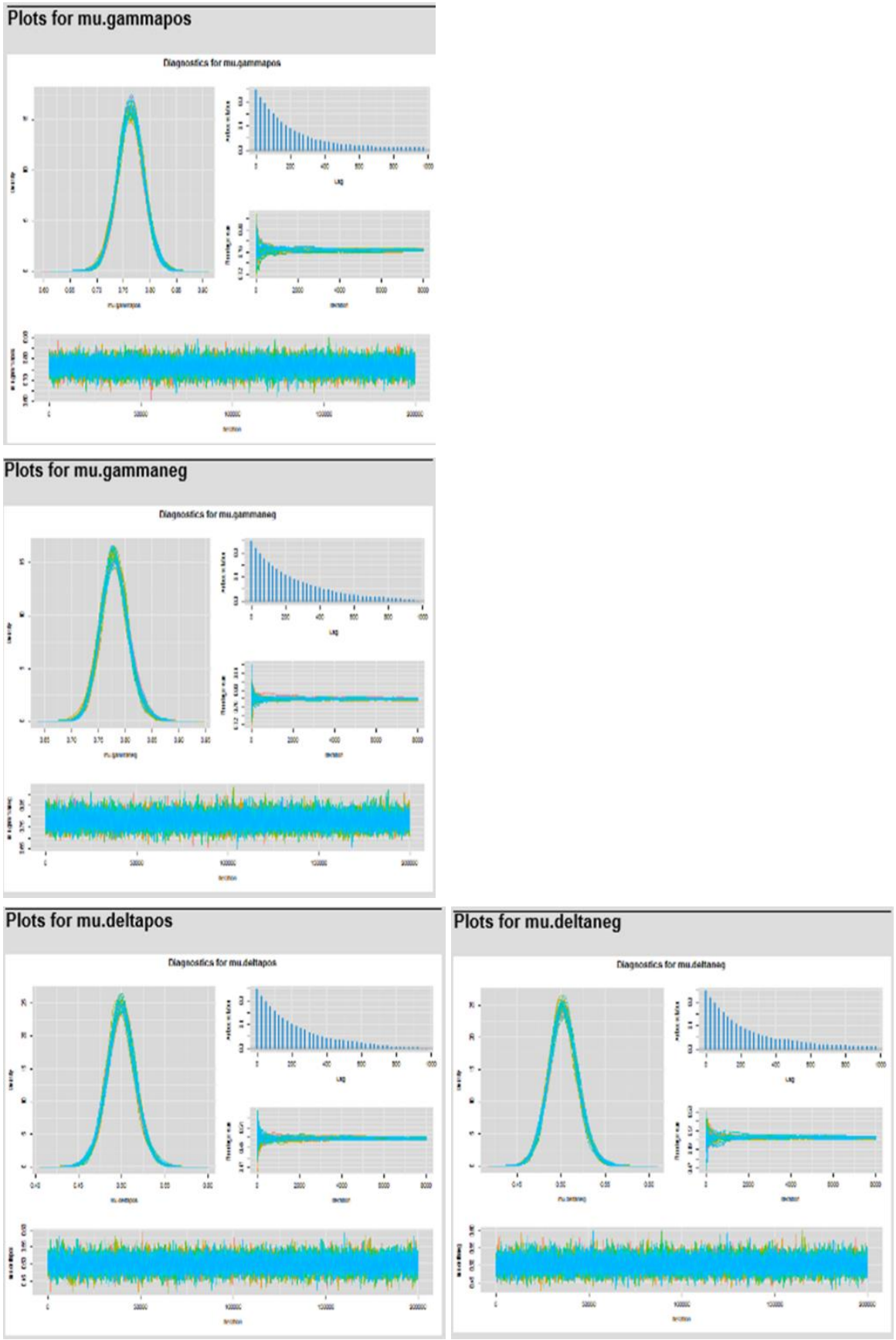


Figure L 5: High numeracy group: MCMC diagnostic plots for rank-dependence with Lattimore probability weighting function (gains and losses)

