A Conceptual Framework for the Use of the Generalised Latent Variable Model in Psychological Research

Trisha Marie Nowland

B. Economics

B. Psychology (Hons.)

M. Research

Department of Psychology

Faculty of Human Sciences

Macquarie University, Sydney, Australia

This thesis is presented for the degree of Doctor of Philosophy (PhD)

April, 2019

Table of Contents

Summary

Through the 20th century to today, latent variable modelling remains predominant as a psychometric technique employed in psychology research practices. Through time it has also become increasingly complex, and dependent upon pre-specified parameters and assumptions coded in computer software, which may remain unreviewed by a researcher. With these trends, there has been a loss of focus on important mathematical and statistical questions for the latent variable model. Further, we remain without a unified solution for presentation of extra-statistical assumptions necessary for use of the model, where it remains the case that decisions researchers must make when using the latent variable model remain unrecorded and perhaps, unreconciled to anything more than researcher intuition. The primary aims of this thesis were to firstly, to document a cognitive-historical analysis of the methodological problems for the generalised latent variable model; and secondly, to use this analysis as a guide to develop a conceptual framework for the use of the latent variable model in psychological research. Contemporaneous developments in philosophy of mathematics and logic were utilised to ensure the proposed structure for the framework was levelled on grounds that facilitate logical consistency and unification, given that diverse methodology accompanies the assembly of data, variables, relations and models of phenomena, in latent variable modelling. Such unification was shown to facilitate assessment of quality uncertainty in psychology study outcomes that make use of latent variable models; facilitating for perhaps the first-time true study comparability and replication assessment, as well as facilitating robust integration in meta-analyses. The principal finding of this thesis was the essential role for a realist constructivist conceptual framework for research projects that make use of the generalised latent variable model, in psychological research.

Statement of Candidature

I, Trisha Nowland, certify that the work in this thesis entitled "A Conceptual Framework for the Use of the Generalised Latent Variable Model in Psychological Research" has not been previously submitted for a higher degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any sources of information used throughout this thesis are acknowledged, including any help or assistance that I have received in my work and preparation of this thesis.

Signed:

Trisha Nowland (Student ID: 30308283)

April, 2019

Acknowledgments

It is right to say that the origins of the work that follows predate its beginnings as a thesis, founded in conversations and arguments around the dinner table of my family, a birthing ground for testing of logic and belief. This thesis is dedicated to my brothers Ian and Jeff, my father Robert who died during the time this thesis came to life, and my mother Lesley who has borne with fortitude our zealous discussions, and who has served to model and provide continued grace and unwavering support.

Without the ongoing support and gifted guidance of my principal supervisor Associate Professor Simon Boag, there would be no thesis which sits before you. It is a privilege to find a mind that continues to step beyond your own and turn with you to investigate what might be in the gap; words do not do justice to the gratitude called forth. With great thanks also to Associate Supervisor Dr Alissa Beath, and for stewardship from Dr Colin Klein, Dr Fiona Hibberd, and Dr Josh McGrane through the preparation of the thesis. Also extending gratitude to Josh and Dr Therese Hopfenbeck, and all the team at the Oxford University Centre for Educational Assessment, for the generous welcome and provision of working space at the Centre for my three visits during the preparation of this thesis.

Finally, I would like to give thanks to the Dharug community and friends at the Aboriginal Drop-In Centre at Aboriginal Catholic Services, in Emerton, particularly to Linda McDonald, and Linda Davis. While the Centre served as the backdrop for my work originally as a counsellor and then as psychologist over the previous years, community here serve as a support network, in ways that can never be quantified, with lessons still unfolding in real time. May this thesis serve in some way to honour those past, present, and future, on whose land, we dwell.

Chapter 1: The Need for a Conceptual Framework for the use of the Generalised Latent Variable Model in Psychological Research

1. Introduction

The latent variable model is cited in psychometric literature as the most utilised of all statistical techniques in present-day psychological research that involves psychometric modelling (Mulaik, 2010). It is also cited as one of psychology's most successful exports to other fields (Maydeu-Olivares & McArdle, 2005). Yet, the latent variable model has introduced challenges for psychology researchers since its first development in the work of Spearman (Thomson, 1916; Wilson, 1928; Wolfe, 1940; Campbell, 1982; Maraun, 1996; Michell, 2008; Markus, 2008). Spearman (1904) presented his ideas about latent variables in factor theory, which he presented as a mathematical technique that aimed to resolve substantive problems in defining the concept of intelligence. With present-day advances in technology and software, researchers using what is now described as the generalized latent variable model (GLVM: Skrondal & Rabe-Hesketh, 2004) find themselves with a considerable array of decisions to make about the model in connection to the empirical situation of their research, and assumptions to check that are statistical rather than mathematical, in nature (Markus & Borsboom, 2013).

The term GLVM is chosen here because it embraces the diverse developments made from the time of the introduction of latent variables into psychological research practices. Latent variables are variables for which the variate values are considered to be not directly available to individuals who are interested in some aspect of what they represent. These variables have otherwise been described in the literature as "hidden" (Skrondal & Rabe-Hesketh, 2004, p. 1) or "unobservable" (Borsboom, 2008, p. 31). Latent variables are differentiated from other variables included in the latent variable model, often described as manifest variables, for which data is available, or as is described in the literature, observable (Bollen, 2002; Borsboom, 2005, 2008). Specifically, a working definition for the GLVM utilised through this thesis includes elements for: i) a response model, which includes the relationship of dependency between the latent and manifest variables; ii) the structural model, which also includes parameters for directional relationships between the variables; and iii) a model for the distribution of errors associated with the variables in the model (Skrondal & Rabe-Hesketh, 2004, p. 95). In this thesis, we begin from the earliest formulations of what is now known as the GLVM, and explore two specific versions of the generalised model, in subsequent chapters, developed in the work of McDonald (1999) and Bartholomew, Knott, and Moustaki (2011).

The aim of this present work is to make clear the need for a conceptual framework for any use of the GLVM in psychological research, given that use of this model deals in a complex combination of empirical, statistical, and mathematical information. A secondary aim is to present a working version of what this conceptual framework looks like in an applied setting. To clarify the need for this framework, this thesis examines the history of the development of the GLVM from the time of Spearman's original proposal of factor theory in 1904, with an eye for the logical structure of the model, when used in psychometric practices. It will be shown that as the latent variable model was developed and generalised to the format we recognise today, the number and complexity of assumptions involved and decisions required in an application of the model increased substantially. There remains at present no standardised format for the presentation of these decisions and assumptions. Even though any one decision related to the model may seem quite straightforward, the complex interaction of mathematical, statistical, and empirical information in GLVM processes means that tracing these decisions for their influence on research outcomes is a complicated undertaking. It then becomes impossible to interpret the effects of the decision or assumption, considering the relatively simplified reporting of latent variable modelling analysis outcomes seen in peer-reviewed journal literature. The conceptual framework aims to specifically target, and address, this concern.

This thesis presents the background, argument, and proposal for a conceptual framework for the use of the GLVM, as well as making a case for a realist constructivist approach to the research situations that support psychometric investigations. The proposal is realist, insofar as it is grounded in the geo-historical context of the research project, and constructivist, in that principles of mathematical constructivism are adopted as foundational to the representation of the research situation and particularly, the psychological phenomena that is of interest. Such an approach yields a theory of expected relations for the research project, where key aspects of the project in terms of adopted principles, assumptions, and decisions are captured in what is called a research element. The links between these research elements are charted via a network systems approach that facilitates logical representation of the relational structure for the research project, in a way shown to be amenable to inference from the best systematization (Rescher, 1979, 2016). The research elements or representations of components of the research situation, and analysis tools that sit in this relational structure are guided by principles of set theoretical formulation for models, following Suppes (1969, 2002). Set theory, almost uniquely from the historical mathematical corpus, offers resources for logical unification for psychometric practices, to the degree that set theory offers the whole of mathematics indispensable resources for unification (Ferreirós, 2016; Maddy, 1990, 2011; c.f. Davies, 2003, p. 113). Set theory is appropriate for the purposes of a conceptual framework for research projects making use of the GLVM, because set theory already accommodates inclusion of empirical and mathematical considerations (Ferreriós, 2016; Maddy, 2011) which are intrinsic to this psychometric model. This thesis explores an approach informed but not limited by axiomatization of research elements (see Suppes, 1967; Costa & Chuaqui, 1988), in a way that supports evaluation of quality uncertainty regarding research report outcomes (Vazire, 2017). Axiomatisation is the most effective means by which series of statements are interpreted together, providing the researcher and the research reader with an understanding of the set of constraints within which research claims are made (Suppes, 2002).

Such an approach has a good deal in common with recent proposals seen in the psychology literature following the outcomes of the Reproducibility Project regarding pre-registration of psychology research studies (Open Science Collaboration, 2012, 2015). In the representation of the research situation in the conceptual framework that is novel to this thesis, researchers record the commitments they hold regarding a number of largely statistical and qualitative assumptions about the world and the phenomenon of interest. They also record relevant theory and as much information as is possible about the context within which the research project takes place. When compared with other pre-registration processes (see for example, van't Veer & Giner-Sorolla, 2016; Cybulski, Mayo-Wilson, & Grant, 2016), the difference is the grounding of this conceptual framework in logical consistency. Logical consistency then serves as a key means by which the contribution to science made by the project can be evaluated (Rescher, 1979). Such an approach to a conceptual framework is also standardized in a way that facilitates both cross-study comparison and outcome integration. Various pre-registration processes themselves can be accommodated within this conceptual framework, which ultimately makes possible true meta-analysis that is reconciled across studies in qualitative and quantitative terms. The aim for this framework is ultimately to be general enough to support any research endeavour in psychology involving the GLVM, and precise enough to yield some solutions to key problems associated with the latent variable model, as they are described in this thesis.

The conceptual framework described above is increasingly essential since researchers now must justify the specific choice of model from the array available within the GLVM framework, as well as competing alternatives (see Schmittman et al., 2013;

Epskamp, Rhemtulla, & Borsboom, 2017). These model types represent new analysis tools for psychological researchers; involving choices from an array of possible model forms in ways never encountered in the field (see Borsboom & Cramer, 2013; MacCallum, Wegener, Uchino, & Fabrigar, 1993). The aims of latent variable modelling are also changing. When Spearman developed his factor theory that underlies the latent variable model at the turn of the 20th century, he did it to address a problem of concept definition – the concept of intelligence (see Spearman, 1904; Horn & McArdle, 2007). Increasingly today, researchers use the GLVM in statistical predictive analyses, which are many steps removed from Spearman's conceptual analyses. Instead, researchers choose between distinct model types to optimize for predictive accuracy, for example (see Schmitt, 2011; Yarkoni & Westfall, 2017). Each model potentially has disparate sets of assumptions say, for example, about the structure of data for the variables in the models (Thompson, 2004; Schmitt, 2011). Yet no standardised reconciliation is demanded for choices made for these in the reporting of study outcomes across peer-reviewed literature, even when two different models are used in justification for an interpretation of psychological phenomena, as is often the case for example with Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) under the GLVM (Schmitt, 2011). These two techniques are for continuous scale structure data, yet involve different assumptions about the relationship between the latent variable and other variables within the model, as well as error terms (Thompson, 2004; Schmitt, 2011).