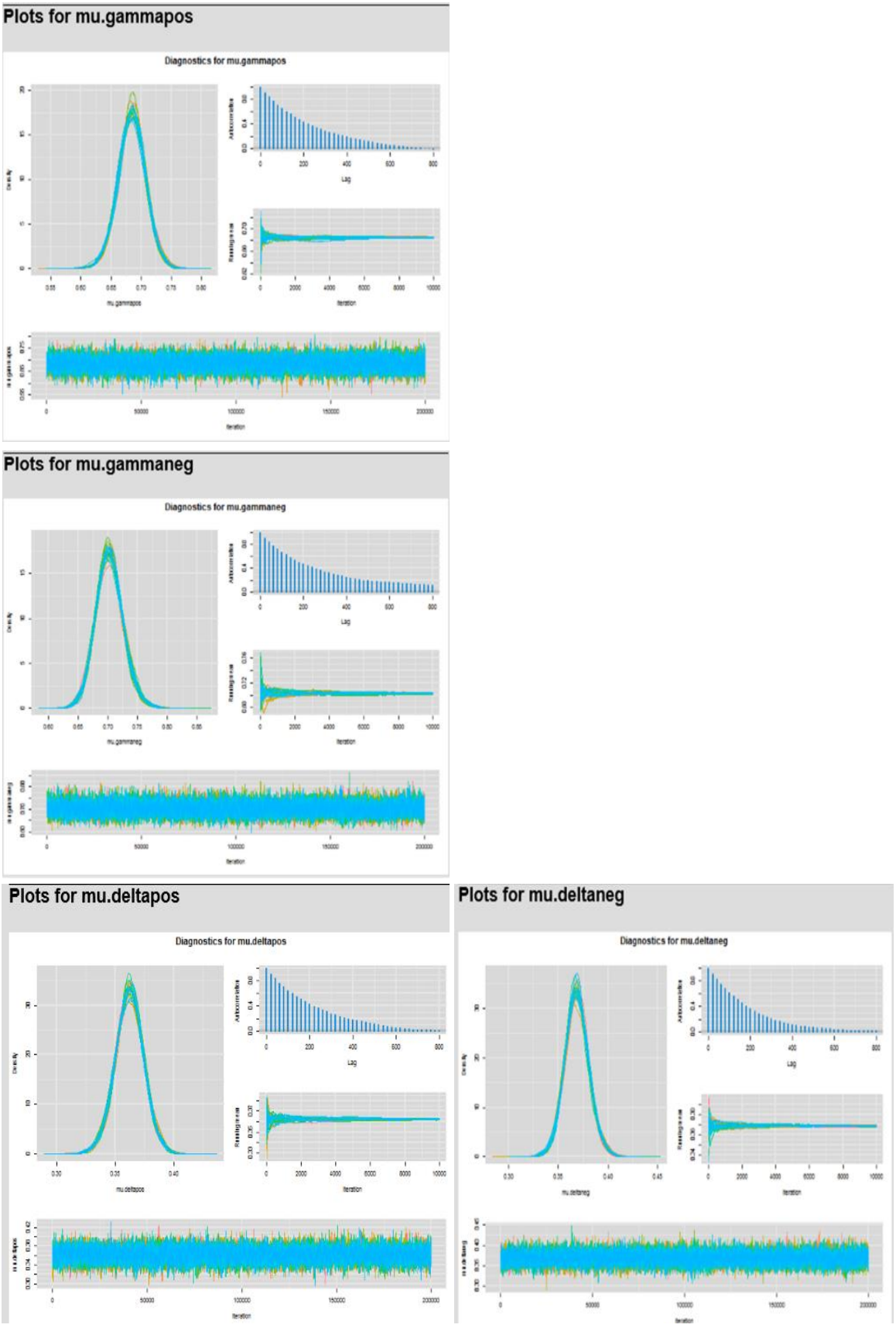


Figure L 6: Middle numeracy group: MCMC diagnostic plots for rank-dependence with Lattimore probability weighting function (gains and losses)