Another recent example of this issue of conflicting assumptions is seen in recent efforts to distinguish patterns of symptoms in psychopathology, where proposals have been made for different overarching factors that subsume DSM diagnoses under hierarchical latent structures (see for example, Krueger & Markon, 2006; Wiecki, Poland, & Frank, 2015). Each different hierarchical structure endorses distinct assumptions about the relations between the latent variables and other variables in the model, yet there are few or no requirements in peer-reviewed reporting processes to account in detail for the choices made regarding a preferred structure for the relationships. These proposals as these have distinct forms that may include for example multiple latent variables, or a single overarching latent variable presumed responsible for any occurrence of psychopathology (see Krueger & Markon, 2006; Krueger et al., 2018; see also example in Chapter 8). The implications of these different structures are significant, in that they may be used to justify distinct forms of therapeutic treatment and subsequent research programs, given that they form part of an evidence base for some set of symptoms in mental health diagnoses (see for example Schretlen et al., 2013; Harvey et al., 2016).

The examples above suggest that techniques developed by Spearman to investigate definitional problems for the concept of intelligence now are used for completely different purposes from those for which they were designed. This means that the sets of assumptions relevant to the original models and structures, particularly in reference to how the model connects to the world, will be substantially different from the assumptions that are relevant to present day uses of the model. Furthermore, in today's psychometric practices, the competing perspectives reflected in distinct mathematical and statistical models need to be evaluated for consistency as substantive scientific claims. The point of the conceptual framework developed in this thesis is to facilitate in both a general and yet particular way evaluation of logical constraints in a way that supports meaningful interpretation, of research outcomes. This applies for any single project, but also facilitates robust between-project comparison and ultimately, integration, where meta-analyses become of interest to research communities.

In this introductory chapter, the research domain for the present project is set via definition of terms that are relevant to material presented throughout the rest of the thesis. A brief definition of the latent variable model is first presented, with short summary-form presentations of major paradigms of influence in the development of this conceptual framework, including cognitive-historical analysis (Nerssessian, 1995, 2008), set-theoretical model-based reasoning for empirical practices (Suppes, 1967, 2002), and cognitive systematization (Rescher, 1979, 2016). Chapter summaries serve to articulate the landscape over which the background, proposal, and argument for the conceptual framework both journeys, and is created. This leads into a cognitive-historical exploration of the key developers in the GLVM space, from the beginning of Spearman's (1904) factor theory, in Chapter 2.

1.1 What is a latent variable model?

The latent variable model as presently used in psychometric practice has its roots in Spearman's (1904) first characterisation of factor analysis. In "*'General Intelligence': objectively determined and measured",* Spearman sought to present evidence that there must be a common factor that was underlying the patterns in test scores for different tests of ability conducted across several different studies. This pattern is described as the positive manifold, and Spearman made use of the idea that a common underlying factor, or *latent variable*, can represent the positive manifold. The positive manifold is a pattern that Spearman observed in the scores achieved by children on the tests, where positive correlations were apparent between ability test scores, across individuals (Deary, 2000). This means, for example, that the only commonality between verbal test-score outcomes and mathematical test-score outcomes is a general factor of intelligence, or what Spearman ultimately names as *g*, for cognitive ability. The original 1904 paper presents the conceptual background for a common factor or latent variable, making use of the previously published literature that had examined different kinds of patterns in group-level data such as a series of test scores. In later work with colleagues, Spearman makes use of linear matrix algebra in an effort to provide mathematical evidence about the partitioning of test scores into two parts (see Hart & Spearman, 1912).[1] The common factor or *latent variable* is reflected in partial correlations between different test score tallies, which in latent variable modelling terms together are the *manifest variables*, or unique factors (Cowles, 2001; Bollen, 2002). This common factor notion as developed by Spearman forms the basis of a general definition of a latent variable model (LVM) relevant to psychology research today. The LVM is a statistical model of some underlying psychological phenomenon that makes use of one or more relationships or correlations between two or more manifest variables, for which empirical realisations, or data, are available (Bollen, 2002).

Mathematically, the LVM in its most basic form is presented in this way (see Mulaik, 2010):

$$Y_j = a_j G + \psi_j$$

Where Y_j is the $_j$th manifest variable, a_j represents the degree to which Y_j is represented in latent variable G, and ψ_j stands for the specific ability Y_j that is uncorrelated with G. The notion of positive manifold is also expressed as the idea that the manifest and latent variables share a joint distribution over some population (Holland & Rosenbaum, 1986). Several important assumptions accompany the use of this model. The assumption of conditional independence today sits as the key assumption that differentiates the LVM from any other type of representational mathematical model (see Guttman, 1955; Suppes & Zanotti, 1981; Holland & Rosenbaum, 1986; Borsboom, 2005). Conditional independence here means a common factor or latent variable is assumed to underlie or influence a set of variables. If we hold this variable constant for its parameter value, all other variables will be independent of each other. This means, for example, that the only commonality between verbal test-score outcomes and mathematical test-score outcomes is a general factor of

[1] The attempt at mathematical proof for the partitioning of scores is explored in depth in this thesis in Chapter 2.

intelligence or what Spearman ultimately names as *g*, for cognitive ability. The psychometrician attributes any remaining variance in test score outcomes to the respective specific tests. On this point, Borsboom (2005) cites the assumption of conditional independence as that which connects the mathematical model to the empirical psychological phenomenon that is of interest to the researcher. The idea here is that there is a common factor, such as intelligence, which is represented as shared variance in test-score outcomes, identified most recently as the 'common cause' of these test-score outcomes (Borsboom, 2005; Markus & Borsboom, 2013).

The assumption of conditional independence carries with it a number of problems that demand further decisions and assumptions on behalf of the researcher. Two key concerns for psychometricians that follow are the mathematical problem of factor indeterminacy, and the logical problem of circularity. Factor indeterminacy follows the assumption of conditional independence in the latent variable model, as a mathematical consequence of the structure of the LVM (Maraun, 1996; Mulaik, 2010). Factor indeterminacy means that it is logically impossible to connect the outcomes for a latent variable from LVM analysis to the psychological phenomena of interest. The researcher instead must adopt a further set of assumptions, to make this connection (Guttman, 1955; McDonald & Mulaik, 1979; McDonald, 2003). The logical problem of circularity comes about because the structure of the LVM itself means that the evidence accumulated from LVM analysis comes only from whatever tests or questions or items were administered in the research setting. As such, there is no other way for latent variable parameters to be determined.

The challenges presented for all latent variable models by conditional independence, factor indeterminacy, and circularity each are addressed in more detail in Chapter 4 of the present thesis. The proposal of this thesis is that because of the number of assumptions and decisions that are demanded in any research project that makes use of the LVM, a conceptual framework that facilitates transparent reconciliation of the researcher's chosen representations in their project reporting is vital. Some example of the kinds of concerns regarding research contexts that Spearman himself reported as essential with respect to the latent variable modelling outcome are included in Chapter 6 and are reconciled back, to the present proposal. To return to the moment of Spearman's work, whatever else we may say about factor theory, the invention of the LVM captured the imagination of psychometricians with startling veracity. Within a generation, different formulations of the model emerge in the psychometric literature - with multiple latent variables, multiple relations and multiple sources for error, all featuring, and with strong competition evident, between protagonists (see for example, Thurstone, 1934; Thomson, 1939; Carroll, 1941).

The computer revolution of the mid-20th century presented psychology researchers with further opportunity - the ability to conduct complex calculations involved in statistical analysis of mathematical matrices for different kinds of variable structures such as continuous or categorical data, with ease and speed (Cliff, 1992). Analytical techniques that made use of this computing power saw the development of a suite of LVM practices suitable for different kinds of research questions where underlying variables were at stake. These include for example exploratory factor analysis (EFA: Jöreskog, 1967), confirmatory factor analysis (CFA: Jöreskog, 1969), structural equation modelling (SEM: Jöreskog, 1993), and item response theory (IRT: Lord, 1950).[2] Not only this, but different methods for arriving at parameter values suitable for the models also emerged - for example, maximum likelihood (Jöreskog, 1969), weighted least squares (Muthén, 1993), and image analysis techniques (Jöreskog, 1969), among others. These techniques made possible calculations for solutions for different model types, given some empirical data for the manifest variables, and a series of assumptions relevant to the model and estimation

[2] The development of these techniques is described in more detail in the historical overview in Chapter 2.

type. What followed the model developments was growth in different indices and yardsticks that facilitated comparisons between different model and different parameter outcomes. For example, most commonly in published studies that evaluate the fit of a CFA model to some study data, we see the use of a combination of the root mean square error of approximation (RMSEA), the Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI) (see Hu & Bentler, 1999; Thompson, 2004). Despite, or perhaps because of, the disparity and ingenuity with respect to the model variations, the solution techniques, and array of evaluative criteria, several problems occur with the application of the LVM in research practices. These are described in more detail in this thesis in Chapter 4.

The sheer number of possible LVM approaches invited several attempts to unify the disparate forms of the model in the last 20 years or so, under a single rubric. Named then as the generalised latent variable model (GLVM: Skrondal & Rabe-Hesketh, 2004), different theorists have given different accounts of the generalisation. Two that are relevant specifically to psychometric practice are covered in more detail in the present thesis in Chapter 2 (see McDonald, 1999; Bartholomew, Knott, & Moustaki, 2011). Each account represents an attempt to address the diversity apparent in the subsequent historical modifications of, and developments for, the LVM, coordinating these under a single scheme or model. These generalised formats however are characterised by diverse theory and perspectives on what constitutes a model, what constitutes measurement, and what constitute psychological phenomena, as we will see in Chapter 3. Such discrepancies are shown to have distinct implications when we come to consider how a researcher will construct their conceptual framework for their project. The conceptual framework then is maintained in line with the principles of local realism, to account for these commitments in a logical, contextualised, yet unified format.

Filling in the details of planned analyses in standardized conceptual framework format, as it is described in this thesis, asks the researcher directly to consider the logical and structural implications of decisions for their research representations, in virtue of the sets of assumptions that undergird the methodology chosen in their analyses. The distinct GLVM proposals can be conceptualised as vying for attention from researchers, academics, teachers, and professionals, with differing advantages that may or may not be apparent to the researcher, or those who read the project report. What is important here is that the set of assumptions relevant to and implicit in: i) the choice to utilise a particular model for any of the response model, structural model, or error model; ii) the method of calculation for the model(s); and iii) the evaluation of model 'fit' potentially remain inscrutable, and in some cases, may involve incompatible principles as philosophical or method-based assumptions. As it works through the history of the development of the GLVM, this thesis aims to demonstrate problems that can transpire when assumptions are not explicit, or available for community scrutiny. It sets out a framework by which clarification of assumptions and decisions can be achieved.

2. Purpose, aim and research objectives

2.1 Research Purpose

The purpose of this thesis is to put in place the foundations for a conceptual framework designed for the use of the GLVM in psychological research guided by cognitive-historical analysis, set theoretical concerns, and network systems constraints. The intent is to produce a set of conventions, which support a researcher in tying together the mathematical, statistical, and substantive goals inherent in quantitative analysis of psychological phenomena. This is important because these goals are very likely to exist in tension with each other, as detailed in Section 4.2 below. The conventions set out here offer the researcher tools for practice, rather than merely offering critique, review or testing, of current modelling practices in psychology research. In so doing, the intent is to begin from the practices already accepted by psychology researchers as a practicing community, without limiting the spectrum of the work to the current or historical practices endorsed, by researchers.

2.2 Research Aim and Research Objectives

The research aims of this thesis are:

1. Examine the historical development of the GLVM, with specific attention to the work of Charles Spearman and Ronald Fisher. Focus is maintained on the concept of proof in their proposals; also pursued is examination of the assumptions, principles that are present for current-day use of latent variable techniques;
2. Identify problems that persist with the use of the GLVM in psychology research that demand researcher decisions in order to carry out analyses;
3. Investigate the development of set theory and concepts co-extensive with set theory and logic in the philosophy of mathematics, drawing out resources that support unification as beneficial for a conceptual framework;
4. Set out conceptualisations of psychological phenomena with attention paid to: (i) the situation of psychological research activity; (ii) the place of this activity in the broader perspective of a contribution in a specific field of practice; and (iii) research decisions and the connections to research modelling practice;
5. Combine points 1., 2., 3., and 4. above into a new conceptual framework for the use of the generalised latent variable model in psychological research;
6. Apply this same conceptual framework in a working example making use of present-day literature that makes use of the GLVM in computational psychiatry (see Krueger et al., 2018).

In the process of following the above steps, philosophical perspectives are adopted that are informed by local realism (Maki, 2005; Kincaid, 2000b) and mathematical constructivism (Ferreriós, 2016). What follows is a conceptual framework broad and basic enough in scope to allow for coherent integration of sophisticated modelling methods into a research project, while working simultaneously to best account for the nature of the psychological phenomenon that is under scrutiny. This is important as the problems associated with utilising the GLVM in psychological research are shown in Chapter 4 to demand decisions that work to bridge the gap between the model, and reality. Social sciences research such as psychology research has been described as overly-reliant on formalism in scientific inferences and statistical practices, with the casualty being attention to the true nature of the phenomenon in question (Kincaid, 2000a). The conceptual framework serves as an evidentiary basis for the kinds of decisions and assumptions entered into with use of the GLVM, with a systematised structure that facilitates both the account of the unique circumstances of a given research project about some specific phenomenon, supporting both researcher self-audit, and communication with the broader scientific community, about a set of findings.

In the remainder of this introduction, terms that carry throughout the rest of this project are introduced in order to set the research domain. These terms include cognitive-historical analysis, set theory for model-based reasoning, quality uncertainty, and cognitive systematization. To begin with, the term ‘conceptual framework’ bears a heavy evidential load, and to the definition of this, we now turn.

3. Key terms for the research domain

3.1 What is a conceptual framework?

Conceptual frameworks are described in a variety of terms in the research literature, and these terms do not necessarily reconcile with each other. Given (2015, p. 59) for example describes a conceptual framework as “an overarching set of beliefs, theories, and perspectives that shape the design of a research project”. Given (2015) states these beliefs, theories, and perspectives are influenced by several factors, which may include extant literature, chosen methodology, prevailing worldview, and preferred epistemological stance. For Given, the key role for a conceptual framework is to provide resources for theory testing, theory generation, theoretical application, and theoretical interpretation across the whole of the project. Berman and Smyth (2015), on the other hand, propose that a conceptual framework anticipates a theoretical framework for a concept or phenomenon under scrutiny, so that the conceptual framework has a limited role in any other aspect of a research project. In Berman and Smyth's account, the conceptual framework has merely a transitional role, where the theoretical framework takes over. This idea of a transitional role for the conceptual framework however is less suitable for projects making use of the GLVM, because a much broader set of research tools must be accounted for as elements of the research project, such as questionnaires, or data which are outputs from earlier studies. These instruments and sources would have no prior theoretical role with respect to the GLVM, itself.

Not distant from the proposal of Berman and Smyth but with perhaps broader range, Shields and Tajalli (2006) distinguish conceptual frameworks as intermediate theory, which sits between the background literature and the ultimate definitions that emerge as new knowledge from a project, where the intermediate theory serves as practice tools, or "logic-in-use" (Kaplan, 1964, p. 14). These authors take a functionalist approach, where a conceptual framework supports the "problem definition, purpose, literature review, methodology, data collection and analysis" (Shields & Tajalli, 2006, p. 313). Essentially, in this view, a conceptual framework forms the backbone or scaffold, to the whole of a research project (Wisker, 2005).

In this thesis, the conceptual framework for a research project making use of the GLVM will be most aligned with the ideas of Given (2015) and Wisker (2005). For both authors, the conceptual framework has some role as connective tissue between different elements of a project, but more importantly, the conceptual framework aims to envelop the whole of the research project, providing something akin to a metatheoretical lens through which the research project can be viewed. The unique conceptual framework associated with each research project described in what follows works to set out researcher commitments in a way similar to the pre-registration processes that have started to become accepted as a part of the production of psychology research outputs. However, the conceptual framework here does not only serve to set out a priori commitments, as is the case for most pre-registration procedures. It stands to serve the researcher throughout the conduct of their project as a means to evidence and reconcile decisions made and assumptions adopted, in a way to serve to minimise the quality uncertainty associated with the project.

Researchers will be working in different contexts whether concurrently or prospectively. Because of this, the conceptual framework constructed here brings attention to what needs to be accounted for with regards to contexts relevant not only to psychological research, but also to the occurrence of psychological phenomena, or events, relevant to questions regarding reproducibility of psychological phenomena (c.f. Boker & Martin, 2013, pp. 242-243). In Chapters 5 and 6 of this thesis, a case is made for including elements that define the research situation and the research phenomena as distinct components in the representation of a research project, drawing on the principle of local realism, in a realist constructivist approach. In a realist constructivist perspective, a conceptual framework will support the contextualization of the research project in such a way as to render clear the implications of the research situation on the research phenomena of interest. Effort is maintained on distinguishing between the phenomena, the concept of interest, the construct that is informed by psychometric principles, and the variable that performs a function largely in the context of the mathematical model. These distinctions will be shown to be vital in taking steps towards overcoming the limits of the problems that are associated with the use of the GLVM.

A conceptual framework thus provides then a philosophically informed lens through which current psychometric practice in psychology research can be viewed. In the section on quality uncertainty in this chapter, researcher self-audit and scientific community scrutiny each are described as playing a role in both the formulation of what goes in to a conceptual framework, and in evaluation of the research project output, on completion of its parameters. This supports then a reflexive relationship between researcher practices, representation of researcher practices, and accountability for those practices in a public domain, insofar as the awareness of the likelihood of community assessment is itself likely to exert an influence on researcher practices. The logical structure of the framework ensures comparability between studies, but more than that, it provides a singular means for integrating the mathematical, statistical, and substantive aspects of a project that are present whenever the GLVM is drawn into research analysis, in a way that can be assessed, by other researchers.

3.1.2 Mathematics, statistics, psychology - one conceptual framework

The benefit of a conceptual framework for psychological research that utilises the GLVM is evident when we consider the degree to which psychometric practices generally take place at the intersection of mathematics, statistics, and psychology. Each stands as a unique discipline, and each has a different goal as a function of its exercise (see Table 1 below for commonalities and differences between these three fields in psychometric practices). These goals, geo-historically situated as they are, can be expected to change with the passing of time, or even within the context in which the practice occurs. Where multiple disciplines serve in the account of a singular psychological phenomenon, or concept that describes the phenomenon or its event, it is likely that tensions between accounts or explanations for the phenomena will result (Teo, 2010). For example, a researcher may make an inference from a statistical analysis via a mathematical model regarding a singular latent variable, which is proposed to represent human cognitive ability. At the same time, there might be very plausible theory or findings from within psychology research that would say that it is impossible that such a singular variable could account for the same phenomenon (see as an example, Gignac, 2008). Statistical evidence may be accumulated for each model that demonstrates satisfactory fit to data, but without further information accumulated in a systematic conceptual framework that facilitates comparability of the underlying substantive assumptions and principles of each account, no guidance may be available regarding which is the most appropriate model, in some given circumstance.

The way that present-day practice is conducted is important for the development of the conceptual framework for use of the GLVM in psychological research. Borsboom (2005), for example, states that because psychology researchers treat latent variables as real insofar as they use latent variables in what he has described as causal inferences in practice (where the latent variable is considered to cause the test scores), latent variables must be real, at least in the eyes of the psychology researcher. Realism and its implications for our conceptual framework becomes an important part of our considerations in Chapter 5. It can easily be seen that claiming something is mathematically real is different from something that is statistically real, or psychologically real, by reference to the differences between these fields (see Table 1 below). An abstract mathematical function is real in a different sense to data gathered to be used in a statistical model (see Ferreirós, 2016; Suppes, 2002), which is different from a psychological event we could describe as an instance of ‘anxiety’, for example. In psychometric practicalism (see Michell, 1999) the sense of mathematical, statistical and psychological distinction becomes conflated into a singular view, expressed as the idea that the latent variable model places the psychological attribute ‘within’ the model (see Maraun & Gabriel, 2013). Some authors state that latent variable modelling offers an improvement over alternative statistical methods because of the way that the mathematical structure models the relationships for the latent variable (see Borsboom, 2005, 2008). In this thesis, transparency regarding decisions made with respect to the phenomena and their representation in the latent variable are key, to overcoming some of the present-day problems that remain, for the use of the model.

Table 1

Model types used in psychological research

Model Type	**Mathematical**	**Statistical**	**Psychological**
Conventions	Deductive proof Syntactic structure	Inductive inference	Describing relationships between theory and phenomena
Characterised frequently by	Axioms	Probabilities	Situation
Key explicit assumptions	The nature of proof; the nature of a priori assumptions	Nature of model-data relations	Nature of relations between empirical phenomena
Likely implicit assumptions	Relevant number systems	Structure or existence of error	Subjective, cultural or clinical biases

To the degree that it makes use of mathematical, statistical, and psychological inputs, one way to conceive of psychometric practice is as one that is intrinsically pluralist. This is because it draws on the fields of mathematics, statistics, and psychology in the conduct of what it describes as psychological measurement. In a 2005 paper, mathematician E.B. Davies sets out one means by which pluralist practice may be analysed, highlighting how background conventions, context, content, and intuitions each have a role to play in any mathematical activity. We will define these terms following Davies (2005) below, and then have a look at possible contradictions or inconsistencies that are immediately observable for psychometric practices that make use of mathematical, statistical, and substantive information and techniques. Then we will consider what this means for the remainder of the thesis and the construction of the conceptual framework. To begin, we breakdown the practice influences into the four analytical concepts as: background conventions, context, content, and intuitions, with some developed examples, in what follows.

3.1.2.1 Background conventions

Background conventions broadly are the procedures by which experts in the given field accept that statements contribute to the body of knowledge within the field (see Davies, 2005, p. 3). In mathematics for psychology research, this would consist in a correct mathematical proof for some model relevant to psychological phenomena, following Davies. Statistical inferences often rely on mathematical models, but formalised processes guide inferences from data gathered from empirical research situations (see Kincaid, 2000a). For background conventions for psychology research, we find ourselves today bound by statistical analysis in garnering what is described as “evidence-based practice” (Levant & Hasan, 2008) – the idea under the scientist-practitioner model that psychological researchers are interested in the relevance of some psychological statement for a population or sample of organisms under study.

This last point brings us to the question of what psychological research is, and what psychological phenomena are taken to be. This thesis adopts a definition of psychology as the study of mind and behaviour as a function of evolutionarily responsive organism-environment relationships (see Brunswik, 1952; Petocz & Mackay, 2013). With this definition in mind, psychometric research practices are those consistent with the study of organism-environment relationships in a way that contributes to the evidence base for psychological claims (see Petocz & Newbery, 2010). Background conventions in this regard draw on mathematics, statistical practices, and philosophy for checks of logic and consistency of inferences, regarding statements made in the derivation of concepts utilised in the field, or for concepts that are new for the field.