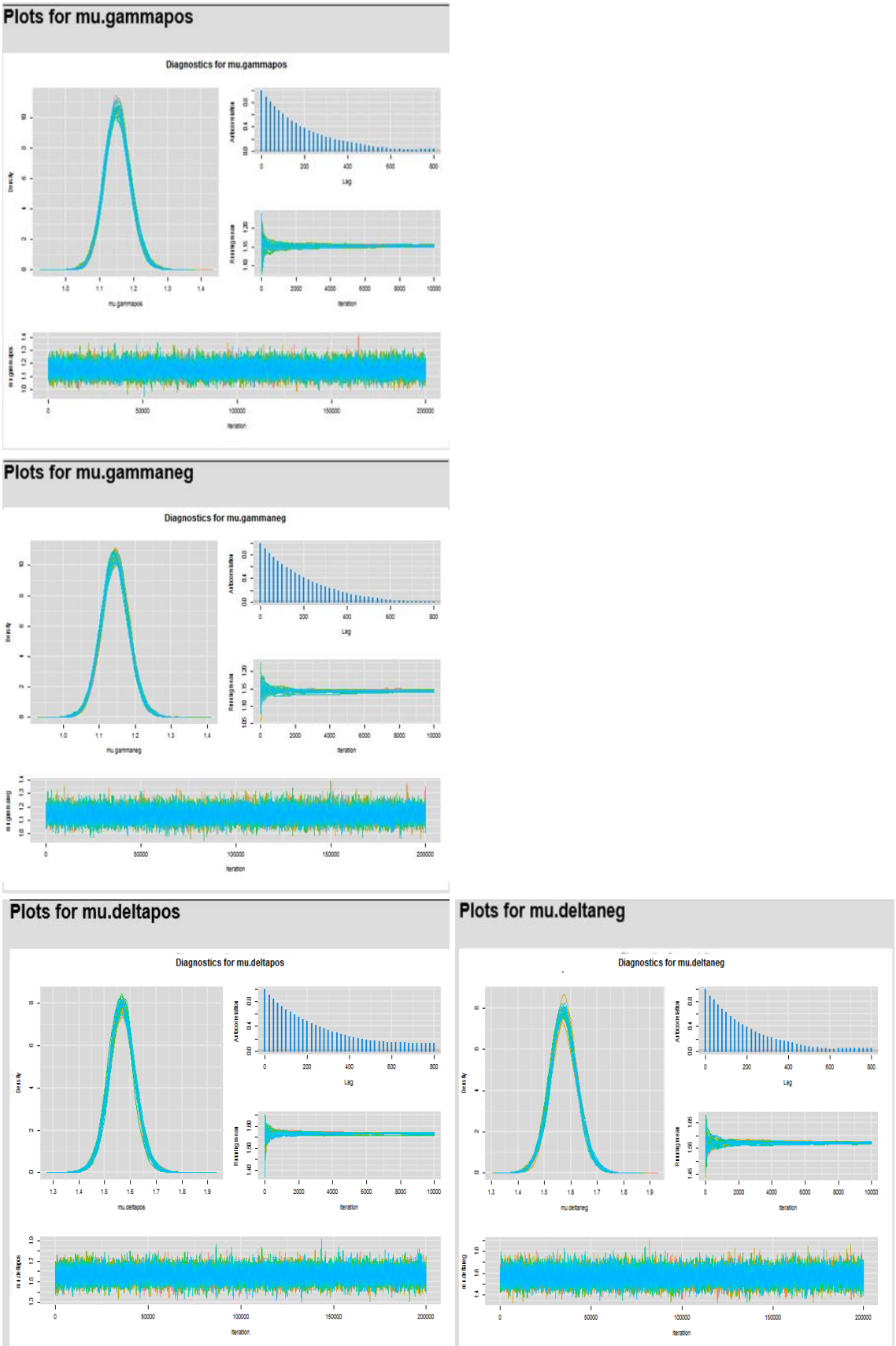
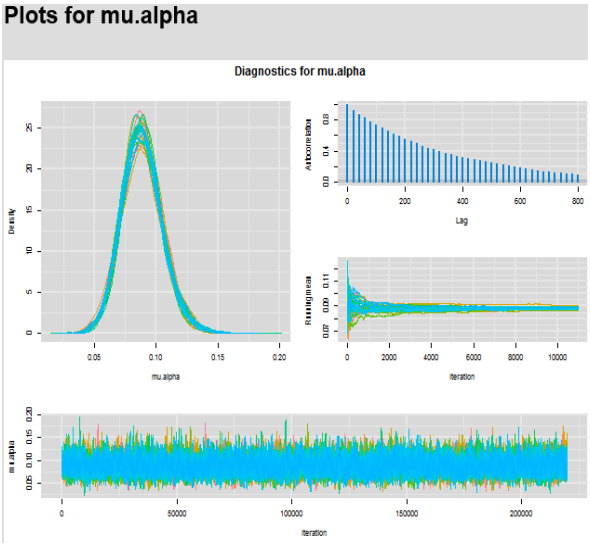
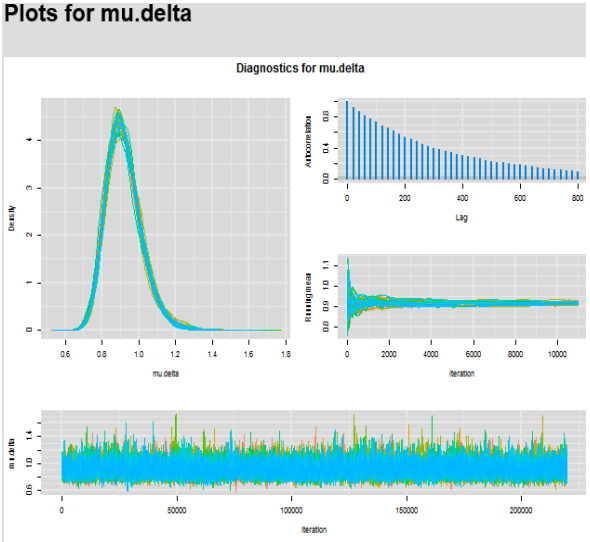
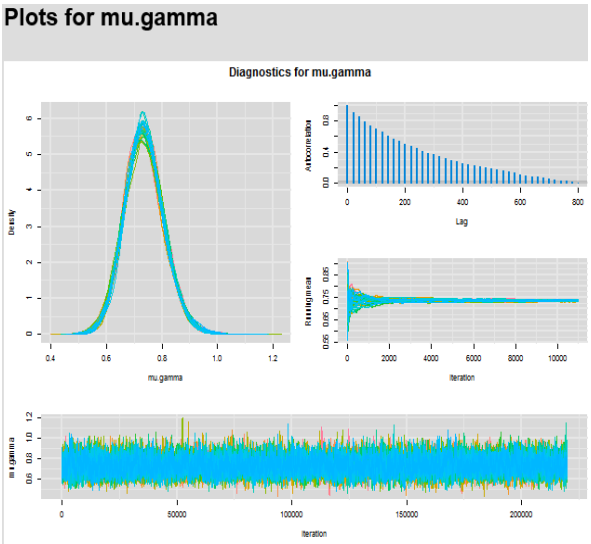