3.1.2.2 Context

Context as stipulated by Davies (2005, p. 2) includes "assumptions, axioms, and definitions" relevant to the field. Davies (2005) identifies the approach he takes within mathematics as contextualism. Because of the advances made in mathematics and logic at the hands of Frege, beginning around the time of Spearman's development of factor theory (see more on this in Chapter 2), it could no longer be taken for granted that mathematical assumptions and logical principles completely coincided. This was true under old principles of formal logic, but now what was needed was clear specification of the context of assumptions within which a set of logical rules were understood to apply. Contextualism in this form will continue to appear throughout this thesis in other guises including as local realism (see Mäki, 2005; Kincaid, 2000b, Ferreriós, 2016, and the section below). Davies (2005) notes that different contexts in mathematical practices stand to create different mathematical universes, where substantially different types of theory are employed. In psychology research, the context of the research in terms of the interweaving between assumptions, principles, and definitions relevant to the research situation may remain unspecified in the reporting of research outcomes. This can have significant consequences, where for example, some psychological phenomenon such as intelligence or anxiety may be assumed to have quantitative structure, but no evidence of quantitative structure exists for that phenomenon in that research situation (Michell, 2004b). Other consequences of adopted models and adopted philosophical stances are presented in Chapters 3 and 5 of this thesis. The conceptual framework developed in this thesis is designed explicitly to provide the tools for accounting for the assumptions, principles, concept definitions, and specific research situation for any researcher making use of the GLVM.

3.1.2.3 Content

Content describes, for Davies, "theorems, both individually and collectively as a theory, which have been proved in a particular context" (2005, p. 3). In the case of latent variable modelling, Spearman's idea about a mathematical partitioning of test scores based on observation of the positive manifold in children's ability test scores provides the theorem essential to the LVM. However, it should be noted that for psychology research, this theorem is one that has never been proven (more on this is presented in Chapter 3). In Chapter 3, we also examine Fisher's failure (Aldrich, 1997) to find evidence for isomorphism between the probability density function that applies to an overall model, and the probability density functions applied in an additive way to variables, for maximum likelihood estimation techniques. Subsequent developers have never secured proof for these two components of psychometric methodology that are both employed in the GLVM paradigm. Yet it is the case that we still make use of both mathematical models in psychometric practices, despite this lack of proof. This is likely because practicalism has driven the usage of these two methods - perhaps with the hope, as is sometimes the case within mathematics, that the proof would eventually appear (see Ferreirós, 2016; Feferman, 1998). Davies (2005, p. 6) notes that in mathematics, it is when some method is investigated for its formal detail that shifts occur in understanding in a way that is relevant to the whole field. Suppes (2002) notes a similar trajectory in the development of theory in the physical sciences, suggesting that formalisation as a process has the benefit of revealing assumptions that may be questionable or ill constructed. Investigation moves practice forward, as conventions that were previously taken as given are pulled apart and explored. The conceptual framework developed in this thesis supports this type of investigation in a way not otherwise truly made available in the field of psychology research to date. This is because it provides a standardised and generalised structure into which all element of information relevant to a representation of the research project can be input. In this way, separating out the implications of theory from assumptions, definitions, and other aspects of the research situation becomes possible in a way that facilitates discernment of their differences and influences on each other.

3.1.2.4 Intuitions

Intuitions for Davies (2005) are not defined explicitly, but they are suggested to provide guidance for reasoning, given what is at play in the background conventions, and what is required, for clear definition of the scientific concept or event that is under scrutiny, when considered by the researcher. Intuitions broadly align with implicit assumptions for a relevant field or paradigm, which represent ideas or information taken to be so obvious as to be given, in a field. Davies describes a closed loop between content and intuition, where intuitions and content can reciprocally influence each other, within the domain of the context. For example, assumptions about the structure of error in statistical modelling may lead to certain ideals being assumed about the nature of data for a phenomenon, whether the idealised features could be said to apply for the phenomena-data relationships in question, or not. Davies notes that particularly in the field of mathematics, mathematicians may be reluctant to relinquish favoured intuitions. This perhaps persists in psychology research as well, to the degree that researchers continue to use statistical methods which may not completely match to the actual qualities of the data for the phenomenon (see Michell, 2004b regarding factor analysis or item response theory; Kincaid, 2000a). In other possible examples, intuitions or 'good guesses' about the importance of some psychological phenomena, for example, may lead researchers to analyse data for psychological events based on the convenience of a normal distribution for the data. This may occur even though the researcher may consciously know that no such normal distribution exists for the dataset in question (MacCallum, Brown, & Cai, 2007).

The conceptual framework developed in this thesis invites researchers to participate in a process of accounting for their research intuitions and implicit assumptions in a way typically demonstrated well in qualitative psychological research reporting (Ponterotto, 2005; Meyrick, 2006). The notion of "researcher as instrument" (Given, 2008) has roots in the introduction of qualitative methods of evaluation exemplified in Guba and Lincoln's

(1981) *Effective Evaluation.* The naturalistic approach of Guba and Lincoln acknowledges that researchers may influence their participants through human interaction in a way that should be specified in accounts of psychological research, but more than this; the researcher's perspective constitutes the original lens through which the research practices and their outputs are interpreted (see also Denzin & Lincoln, 2000). Biases may influence both the assumptions made in respect of representation of the project and its outcomes, as well as the actual set of events in a research situation, influencing the data collected. What is important in qualitative research accounts is transparency regarding researcher commitments (Levitt et al., 2017). For the contextualism endorsed in the formulation of the conceptual framework in this thesis, researcher intuitions, as well as institutional or ideological commitments are shown to have important roles, in the ultimate interpretation of research outcomes for a project making use of the GLVM.

3.1.3 Interim Summary

In the above material, a working definition for psychological phenomena suitable for the rest of this thesis is set out. Advocacy for a contextualist approach that endorses representation of both researcher perspectives and breakdown of theoretical or content-focused aspects of the research project, which may otherwise remain abstract or unknown, is endorsed. In the next section, terms that shape and formulate the body of the conceptual framework are briefly defined for working purposes as the exploration of the history of the development of the GLVM continues, and prior to rich explication in subsequent chapters 5-8.

3.2 Cognitive-historical analysis

In this thesis, a cognitive-historical analysis of the GLVM is presented. Cognitive-historical analysis as developed by Nersessian (1995, 2008) is a methodological naturalist approach (Schwarz, 2017) that is interested in the process of conceptual change in science. Distinct from a logical positivist approach, which would seek a rational reconstruction of scientific practices, methodological naturalism works from the ground up, informed directly by what takes place in practices, as they occur in science (see also Haig, 2014; Maddy, 2011 for similar yet distinct perspectives). Such a position is relevant to the GLVM in two respects. Firstly, as we have noted above, the very meaning of the GLVM itself has changed within psychometric practices, in virtue of the changing aims for its applications, in time. In this way, this thesis tracks the historical development of the GLVM, and looks to the changing criteria both within science generally and psychometrics particularly about how scientific practices contribute to the formulation of knowledge. Secondly, the GLVM may be used in processes that aim to justify changes for psychological concepts, when for example a GLVM analysis is utilized as justification for changing the items that are included in a self-report questionnaire for some particular psychological phenomenon (see for example Markus, 2008; Markus & Borsboom, 2013). Throughout what follows, then, we navigate both a meta-theoretical perspective on the GLVM, as well as a consideration of the direct application of the GLVM in the real context of what is done in research.

Conceptual change, in Nerssessian's account, asks that we be aware of the cognitive implications of types of scientific practices, as much as we might describe histories of development of these same practices. This is because the practices themselves have a secondary impact on conceptual, model, or theoretical change (Nerssessian, 2008; Schwartz, 2017). In psychology research that makes use of the GLVM, model-based reasoning comes to occupy a substantial role on psychological research practices, through time (Mislevy, 2009). The principles behind the GLVM, particularly the idea of underlying latent variables which represent the operation of underlying psychological phenomena can be seen to be particularly influential in their role which formulates in part, how psychologists think about what is needed in the development of their research projects and protocols (see Maraun, 1996; Borsboom, 2005).

To date, to this researcher's knowledge, no systematic conceptual framework has appeared in the psychometric or psychology literature that works to articulate the details of the representations of the phenomenon that is of interest, or the context of the research, itself. This is unfortunate, as, per Nersessian's account, *thick descriptions* of both the representation and the context support robust claims about the actual outcomes of scientific processes, particularly those that are invested in conceptual change. The term thick description (Ryle, 1971; Geertz, 1973) has a home within qualitative research investigations (Ponterotto, 2006) and refers to rich data that result as a function of maintaining close attention to the nature of the phenomenon that is under scrutiny (Geertz, 1973; Ponterotto, 2006). As this thesis unfolds, there will be continual return to practices informed by qualitative research literature, conceptualisations that stand to be of great support in a deepening and broadening of the reporting of qualitative aspects of research projects, which are of significant import to quantitative processes such as those that accompany the use of the GLVM in psychology research. Such an approach seeks to integrate learnings and principles endorsed in psychological science as demonstrated by the admission of qualitative inquiry into the American Psychological Association's Division 5, with the change of name for this Division from Evaluation, Measurement, & Statistics to Quantitative and Qualitative Methods (Gergen, Josselson, and Freeman, 2015). The conceptual framework established in this thesis integrates the logic that supports robust inferences in psychological science with qualitative information that is a necessary adjunct to any quantitative aspect of psychological phenomena reported in research project outputs.

3.3 Set theoretical principles for representation and invariance

Integration of qualitative and quantitative aspects of a research project in a logical form that is amenable to distinct circumstances relevant to representation of different types of research projects demands a structure that is universal and precise, in its definition. The lifetime work that is collected in Suppes (2002) *Representation and Invariance of Scientific Structures* provides guidance on how the two paradigms of empirical information and abstract representation together feature in scientific practices, through time (Ferrario & Schiaffonati, 2012). Considering logical concerns, this is important because these two paradigms lend themselves to distinct kinds of logical description and analysis (Kaplan, 1964; Suppes, 2002, 2014).

With some parallels to the representational theory of measurement set out in the classic *Foundations of Measurement Volumes I, II, and III* (Krantz, Luce, Suppes, & Tversky, 1971, 1988, 1990), Suppes outlines the means by which constraints, articulated in abstract structures informed by set-theoretical considerations, provide guidance about invariances as they arise in the context of any research project. These invariances are important, as in Suppes's terms, they constitute exactly what science is interested in. Where once such invariances in science were marked with the character of eternal and immutable law, developments in the philosophy of science particularly through the 20th century saw the emergence of model-based accounts of scientific strategy (see for example, Godfrey-Smith, 2006; Levins, 1966). Notably as time has gone on, less attention in these strategy-based approaches is paid to logical structure, and greater focus has been afforded to pragmatic views of models as approximations (Godfrey-Smith, 2009).

For the conceptual framework developed in this thesis, logical structure serves as the very ground against which research claims may be evaluated. Representation of a research project takes place in research elements, which are sets of ordered information that describe the domain, the formulation, and the variables for each specific aspect of the research project. Research elements are linked by setting one specified element as the domain sub-element of a second element, and so on. Degrees of freedom are reserved for the researcher in this respect, as the researcher may nominate the form and type of links that are made in their project. In this way, the conceptual framework is set up to serve transparency regarding researcher decisions, as much as the meta-theoretical structure serves to facilitate integration of disparate study outcomes. This is a process typically undertaken through statistical meta-analyses across peer-rerviewed reports of psychology studies, which involve no set of standards for reporting of outcomes, assumptions, or contexts of research. Two other features of research reporting may facilitate good quality scientific inference from psychological research making use of the GLVM, these include quality uncertainty, and cognitive systematization with its role in inference from the best systematization. These are briefly described, in the next sections.

3.4 Quality uncertainty

The sheer volume and complexity of assumptions associated with the application of the GLVM in any research setting places some burden on quality uncertainty associated with the research project reported outcomes. Quality uncertainty is described as “the inability to make informed and accurate evaluations of quality” (Vazire, 2017, p. 1). Originally coined by Nobel Prize-winning economist George Akerlof (1970), the term was used to describe the effect of decreased market trust on price outcomes, particularly in his original paper, for second-hand cars, or “lemons”. Vazire (2017) notes that in recent times there has been a decrement of trust in the outputs of scientific products (see Ionnidis, 2005). This is partly driven by forces of peer-reviewed journal publishing which may increase adoption of questionable research practices (Simmons, Nelson, & Simonsohn, 2011) and other forms of information obscuration that enhance likelihood of publication (Crede & Harms, 2019).

Quality uncertainty is built from information asymmetry, and failures to engage in adequate quality control regarding the type of information made available (Wankhade & Dabade, 2006). Vazire (2017) proposes that improved transparency regarding scientific practices provides a direct amelioration targeted to quality uncertainty. She does not address questions of quality control directly. In this thesis, what is explored is a role for the conceptual framework not only in supporting transparency, but also in negotiating outcome values assessed by both the researcher and research peers, present and future, regarding quality control for the representations made in the research project. The process relies on the logical structure of the conceptual framework, and an assessment of the systematicity of the project elements, and to an introductory description of cognitive systematization as exemplified in the work of Nicholas Rescher (1979, 2016), to which we now turn.

3.5 Inference from the best systematisation

For a conceptual framework that remains open to diverse philosophical perspectives, delineation of a clear structure within which the framework can be articulated is vital. Local realism takes the specification of the principles and context involved in the production of some research outcome as essential elements of that outcome (Kincaid, 2000b; Maki, 2005; Ferreriós, 2016). No matter the type of scientific work that is underway, whether it be description, explanation, or prediction (see Chapter 2 of the present work), well-specified constraints formulate a space within which interpretation of both the scientific practices and research outcomes can take place. At a minimum, these constraints are the principles and context of the research project. For a project that involves the use of the GLVM, more than just principles and context make up the constraints for the research outcomes – as seen following Chapter 4 of this thesis, several statistical, mathematical, and substantive assumptions are involved in the production of a research outcome.

Rescher (1979) proposes that the key feature that for a contribution to science is demonstrated *systematicity* between principles, assumptions, and aspects of the substantive situation of science. Systematisation has different possible forms (such as coherentism or foundationalism), but in common for any system that contributes to knowledge are three cognitive features: i) that the system renders the context intelligible to us; ii) it serves in an organizing capacity; iii) it constitutes the means of verification of knowledge. For Rescher, the degree to which an account of scientific practice is systematic is the degree to which the reach of ignorance and error are reduced. In *Cognitive Systematisation* (1979), Rescher draws Aristotle's preference for systematisation; Kant's account of rationality; and what he describes as the Hegelian Inversion (described in Chapter 7), into a test of the truth of a claim. One advantage of inference from the best systematisation is that it embraces the broad scientific goals that the GLVM serves for psychology researchers, including description and prediction, beyond explanation. Advocacy for inference to the best explanation and causal explanatory accounts of the GLVM are apparent in the psychometric literature (Borsboom, 2005; Haig, 2009). However Chapters 2, 4, and 7 of the present work show in a number of examples the ways that a broader view of scientific activity is already adopted in the ways that the GLVM is employed, in psychology research practices.

4. Chapter outlines

In Chapter 2, a historical account of the development of the GLVM is presented, including looking to the logical structures of the distinct scientific processes of description, explanation, and prediction in model development. Chapter 3 initially presents two specific accounts of the GLVM as proposed by McDonald (1999) and Bartholomew et al. (2011). These are contrasted to arrive at a perspective of how these different generalised accounts insist on different philosophical orientations to the model and phenomenon, with different requirements resulting for the framing of research outcomes, for each. Then in the second half of Chapter 3, the mathematical work of Charles Spearman and Ronald Fisher is reviewed exploring the question of mathematical proof, looking at the degree to which mathematical proof was produced for the factor analytical model of Spearman or the maximum likelihood method, of Fisher.

Chapter 4 presents consideration of problems to do with the GLVM in current day practices. These are broken down into theoretical concerns intrinsic to the model that include conditional independence, factor indeterminacy, and circularity, and then common problems that present for applications of the GLVM, such as resolution of between versus within subject distinctions, causation claims, equivalent models, variable structure and the nature of infinite populations. A further set of concerns regarding model testing are also investigated including goodness-of-fit heuristics, measurement invariance techniques, reliability and validity roles and processes, and other applications aspects of modelling. Specifically of focus are the logical concerns present for each of these. A view is drawn in the closing of this chapter of the role of quality control in consideration of reliability and validity assessments.

Chapter 5 recounts interpretations of the practices of modelling and measurement made under present-day philosophy of science-informed rationales for latent variable modelling methodology and practice, with attention paid to the account of realism endorsed in Borsboom (2005) and Haig (2014). These two distinct accounts of notionally the same philosophical orientation are demonstrated to produce diverse consequences for the kinds of constraints that would ultimately apply to the claims made following a GLVM analysis. Following this, set theory is introduced for its structures that provide a unifying framework that stands to offer tools for accounting for the research situation in a way that provides some foothold for ontology for the psychological phenomena, or recording of invariances that are of most interest to the researcher. A version of realism identified as local or thin realism is discussed as appropriate to this grounding of the research situation in a realist perspective.

In Chapter 6, the conceptual framework itself is introduced, with reliance on the work of Patrick Suppes to draw out his roles for axioms in models. The question of measurement for the GLVM, and the role of pre-registration in respect of evaluation of research claims also feature in this chapter. Chapter 7 introduces the concept of cognitive systematicity as developed by Nicholas Rescher (1979), and answers how inference from the best systematization may serve in psychological science, with reference to the principle of logical consistency. Logical consistency is also investigated in reference to the development of mathematical constructivism. Chapter 8 presents a worked example from the computational psychiatry psychometric literature, looking to the constraints we must articulate for the research situation and phenomena, data, variables, models, theory, and researcher stance, incorporating both qualitative and quantitative insights.

5. Summary

This chapter has introduced the topic of this thesis, described the purpose, aim, objectives, and rationale for the study, and set out some of the preliminary background, which will be expanded in later chapters. Summaries of later chapters are also included, and focus has been maintained on the benefits of a set-theoretical approach to psychological modelling. Rather than taking a mathematical approach to modelling psychological phenomena limited to merely setting out the formula that represents a theoretical model, this thesis proposes that mathematical set theory best underpins the whole of any research project, from its earliest or most fundamental ontology, to its most particular statement or universal generalisation.

The next chapter sets out the present research domain in detail and draws together a framework describing the core terms that are necessary for what follows in later chapters. Chapter 2 sets the scope of the study, describes criteria by which the conceptual framework may be both set out, and then, as the research project unfolds, judged. Chapter 2 also excavates history relevant to the development of latent variable modelling techniques.

Chapter 2: Research Domain and Historical Developments in Latent Variable Modelling

1. Objectives

The objective for this chapter is to set out the historical development of the GLVM with a view to further specifying the domain of enquiry for the project. Clarity regarding the domain of application for the conceptual framework is important for three immediate reasons. Firstly, latent variable modelling techniques have been adopted across diverse fields beyond psychology, including computer science (Blei, Ng & Jordan, 2003), medicine (Yang & Becker, 1997; Rabe-Hesketh & Skrondal, 2008), economics (Bauwens & Veredas, 2004; Muthén, 1983), and financial market analyses (Diebold & Nerlove, 1989; Hodrick, 2014). In this thesis, focus is maintained specifically on the development and use of the GLVM in the field of psychological research. While there may be brief reference to use across other disciplines, there is no attempt to set a conceptual framework for any field outside of psychology.

Secondly, as seen at the end of Chapter 1, the theoretical diversity just across the three fields that are vital to psychometrics - mathematics, statistics, and psychology - brings with it concerns regarding the distinct goals associated with each. For example, mathematical theory demands closed models and formal proof (Krantz, 2007), while substantive psychological theory demands open models which interact with empirical situations (Teo, 2015). A unified framework that seeks to ground psychometric practices must remain broad enough to meaningfully integrate contributions from at least these three fields, while remaining open to future methodological developments in any of these.

A third layer of concern arises with the distinct goals of modelling present at different historical moments, per cognitive-historical analysis (Nersessian, 1995, 2008). Relatively recently, claims have emerged in psychometric literature that latent variables serve in an explanatory capacity, as causal entities that influence the outcomes seen in test scores (Borsboom 2005; Hood, 2013; Haig, 2014). To understand the logical consequences implicit in such a claim for use of the GLVM, we need to return to some earlier moments in the development of the method. As we do this, we can raise question as to whether psychology research through time was only interested in the goal of explanation in respect of scientific discovery, or does it include other processes relevant to science, including at a minimum, description, and prediction? How would evidence for, or about psychological phenomena garnered via GLVM analysis amount to an account that satisfies any or all these kinds of scientific goals? These questions are of interest to us insofar as they help us to understand exactly what the conceptual framework will need to address.

As a guide to what follows, this chapter begins by examining how description, explanation, and prediction feature in psychology research. The historical approaches to latent variable modelling are then characterised as interested in any, or all of, description, explanation, or prediction of psychological phenomena. The chapter also examines the role for logic that is apparent in the work of the psychometricians who come to be interested in GLVM techniques, historically. This will clarify terminology for logical tests used in the conceptual framework, informed by an understanding of the *laws of thought*, laws that have had a role for knowledge since the time of Aristotle (see Boole, 1854; Maddy, 2012).

1.1 A review of the structure of the latent variable model

Before we progress too far forward, it is helpful to review the essential elements that need to be in place to construct the GLVM, as we have set them out in Chapter 1. These are, following Howe (1955):

1. Is there a pattern of correlation between manifest variables, and, does this pattern differ from a correlation of one, or unity, in all pairwise correlations? (*positive manifold*)

2. If a correlation pattern is discernible, is there also a latent variable f_1 such that, if we hold this constant, all partial correlation coefficients between the manifest variables reduce to zero? (*conditional or local independence principle, or assumption*)
3. If not, are there two latent variables f_1 and f_2 so that, if these together are held constant, the partial correlation coefficients for the manifest variables reduce to zero?
4. Does some iteration of step 3. at the n^{th} point involving n latent variables, reduce the partial correlation coefficients to zero?

Important to keep in mind in what follows is the degree to which the latent variable relies on correlations between test scores, as brought together by some researcher, at some point in time, for some specific purpose, or question. This means that the set of correlations that are foundational to the development of any GLVM for a research project are a product of a *theory of expected relations*, as described in Chapter 1. There are consequences for what needs to be considered, then, regarding these correlations, or a test of reducing partial correlation coefficients to zero, in reference to the specifics of the situation in which the correlation occurs. As Guilford (1965) observes

> . . . *a correlation is always relative to the situation under which it is obtained, and its size does not represent any absolute natural fact.* To speak of *the* correlation between intelligence and achievement is absurd. One needs to say *which* intelligence, measured under *what* circumstances, in *what* population, and to say *what kind* of achievement, measured by *what* instruments, or judged by *what* standards. *Always, the coefficient of correlation is purely relative to the circumstances under which it was obtained and should be interpreted in the light of those circumstances, very*

rarely, certainly, in any absolute sense. (p. 105, italics in original).