Figure L 7: Low numeracy group: MCMC diagnostic plots for rank-dependence with Lattimore probability weighting function (gains and losses)







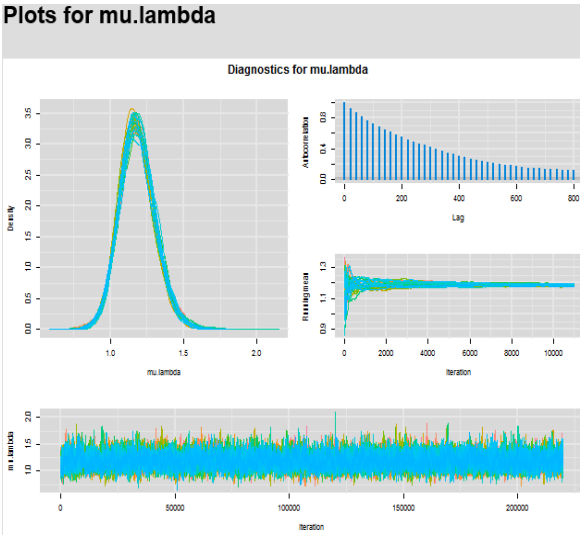


Figure L 8: High numeracy group: MCMC diagnostic plots for cumulative prospect theory with Tversky and Kahneman probability weighting function