One concern of interest for the GLVM is the process of reasoning from correlations between test score outcomes and the positive manifold as observed by a researcher to ideas about measurement of underlying psychological phenomena, which occurs relative to a set of circumstances. A key question that follows then is how the reasoning takes place from evidence for correlation among a certain set of variables as occurs in a specific research context, to the kinds of generalisations worthy of the label scientific discovery. These questions are particularly important for latent variables in psychology research, because they are precisely *latent* – unavailable for direct or independent confirmation, regarding their status (see Maraun, 1996; Borsboom, 2005). There are a series of problems associated with the generation of data for a latent variable, which are covered in Chapter 4, and some concerns from a mathematical perspective regarding the logic of the proof for a latent variable, which are set out in Chapter 3. In this chapter, the historical development of the model with a view to logical structure is examined. Logical structure for the GLVM is pivotal, because outcomes of statistical analysis at best add to an already existing evidential base for some phenomenon (Abelson, 1995). Statistical inferences are best made where there is firm logical footing, as a starting point. Outcomes from tests of logical implication, where they are found, are irrefutably true, in all contexts or circumstances. A logical outcome cannot be otherwise than what it is, whereas a statistical outcome is always made with reference to chance – it is, at best and always by definition, probabilistic (Achinstein, 1971). With the use of the GLVM as a notional instrument in psychological measurement[3], clarity is demanded for about the logical conditions sought and met by

[3] See Raykov and Marcoulides (2011) and Boateng et al. (2018) regarding use of the word instrument for the GLVM- typically this follows from the idea that the latent variable represents a construct, and the manifest variables or items in a questionnaire in some way ‘tap’ this construct, with an ideal of measurement in view. This is challenged in the present thesis as an interpretation of the GLVM, in Chapter 5.

instrument developers, through history. We are interested then in the use of reasoning made by the engineers of GLVM techniques, noting that logic would provide founding ground for the most robust truth claims, made from use of the model.

In this chapter, a detailed overview of the scientific processes of description, explanation, and prediction is set out. Initially, elements of logic presented in the laws of thought are presented. The chapter then connects this analysis to the role of logic in cognition, and further makes links to the histories of development of logic and set theory in mathematics. These are taken into the analysis of the structures of description, explanation, and prediction in psychology research and each of these scientific processes is matched to the history of the development of GLVM techniques. This sets basic terminology for the breadth of the conceptual framework articulated in Chapters 6-8.

1.2 Laws of thought

In Chapter 1, it was noted that Spearman first published on factor theory at the turn of the 20th century. Factor theory becomes the basis of all that follows for the GLVM. At the same historical moment, remarkable change was occurring in logic and mathematics as fields, largely informed by two intersecting undercurrents – change in logic, brought about by the developments of Frege (Gillies, 1992), and change in conceptualizations of the foundations of mathematics and infinity, informed by the development of set theory under Cantor (Grattan-Guinness, 2000). Frege institutes a shift from the confines of formal syllogistic logic that marked the practice of mathematics and philosophy through to the nineteenth century (Moore, 1988). Formal logic was characterised until this point in Western intellectual history by the theory of the syllogism, as first found in Aristotle's *Organon* (Kneale & Kneale, 1962). Syllogistic structure for a series of statements or propositions allowed that the force of deductive argument could be claimed for two or more premises to a logical conclusion – the statements, taken together, formed a closed system, by which the conclusion would logically follow (Khemlani & Johnson-Laird, 2012). An example from Khemlani and Johnson-Laird (2012, p. 429) is the following:

> Some artists are bakers.
> All bakers are chemists.
> Therefore, some artists are chemists.

Logical structure for any statement was ascertained according to coherence with this pattern of argument with subject-predicate form (Corcoran, 2003). To ascertain whether logical structure was present for a syllogism, a system where the truth or falsity of statements that made up the premises was needed; with this, the overall truth or falsity for the conclusion of the syllogism could be determined. The basis of such a system is known as the laws of thought, and these are introduced, below.

Frege is credited with instituting a shift from the subject-predicate form of syllogistic logic, a structure that had meant that all arguments needed to be made about specific properties of entities, or subjects (Grattan-Guinness, 2000). Both Frege and C.S. Peirce are noted in the literature as independently developing systems that go on to become the basis of what is described as first-order logic, with roles for quantifiers of properties, and roles for relations and functions which had not been seen in the subject-predicate logic and syllogistic structures of Aristotle (Moore, 1988; Kneale & Kneale, 1962). The innovations took shape in the form of an argument-function structure clarified using symbol systems, so that arguments could be set out in a generic or universal form, and then applied to specific situations or contexts (Grattan-Guinness, 2000). Frege was concerned with developing a universal language, one that allowed for the fact that two individuals could have different sense impressions or cognitions, and yet be looking at the same entity and making commensurate judgements about it (Ricketts & Levine, 1996). This for him meant that there needed to be a logic broad enough that it would be able to substitute, roughly, each of their perspectives into the functional form and still derive the same outcome, which was the truth of the situation as it presented itself to the observers (Frege, 1918/1956). Frege through his pursuit of absolute logicism, or the idea that all number systems are founded in logic, comes to define the process of domain setting, for an argument-function structure (Wilson, 2010). With a domain set, the different argument-function forms from two distinct perspectives could be compared. A distinct role for cognition was thus set for logic in the comparison and determination of truth and falsity with the innovation of argument-function relations, different from the straightforward machinery of deduction operable in syllogistic reasoning structures (Bell, 1981; Grattan-Guinness, 2000).

Frege's invention constitutes a major advance both for logic and for mathematics (Grattan-Guiness, 2000; Wilson, 2010). The idea of a single abstract argument-function form for different interpretations is consonant with argument that the GLVM, for example, represents a syntactical model that can be used in different research situations (Borsboom, 2005), whether the phenomenon of interest is intelligence, personality, or psychiatric illness. An important aspect of Frege's function-argument form, however, was that it meant that statements could be designated as true or false, insofar as the function-argument form described a map for ranges, values, or concepts (Wells, 1951). Determination of what is true and what is false relies on the laws of thought. In the next section, the laws of thought are described, and then the chapter examines how Frege's basic infrastructure is employed in processes of logical deduction and induction that are relevant to scientific practices, even today.

1.3 Logical laws of thought

Welton (1896) describes the Aristotelian laws of thought as those that are a priori, or essential, in the first instance, for valid thought. This essentiality implies that the laws of thought were necessary in a formal sense. In this way, the laws of thought provide some benchmark or test, regarding the nature of things – no real entity, situation, or event could exist in contradiction with the laws of thought. As such, the laws of thought provide us with some logical constraints (see Lange, 2013; Ferreirós, 2016) by which we can reason about ontology, or questions regarding "what is" in psychology (see Hibberd, 2014). This is important for the GLVM, as we will see in what follows, since the nature of the latent variable in respect of ontology has been the subject of significant theorising recently regarding psychometric practices (see Borsboom, 2005; Hood, 2013; Maraun, 2017).

The laws of thought typically are cited as first set out in Aristotle's *Metaphysics*, although some trace their history back to the original work of Plato's *Theaetetus* (see Hamilton, 1860; Kneale & Kneale, 1962; Corcoran, 2003). The first law is the principle of identity, which says a sentence is strictly equivalent to itself. Another way to state the same thing is to claim, what is, is (see Russell, 1912), or to note that when something is said, the opposite is not meant, by what is said (see Boole, 1854). The second law is the principle of non-contradiction, which states that it is impossible to assert simultaneously and in the same context, that something is both true, and not true. In this scenario, we have either the proposition P, or the proposition not-P, and these are mutually exclusive. Another way to say this is that two contradictory statements, say P, and not-P, cannot both be true at the same time (Russell, 1912; Maddy, 2012).

In addition, the third law is the principle of the excluded middle - which implies that for any statement, it is either true, or not true, and there is no third (or middle) option (Cooper, 1978). Aristotle clarifies this best:

> But on the other hand there cannot be an intermediate between contradictories, but of one subject we must either affirm or deny any one predicate. This is clear, in the first place, if we define what the true and the false are. To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true; so that he who says of

anything that it is, or that it is not, will say either what is true or what is false.

- Aristotle, *Metaphysics*, Book IV, Part 7 (translated by W.D. Ross)

These three conditions, taken together as the principles of identity, non-contradiction, and excluded middle, mark logical structures that can be found through the history of Western intellectual history. They are taken to mark, as Frege (1884) puts it, not something of the external world, nor something merely of cognition, but the "laws of the laws of nature" (Kneale & Kneale, 1962, p. 739). As such, these principles give the most general set of constraints on a universe of discourse, or what Frege (1884) would call, rational thought.

As stated above, empirical scientific research that makes use of mathematics, statistics, and substantive theory employs something of both abstract logic and empirical evidence, in its reasoning and inferences (see Suppes, 2002), not one to the complete exclusion of the other. When making inferences, no matter the goal of the scientific processes or commitments to different justificatory systems such as correspondence versus coherence theory, scientists presumably have an investment in making representations that are consonant with truth (Haack, 2011). When scientists aim at truth in their practices, they do so in order to state something about the way things really are (Michell, 2004b; Mackie, 1973). In the history of science, where prior to Frege the only way to get to truth was via syllogistic logic, the growth of relative logicism in the 19th century (Wilson, 2010) and then logical positivism in the 20th century (Maddy, 2012) undermined the idea that there was only one scientific method which could lead, to truth. Logical positivism demanded verification with empirical observations for truth claims (Friedman, 1999). Following Michell (2004), logical positivism was successful in doing this to the degree that it relativised logic. Distinct from the relative logicism of the 19th century, which sought to understand different systems from the singular perspective of logic (Wilson, 2010), logic in the 20th century era was made into another linguistic framework among many, following the work of Carnap (1950). As probabilistic frameworks came to dominate in empirical scientific practices, relativism took hold of scientific imagination – one inferential process was just as likely to be as good as any other, anchored by probabilistic analysis of empirical observations (Suppes, 1972).

Today in psychology research, rather than agreement over what makes the best inferential process for truth statements about psychological phenomena, there is advocacy in the field for different techniques. These include for example, abduction in respect of identifying new phenomena (Haig, 2005, 2009), factor analytic approaches for locating underlying phenomena under existing symptom sets (Huys, Maia, & Frank, 2016); or the newly emergent dynamic network techniques to cluster phenomena in new ways (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). Researchers using these processes may have distinct ideas about what makes empirical phenomena empirical, but there remains at present no way to compare either the empirical process or the analytical statistical process, employed by researchers across studies. In this thesis, a process of inference from the best systematization (Rescher, 1979, 2016) is presented as supported by a logically structured conceptual framework that facilitates transparency regarding the assumptions and principles adopted by a researcher in their reasoning processes. Because of the unified structure, inter-study comparison becomes possible with respect to the principles assumed, the assumptions adopted, and the analyses carried out. Quality uncertainty is reduced when transparency is supported (Vazire, 2018), and a logical structure provides the consistent ground on which distinct research protocols can be evaluated. More detail follows regarding this technique in Chapters 6 & 7 of this thesis.

In summary, what is needed for psychology research today are techniques which: (i) facilitate comparisons between different model systems; (ii) allow researchers to overcome the limits of relativism and evaluate claims; (iii) allow for judgements about the processes employed in scientific research projects, and; (iv) provide structure for reporting of outputs of these projects. Basic structures of logic, for example, facilitate discernment between truth and falsity for a set of statements. We rely on this basic discernment, no matter the kinds of logical reasoning undertaken – and no matter the method employed to target scientific truth. This reliance on the structure of logic lies at the heart of the process of critical inquiry, which is interested in the maintenance and pursuit of systematic doubt (Michell, 2004b; Cohen & Nagel, 1935). A stance of critical inquiry is important in a time where it remains challenging to evaluate truth claims made in virtue of different research analysis processes against each other. Michell (2004) cites critical inquiry as a battle against a kind of natural epistemic inertia, which comes about at least in part because there is natural tendency to accumulate evidence that has built within it our own biases, rejecting the uncongenial as false. Critical inquiry works to find alternatives in respect of what is believed to be true and evaluate these possibilities in a way that seeks to defeat uncritical intellectual conformity. This works not only at the level of the chosen model, with a tendency noted in the GLVM literature for researchers to fail to present plausible alternative models to the one chosen by the researcher (Crede & Harms, 2019), but also applies as an attitude of scrutiny towards assumptions that may not align with actual research situations (see Michell, 2004b). As will be developed in subsequent chapters, critical inquiry is a quality relevant to the proposed conceptual framework for the use of the GLVM in psychological research. Critical inquiry also serves in examination of claims made in the methodological literature that addresses use of the GLVM, where for example, recent claims that the latent variable is a causal and explanatory phenomenon can be observed (Borsboom, 2005; Bollen, 2011). As described briefly in Chapter 1 and elaborated below, the GLVM can however also be used in descriptive and predictive capacities, raising questions concerning any interpretation of the latent variable as merely explanatory. In the next section, this thesis examines the roles for description, explanation, and prediction in scientific research for psychology, to help clarify the domain relevant to psychometric practices that make use of the GLVM, to understand how logic may be used in each.