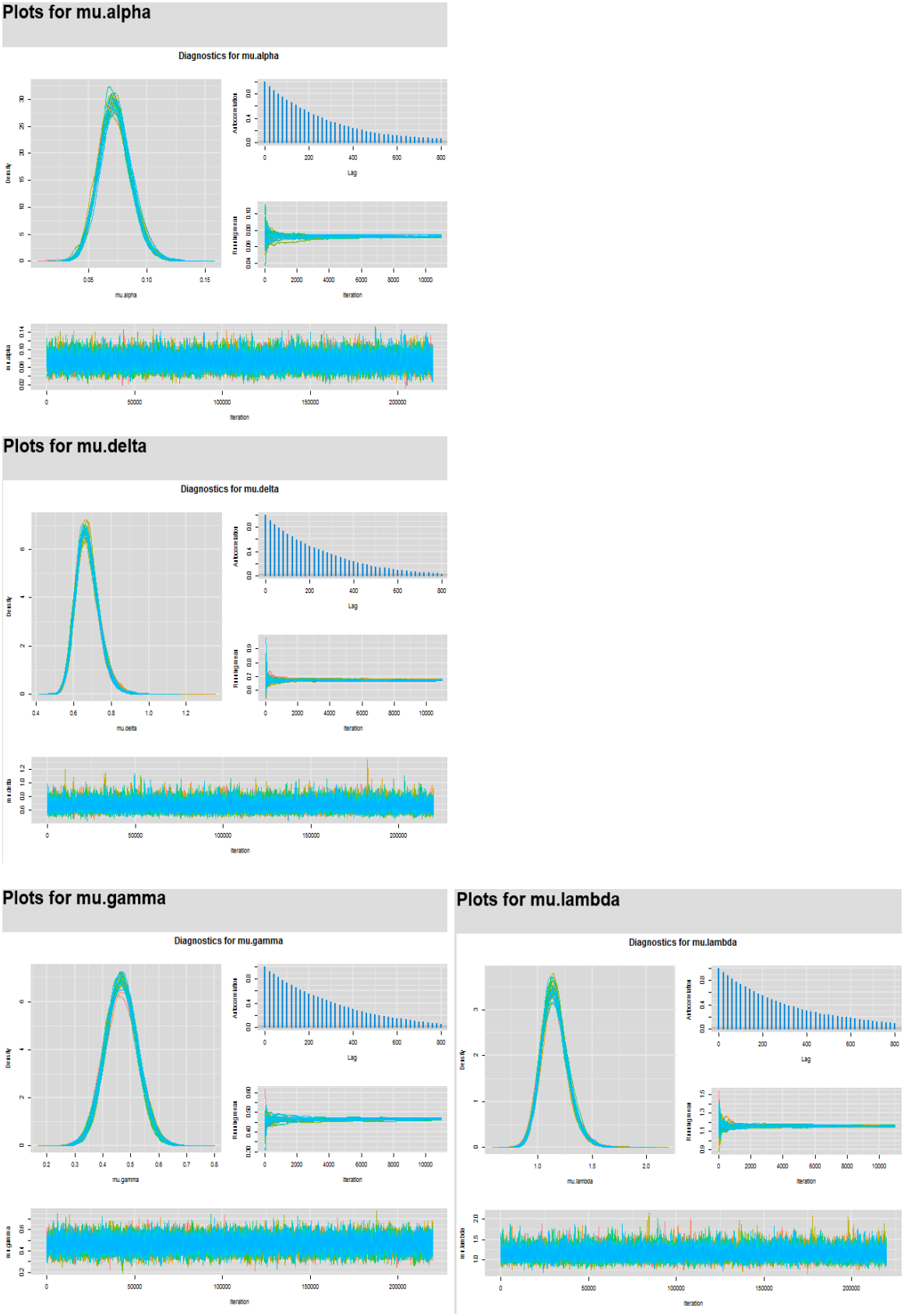


Figure L 9: Middle numeracy group: MCMC diagnostic plots for cumulative prospect theory with Tversky and Kahneman probability weighting function

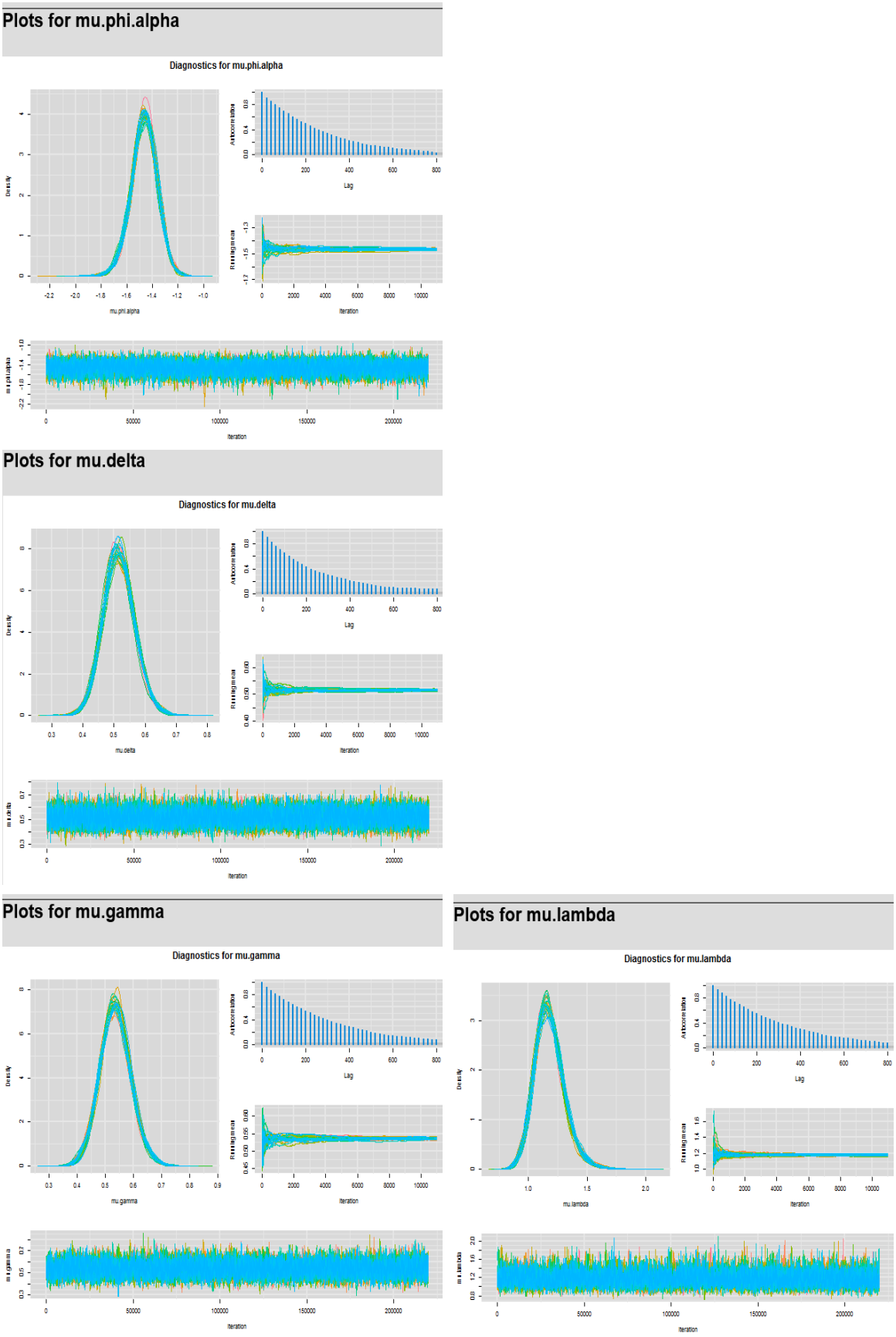


Figure L 10: Low numeracy group: MCMC diagnostic plots for cumulative prospect theory with Tversky and Kahneman probability weighting function

Appendix M: Validity checks of models' in-sample estimates and out-of-sample predictions

Table M 1: Validity check of expected value predictions of GAINS2 and GLM2 responses

Responses to GAINS2					
	Context: R200 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	44.9%	6.4%	32.2%	57.5%	44
Middle numeracy	42.3%	6.0%	30.5%	54.2%	50
Low numeracy	43.2%	6.0%	31.3%	55.1%	50
Numeracy group effect	$F(2,141) = 0.370$, $p = .692$				
η_p^2	<.01				
	Context: R300 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	40.9%	7.4%	26.3%	55.6%	44
Middle numeracy	39.9%	7.0%	26.1%	53.7%	50
Low numeracy	38.8%	7.0%	25.1%	52.6%	50
Numeracy group effect	$F(2,141) = 0.823$, $p = .441$				
η_p^2	0.011				

Responses to GLM2					
	Frame: Gains				
Consistency means	mean	s. e.	95% CI		n
High numeracy	52.5%	1.7%	49.1%	55.9%	64
Middle numeracy	45.1%	1.7%	41.7%	48.4%	65
Low numeracy	45.9%	1.7%	42.5%	49.3%	65
Numeracy group effect	$F(2,191) = 5.575$, $p = .004$				
η_p^2	0.055				
	Frame: Losses				
Consistency means	mean	s. e.	95% CI		n
High numeracy	68.2%	1.5%	65.2%	71.1%	64
Middle numeracy	61.1%	1.5%	58.2%	64.0%	65
Low numeracy	56.0%	1.5%	53.1%	58.9%	65
Numeracy group effect	$F(2,191) = 16.826$, $p < .001$				
η_p^2	0.15				
	Frame: Mixed				
Consistency means	mean	s. e.	95% CI		n
High numeracy	60.2%	1.9%	56.4%	63.9%	64
Middle numeracy	55.4%	1.9%	51.6%	59.1%	65
Low numeracy	50.3%	1.9%	46.6%	54.1%	65
Numeracy group effect	$F(2,191) = 6.691$, $p = .002$				
η_p^2	0.065				