2. Logical structures relevant to description, explanation, and prediction

In what follows it is shown that description, explanation, and prediction each have a distinct logical structure. In the GLVM literature, all these three terms are used in connection to the GLVM, but in different ways, or sometimes, interchangeably. Partly this may have to do with the rapid development of methods that do not strictly adhere to the structure of Spearman's original approach to latent variables as factors. For example, in regression modelling, latent variables are recast in a model that makes use of the language of 'predictors' for the latent variable (c.f. Thurstone, 1935; Yarkoni & Westfall, 2017). This is quite distinct from Spearman's work, which aimed originally at using factor analysis to clarify a definition for intelligence – it was not directly interested in *predicting* intelligence. It is helpful then to return to the scientific literature to articulate some first principles in working out what makes description, explanation, and prediction, as processes, different from each other.

2.1 Description

A description will typically involve some account of some event or phenomenon of interest, which will involve activities such as classification and identification (Boag, 2010). Boag also notes that description necessarily precedes explanation, as some identification of elements is necessary to be able to begin to distinguish the variables, relations, and processes that can be used in an explanation. Description may involve close identification of aspects of some occurrence or thing, with an ideal of remaining as theory-free, or theory-neutral, as possible. Haack (2007, p. 125) outlines a role for empirical observation with description, and she examines the constraints of subjective experiencing and perception which are both present with any event of observation. Haack notes the term "observation" may stand for an attempt to graft scientific seriousness to a process which may be either *simple*, in terms of actually looking at something and recording an account of it, or *complex*, involving chains of background beliefs interwoven in scientific theories that may have a role in, say, what occurs in an experiment. For Haack (2007, p. 128), "[i]n every perceptual event. . .there is something received, something resistant to one's will and independent of one's expectations". To engage the activity of description, though, there is a second step involved. To describe something, a cognitive judgement is needed about identity of the phenomenon - for example, deciding what characteristics are present in order to apply a category, or property-value (Haack, 2007). Instruments in psychometric practices have roles in both processes – both in the initial perception, where for example in GLVM practices, the researcher's intuitions guide the choice of a series of questionnaire items that feature in some factor analytic process, and the judgment, where in GLVM techniques, model fit may be assessed for data. Haack (2007) notes that the judgment step has different characteristics to the perception step, as it involves interpretation between the research situation and the representation made in the model. In any case, what can be seen is that the scientific activity of description may seek to track reality as it is, but cognitive processes, instrumentation, and supporting assumptions and theories all play a role in mediating any assertion made, as a description.

Connecting the above to the GLVM, Mulaik (1985, 1987) characterises the nature of scientific description, which he argues influences the development of the GLVM. Mulaik references Sir Frances Bacon's 1620 *Novum Organum* for its non-metaphysical, descriptivist approach to scientific enquiry. Mulaik (1987, pp. 412-413) then goes on to describe "Baconism" as a series of steps that the scientist completes: 1) identifying an essential form for a particular phenomenon of interest; 2) preparing a table listing the instances which have this phenomenon in common; 3) seeking out properties common to all instances; and 4) constructing a similar table of instances, where the phenomenon was absent. If a nature or property was present in the first table and present in the second table, it was excluded from the list of natures or properties considered relevant to the phenomenon in question, as the property was considered not to discriminate the phenomenon well under a logical test. Then another table was constructed, where the nature or property varied in degree for instances. The scientist checked the original list for natures or properties expected to vary in line with the first phenomenon, excluding any natures or properties that did not so vary.

This process is relevant to Spearman's (1904) original investigations into the nature of intelligence. He states an interest in identity here, which for psychological phenomena he took as constituted in uniformity of function, and, "conceptual uniformity" (p.204). The first was amenable to proof, in his opinion, whereas the second could at best be investigated using correlational techniques. In defining intelligence as that which was common amongst factors, in line with ideas about the uniformity of function, Spearman reviews 28 previous studies, looking for the degree to which conceptual uniformity was or was not present among them. For example, he notes in the work of Boas that intelligence was taken to be made up of factors including sight, hearing, memory, and 'intellectual acuteness' (scores for intelligence from teachers). However, in studies conducted by Seashore, Spearman notes that there was no correlational evidence for scores for memory, hearing, or sight being associated with a measure of intelligence. Subsequently, the definition of intelligence could be amended to exclude measures of hearing, or sight, or memory, under the Baconism described by Mulaik.

For Spearman, logic served as a test in the relationship between functional uniformity, taken as a constant for a particular property or phenomenon, and conceptual uniformity, or which tests seem to correlate best with the functioning of intelligence. A logical problem of circularity becomes apparent in this relationship, insofar as the functioning of intelligence is only evidenced by the test outcomes. The implications of this logical problem are addressed in chapter 4. For now, it is possible to note the reliance on statistical theory about correlation, in the processes that make up description, from Spearman's perspective. Correlations form part of Spearman's empirical description, but theory was involved in the idea that intelligence or ability could be quantified in the first place (Kane & Brand, 2003). Mulaik (1987) notes that description typically remains free of hypotheses, where no conjectures are at play about the nature of what is. Rather, the effort is directed to recording or representing some aspect of our encounter with reality. But what is actually the case is that hypotheses are already assumed to be true in theory that is adopted in the assumptions involved in the process of producing empirical observations (see Michell, 2004a), such as for example, the notion that ability can be quantified using certain scale structures. Without a conceptual framework to record such assumptions, future researchers may not realise the full extent of theoretical assumptions that underlie a reported research claim. As described above, such descriptions may be absorbed into explanations, the structure of which we turn to, next.

2.2 Explanation

As discussed above, description can be logically distinguished from explanation because of the prior nature of its status, insofar as description is necessary before explanation can properly begin (Boag, 2010). Explanation generally is defined as an effort to make something understandable or intelligible (Kim, 1995). Explanation in the context of science for Rescher (1958, p.285) is defined as offering a comprehensive and conclusive account of *why* a situation or a something is, as it is. In this sense, an explanation inductively or deductively establishes or demonstrates its conclusion, so that there is little doubt about fact bearing for the conclusion. In Spearman's time, explanations were rendered under what was described as the covering law or deductive-nomological (D-N) model (Nagel, 1961), having both logical and evidential elements. The logical element was the way that induction or deduction was employed in implication of the conclusion, while the evidential element involved empirical regularities or patterns observed in nature (Achinstein, 1971; Rescher, 1958).

Spearman (1904) pursues his development of factor theory using correlational analyses with scientific laws in mind - he describes the positive manifold as evidence for the "law of the universal unity of the intellective function" (p. 273). The pattern provided evidence for the functioning of a law, and thus, explained something about intelligence or cognitive ability, for Spearman. Spearman describes a likeness between physics and psychology, where for him, there existed a shared characteristic that he describes as "functional uniformity". This meant there were like reactions under like conditions, or, empirical regularities (see p. 204).

These ideas about empirical regularities are connected to notions of causation, and causation itself becomes closely associated with explanation throughout 20th century science and psychology research (see Borsboom, 2005; Haig, 2009). This shift has its origins in the demise of the D-N model in science, due to its inability to distinguish explanatory asymmetries (see Salmon, 1989), and the rise of inductive statistics and probabilistic accounts of evidence in explanations (Rescher, 1958, 1962; Salmon, 1989). The advent of probabilistic claims in science signals a shift in understanding across paradigms. There is an exit from the era of emphasis on scientific laws in scientific explanation, as it became understood that science could not have total knowledge of the set of factors and their interactions that have influence or bearing on any one situation or event (Rescher, 1958; Reichenbach, 1971; Salmon, 1989). By the beginning of the 21st century for the GLVM, the latent variable is described as representing a common cause (Borsboom, 2005, p. 68). In this account, some underlying latent variable is taken as a causal agent, influencing the scores seen in test outcomes. Such claims for causation for the latent variable are explored in detail in Chapter 4.

While causal modelling may define some of what is done with a GLVM in psychological research, use of the GLVM is far from limited to causal modelling. For example, recent research in computational psychiatry examines clustering of variables deliberately without reference to overarching diagnostic causes, and so is less explanatory than descriptive in function (Kotov et al., 2011; Krueger & Markon, 2006; van Dam et al., 2016). It is notable, as well, that in very recent philosophy of science literature there arises some challenge to the idea that all explanations must be causal in structure (Lange, 2013, Skow, 2014). Causation and explanation can and do, come apart, even though both involve interactions between phenomena. This is different from description, which aims only to characterise the occurrences or phenomena, as they are. The GLVM is used in psychology research in an explanatory capacity for invariances that are of interest to the researcher. The latent variable in the model is interpreted as a representation of some invariant phenomena, which can serve in an explanatory capacity for correlations between manifest variables. In this way there are several layers of assumptions to address in remaining transparent regarding research processes – from phenomenon to data, from data to variable, from variable to psychological model, and from psychological model to psychometric model, with different theories relevant, for each research process step. Such relations between research elements may be distinct from those that are needed in a scientific process of prediction, however, and to this, we now turn.

2.3 Prediction

Explanation and prediction can be observed to be employed in the psychology literature in interchangeable ways (Yarkoni & Westfall, 2017). There is some good reason for this, as a good explanation about why something is the case stands to facilitate a good prediction, regarding the outcomes that are expectable given a satisfactory explanation. Yet it should also be noted that there is no logical reversibility between the two. A good prediction does not necessarily entertain any answers to ‘why’ questions, at all (Rescher, 1958).

Rescher (1958) sets out some considerations for the logical differences between explanation and prediction. While prediction and explanation have been in the past identified as logically identical (see for example, Hempel & Oppenheim, 1948), Rescher (1958) demonstrates the ways they are logically distinct. For prediction, Rescher (1958, p. 286) notes that there are different logical criteria for evaluation compared to those which are relevant for explanations. The reasoned arguments only need to support a conclusion that is significantly more likely than any reasonable alternative, for a prediction. In other words, a prediction only needs to present itself as a sufficient argument, while an explanation demands a conclusive reason. Helmer and Rescher (1959, p. 32) further note it is completely possible to have unreasoned predictions. What is important for a prediction is that it can target real world outcomes, which do not demand in their structure any justifying argument.

Yarkoni and Westfall (2017) propose that psychology should orient itself away from explanation in its modelling practices, towards prediction. In their reasoning psychological science should be most invested in matching itself to real-world outcomes for individuals, and less about the accuracy of scientific explanatory conjectures. They point to several practices employed in GLVM, such as the use of goodness-of-fit heuristics for a theoretical model, to highlight the ways that such processes may fail completely to target real-world outcomes. That is, explanatory inferential processes are aimed at theory-to-model relationship, formulated in a GLVM as relations between variables, rather than at real-world facts regarding what is the case. This inferential process may lead researchers to overlook the role of background assumptions in the theory and the model, and the actual situation of real-world conditions. Maximising goodness of fit for data from a sample, for example, may result in an unrepresentative predictive model because of underlying background assumptions (Myung & Pitt, 2002). These assumptions may have roles between the phenomenon of interest and data, data and model, theory and any or all of phenomena, data, model, and situation, in which research occurs. There is no evidence of interest in prediction in the early work of Spearman (1904), however the use of statistical inference in the process of prediction plays an important role in today's psychology research, in for example, computational psychiatry (see Huys, Maia, & Frank, 2016; Chapter 8). The point that follows is that prediction has different goals, from explanation and description, and that the GLVM is employed in processes of prediction, as well as explanation, and description.