Table M 2: Validity check of expected utility in-sample estimates and out-of-sample predictions

Expected utility: Power function: Gain-frame					
	In-sample responses to GAINS2				
	Expected utility				
	Context 1: R200 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	67.4%	6.9%	53.8%	80.9%	44
Middle numeracy	66.5%	6.4%	53.8%	79.2%	50
Low numeracy	63.0%	6.4%	50.2%	75.7%	50
Significance					
Numeracy group effect	$F(2,141) = 0.568$, $p = .568$				
η_p^2	<.01				
	Context 2: R300 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	58.1%	4.9%	48.4%	67.8%	44
Middle numeracy	59.4%	4.6%	50.3%	68.5%	50
Low numeracy	61.2%	4.6%	52.1%	70.2%	50
Significance					
Numeracy group effect	$F(2,141) = 0.773$, $p = .464$				
η_p^2	.011				
Significance					
Context*numeracy effect	$F(2,141) = 3.683$, $p = .028$ (Huynh-Feldt)				
η_p^2	.050				
Context effect	$F(1,141) = 27.561$, $p < .001$				
η_p^2	.164				
	Out-of-sample: responses to GLM2 gain frame*				
	Context 1: R200 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	65.4%	2.2%	61.1%	69.6%	64
Middle numeracy	68.8%	1.3%	61.2%	71.4%	170
Low numeracy	67.4%	2.2%	63.1%	71.6%	65
Significance					
Numeracy group effect	$F(2,296) = 0.390$, $p = .677$				
η_p^2	<.01				

***** Excluding results of triangle with steepest indifference curve gradient

Table M 3: Validity check of rank-dependent utility in-sample estimates in gain frame

Rank-dependence: Gain frame*										
In-sample responses to GLM2* gain frame										
	Tversky and Kahneman (1992)					Lattimore et al. (1992)				
Consistency means	mean	s. e.	95% CI		n	mean	s. e.	95% CI		n
High numeracy	67.8%	2.3%	63.4%	72.3%	64	59.9%	1.6%	56.7%	63.0%	64
Middle numeracy	70.1%	2.2%	65.7%	74.5%	65	56.3%	1.6%	53.3%	59.4%	65
Low numeracy	67.4%	2.2%	63.0%	71.8%	65	52.6%	1.6%	49.5%	55.7%	65
Significance										
Numeracy group effect	$F(2,191) = 0.438$, $p = .646$					$F(2,191) = 5.320$, $p = .006$				
η_p^2	<.01					0.053				

* Excluding results of triangle with steepest indifference curve gradient

Table M 4: Validity check of rank-dependent utility out-of-sample predictions in gain frame

Rank-dependence: Gain frame										
Out-of-sample responses to GAINS2										
	Tversky and Kahneman (1992)					Lattimore et al. (1992)				
	Context 1: R200 maximum									
Consistency means	mean	s. e.	95% CI		n	mean	s. e.	95% CI		n
High numeracy	69.0%	2.9%	63.3%	74.7%	44	54.7%	2.1%	50.7%	58.8%	44
Middle numeracy	64.9%	1.6%	61.7%	68.0%	146	53.4%	1.1%	51.1%	55.6%	146
Low numeracy	63.6%	2.7%	58.2%	68.9%	50	53.4%	1.9%	49.6%	57.2%	50
Significance										
Numeracy group effect	$F(2,237) = 1.042$, , $p = .354$					$F(2,237) = 0.178$, $p = .836$				
η_p^2	<.01					<.01				
	Context 2: R300 maximum									
Consistency means	mean	s. e.	95% CI		n	mean	s. e.	95% CI		n
High numeracy	60.1%	2.6%	54.9%	65.3%	44	45.3%	2.3%	40.8%	49.9%	44
Middle numeracy	60.9%	1.4%	58.0%	63.7%	146	47.6%	1.3%	45.1%	50.1%	146
Low numeracy	62.4%	2.5%	57.5%	67.2%	50	52.4%	2.2%	48.2%	56.7%	50
Significance										
Numeracy group effect	$F(2,237) = 0.221$, $p = .802$					$F(2,237) = 2.786$, $p = .063$				
η_p^2	<.01					.023				
Significance										
Context effect	$F(1,237) = 6.349$, $p = .012$					$F(1,237) = 2.092$, $p = .149$				
η_p^2	.026					<.01				

Table M 5: Validity check of rank-dependent utility in-sample estimates in loss frame

Rank-dependence: Loss frame **In-sample: responses to GLM2 loss frame**										
	Tversky and Kahneman (1992)					Lattimore et al. (1992)				
Consistency means	mean	s. e.	95% CI		n	mean	s. e.	95% CI		n
High numeracy	48.2%	2.7%	43.0%	53.5%	64	67.2%	1.5%	64.2%	70.2%	64
Middle numeracy	41.9%	2.7%	36.6%	47.1%	65	54.9%	1.5%	51.9%	57.8%	65
Low numeracy	37.2%	2.7%	32.0%	42.5%	65	46.1%	1.5%	43.1%	49.1%	65
Significance										
Numeracy group effect	$F(2,191) = 4.271$, $p = .015$					$F(2,191) = 49.221$, $p < .001$				
η_p^2	.043					.340				

Table M 6: Validity check of cumulative prospect theory in-sample estimates and out-of-sample predictions in gain frame

Cumulative prospect theory: Gain frame					
In-sample responses to GLM2 (gain frame)*					
	Tversky and Kahneman (1992)				
Consistency means	mean	s. e.	95% CI		n
High numeracy	62.5%	2.4%	57.8%	67.2%	64
Middle numeracy	70.1%	2.3%	65.5%	74.7%	65
Low numeracy	67.4%	2.3%	62.8%	72.0%	65
Significance					
Numeracy group effect	$F(2,191) = 2.699$, $p = .07$				
η_p^2	.027				
Out-of-sample responses to GAINS2					
	Tversky and Kahneman (1992)				
	Context 1: R200 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	68.2%	3.0%	62.3%	74.1%	44
Middle numeracy	64.9%	1.6%	61.6%	68.1%	146
Low numeracy	63.6%	2.8%	58.1%	69.1%	50
Significance					
Numeracy group effect	$F(2,237) = 0.697$, $p = .499$				
η_p^2	<.01				
	Context 2: R300 maximum				
Consistency means	mean	s. e.	95% CI		n
High numeracy	56.9%	3.5%	50.0%	63.7%	44
Middle numeracy	61.5%	1.9%	57.7%	65.2%	146
Low numeracy	60.6%	3.3%	54.1%	67.0%	50
Significance					
Numeracy group effect	$F(2,237) = 0.675$, $p = .510$				
η_p^2	<.01				
Significance					
Context effect	$F(1,237) = 27.442$, $p < .001$				
η_p^2	.104				

* Excluding results of triangle with steepest indifference curve gradient

Table M 7: Validity check of cumulative prospect theory in-sample estimates in loss frame

Cumulative prospect theory: Loss frame					
In-sample responses to GLM2 loss frame					
	Tversky and Kahneman (1992)				
Consistency means	mean	s. e.	95% CI		n
High numeracy	43.6%	2.7%	38.2%	48.9%	64
Middle numeracy	41.9%	2.7%	36.5%	47.2%	65
Low numeracy	37.2%	2.7%	31.9%	42.6%	65
Significance					
Numeracy group effect	$F(2,191) = 1.444$, $p = .238$				
η_p^2	.015				

Table M 8: Validity check of cumulative prospect theory in-sample estimates in mixed frame

Cumulative prospect theory: Mixed frame					
In-sample responses to GLM2 (mixed frame)					
	Tversky and Kahneman (1992)				
Consistency means	mean	s. e.	95% CI		n
High numeracy	46.5%	2.7%	41.3%	51.7%	64
Middle numeracy	44.4%	2.6%	39.2%	49.6%	65
Low numeracy	41.8%	2.6%	36.6%	47.0%	65
Significance					
Numeracy group effect	$F(2,191) = 0.793$, $p = .454$				
η_p^2	<.01				