2.4 Interim Summary

What we can note from above is that it is likely that we need our conceptual framework to be broad enough to encompass any or all the scientific activities of description, explanation, and prediction, in applications of the generalised GLVM. We will observe in what follows as an historical development of the generalised GLVM a slow shift in focus, from psychological, to mathematical, to statistical influence, marking approximately distinct interests in description, explanation, and prediction. We witness a flow from endorsement of notions of scientific law and universalism for psychological processes under the economic pressures associated with eugenics, to nuanced recognition of individual differences and the means by which psychological phenomena come into being, to sophisticated but perhaps, correctly characterised, low-dimensional techniques.

3. Introduction to historical overview

In this section, the development of the GLVM is reviewed, looking to the original work of Charles Spearman, and then exploring generalisations made to his work, across time. In Chapter 4, a detailed analysis of the problems remaining for the generalised GLVM is presented, given the historical development and the recent generalized accounts. The aim of this section then is wedge open the doorway to history and find out whether psychometricians through time referenced logic in their use of mathematical and statistical modelling, as well as asking seeking answers regarding how description, explanation, and prediction have featured in GLVM through time.

3.1 Spearman (1863-1945)

In 1904, Spearman published two papers cited as seminal in psychometrics literature (see for example, Cowles, 2001; van der Maas, Kan, Marsman & Stevenson, in press). Spearman's second paper, *General Intelligence, Objectively Determined and Measured*, sets out what is of core concern to us – the factor analytical model that becomes the basis of the GLVM. In his factor analysis, Spearman was looking for a way to prove, for a given test score outcome:

Manifest score = proportion due to common factor + proportion due to specific factor.

Chapter 3 explores in detail how Spearman formulated this analytical approach, and whether the partitioning succeeds as a formal mathematical proof. Lovie and Lovie (2010) note that his data for the two 1904 papers seem to have come from a self-initiated study of three groups of school students and one of adults in Berkshire, England, where Spearman's interest was in a hypothesised overall capacity or cognitive ability. Despairing over the lack of agreement in the literature over the definition of intelligence reviewed in the first half of his 1904 paper, Spearman's aim in his empirical work was to make the concept of intelligence precise, by using what he described as measurement, in correlation techniques.

Perks (2010) notes that at that historical time, schooling had just become compulsory in England under an Act of Parliament, and because of the costs involved in schooling pressure was on to quickly and easily identify educability, in children. Influences of eugenics, then deemed as progressive, are relevant to the development of measures that could enhance the selective breeding of good, educable stock in British society (MacKenzie, 1981). Charles Darwin's cousin and architect of the eugenics movement, Francis Galton, with an interest in the way that family pedigree might influence human ability (see Galton, 1869), had set up statistician Karl Pearson, who was Professor of Mathematics and Head of Biometrics Laboratory at University of London, as Chair of Eugenics in the Galton Laboratory. While Pearson was not a member of the eugenics movement in formal capacity (Cowles, 2001), biometrics, statistics, and the goals of eugenics together shape the evolution of Pearson's methods, and inform the beginnings, of Spearman's efforts.

Spearman drew upon Pearson's correlational approach to develop his account of the common factor theory (Dodd, 1928). Spearman's (1904) correlations collected information directly from tests in the educational setting and included teacher evaluation or ranking of the child's ability within their cohort, as well as tests of abilities believed to be intrinsic to the child, for example, sensory discrimination (Spearman, 1904, 1927). As Spearman's work develops, his mathematical methods became subject to critique in both supportive (for example, Wilson, 1928) and combative ways (for example, Pearson, 1904, 1907). These critiques trace problems with study methods (Wilson, 1928), study measurement techniques (Dodd, 1928), and with the use of linear matrix algebra in the method of tetrad equations differences (see Wilson, 1933; Thurstone, 1947; more will be described on this in Chapter 3).

Whatever else remains at question regarding Spearman's innovations, his attempted inclusion of sensory discrimination as a proxy for intelligence made clear an understanding of the need for independent confirmation of the existence of *g*, beyond mathematical factorisation of manifest test scores. In Spearman's subsequent works (1927, 1946), we see him casting about for different bases by which *g* could truly and independently be said to establish the 'missing link' for experimental psychology - some biological evidence, for this idea of an overall cognitive ability. This he conceived as the work of establishing the 'identity' of *g*, a phenomenon that he reckoned occurred in a fixed amount, as mental energy for any one person. Spearman (1927) had elaborated on a theory of the laws of "noegenesis" in respect of cognitive ability. This pertained to: i) apprehension of experience – a person has more or less power to know what they experience; ii) eduction of relations, as the idea that a person has more or less power to attend to any essential relations that hold between two ideas; and iii) eduction of correlates, or the notion that once an individual can bring an original percept and its relations to mind, the person has more or less power to bring to mind a correlative idea. Spearman's assertion of the necessity of independent evidence for psychological phenomena has consistently been overlooked by psychometric methodologists through history (see Michell, 2017; Humphry, 2017; see also Chapter 6).

By 1946, Spearman, perhaps along with many of his scientific colleagues of the time, draws a hard line between what he calls philosophical psychology, which answers questions about whether psychological phenomena "really 'exist'" (p. 117), and what he describes as empirical psychology, where "the observation of '*g*' reduces itself to that of the validity of certain 'hierarchic' equations" (p. 117). Such a fault-line between the philosophical and empirical seems devised to deliver import to what in 1946 Spearman describes as "primary evidence" of definitions of ability as supported by high factor loadings in the studies he chooses to review. The mathematical proof used to justify the fault line is examined in the next chapter, and Spearman features again in Chapter 6 in reference to what is needed to define a construct in the conceptual framework, drawing on the principles he had adopted in his early work, of identity and independence. In what follows next, this thesis explores the developments derived in the beginning from some critique, of Spearman's work.

3.2 Sir Godfrey Thomson (1881-1955)

Thomson is among one of the first to challenge Spearman's factor analysis, with his proposal of what he labelled as a bonds model (Thomson, 1916). A bonds model argues that different elements of cognition support each other without any singular underlying latent variable that accounts for their development (Bartholomew, Deary, & Lawn, 2009). Bartholomew et al. note that Thomson identified the positive manifold and the hierarchical arrangement necessary for the tetrad equations in Spearman's method as elements that could completely describe some *other* pattern in nature rather than a common factor. Thomson pursued analyses with Thorndike's (1903) theory of many abilities in mind (see Bartholomew et al., 2009). This alternative theory indicated that several specific factors could simply overlap, when it comes to correlational analyses of test outcomes, so that there was no common factor underlying them. Thomson (1916) further argued that performance on any test must involve many distinct processes, such as association, retention, perception, apprehension association, reasoning, reflection, retrieval, and so on. These processes may be drawn into service in any combination of interactions that could overlap to produce the performances on different measures that make up the battery of tests. For Thomson (1916), then, there was no way that the mathematical model could discriminate to say whether the correlation matrix simply showed overlapping group factors or any single general factor. Consequently, while Spearman's factor theory appeared to sufficiently explain the positive manifold and outcomes from the hierarchical arrangement and tetrad differences analysis, it was not necessary, for either analytical outcome. Some different form of description from the common factor could equally be ascribed to the phenomena in question, and this, for Thomson, was the bonds model.

Thomson utilised some techniques that made use of random outcomes to demonstrate the failures of Spearman's analysis - showing how the 'common factor' or latent variable could be an outcome of simple chance. Spearman (1916) intriguingly notes this demonstration as a problem, and goes on to say that if a different series of outcomes had been produced, a different common factor would have resulted (Bartholomew et al., 2009). This is intriguing as it evidences some insight into the problem of factor indeterminacy that we will cover in Chapter 4, a problem that remains for GLVM to this very day. Also consistent with an understanding of factor indeterminacy, Thomson notes that it is possible to assign any value to the latent variable or factor, which leaves as many unknowns to solve the equations as had been there to begin with. Thomson (1935) further argued that positive intercorrelations among abilities is just as compatible with a theory of many common factors as it is with a theory of one common factor, so that it is always possible to add another test to the manifest variables, and still have a common factor. Logical resolution to this problem for Thomson (1951) implied maintaining a stance where independent evidence may provide evidence for factor theory, or his bonds model. Despite the evidence of disagreement with Spearman, there is shared understanding here in respect of the independent evidence needed as justification for the interpretation of a latent variable, as having underlying influence on test scores.

3.3 L.L. Thurstone (1887-1955)

For Thurstone (1934), Spearman's factor theory did not imply there was just a single common factor underlying test scores. Thurstone examined Spearman's tetrad equations, which are described in detail in Chapter 3, and concluded that tetrad equations method had not been extended far enough. What Thurstone meant here was that he could not find good reason to stop at a single factor. His engineering background and understanding of linear matrix algebra (see Thurstone, 1952, pp. 313-314) meant he saw that the tetrad approach could be extended to rankings or minors of second, third, fourth and so on, in orders. Thurston is perhaps the first theorist to utilise vector space analysis, to explain the nature of the mathematical problem in Spearman's theory (this will be described in detail in Chapter 3). With an engineering background, Thurstone intuited that one way to address the problems in Spearman's theory was to attempt to 'fix' the domain from which the variable values could be selected, or the space within which the vectors could be said to occur, and then perform a rotation of the factors (Thurstone, 1947; Jöreskog, 1979). Careful assessment and explication of the domain for any series of tests, including establishing their independence from each other prior to beginning a research study, features throughout Thurstone's lithographs and subsequent textbooks. In this way, Mulaik (1986) suggests that Thurstone's latent variable methods were of a quasi-confirmatory nature, in the vein of a "scientific realism that insisted that theoretical entities were real and to be distinguished from artefacts of method and mathematical representation." (p. 24)

Whatever else may be said about his approach to science, the repeated refrain we see throughout Thurstone's contributions is that the latent variable model or factor analysis can be used to reduce the number of variables a scientist must deal with. His awareness of the problem of factor indeterminacy via vector space analysis as described above made him cautious about stating what could be claimed following factor analysis. As he writes, he did not "know any sure way to guarantee that a given test battery satisfy this condition. We must rely on psychological intuition to assemble a test battery with sufficient restriction in content so as not to cover too many common factors" (1937a, p. 74).

Thurstone seems not to refer to the kinds of logical concerns that had been pursued by Spearman in his development of factor analysis to define a concept. Thurstone (1937b) argued that *g* was a statistical artefact resulting from the mathematical procedures used to study it, although he is careful to characterise his work as an extension of Spearman's contribution. Thurstone (1937b) points out that Spearman's technique *can* reveal just one common factor, but also that nothing prevents the application of the same method to demonstrate the presence of multiple factors, or multiple latent variables.

Thurstone's approach, with its rationale of reduction of number of variables as championed by later researchers such as Cattell (1966) and Gorsuch (1983), and its emphasis on reliability and validity procedures, becomes the basis of Kuhnian "routine science" involved with the use of GLVM techniques in psychological research (Mulaik, 1986). Thurstone's role in the professionalization of psychology in the USA also cannot be underestimated (see Brown, 1992), and his advocacy for the use of factor analytical methods cannot be understood without reference to his influence in this regard. What happens from this historical moment of the mid-point of the 20th century on is the development of algorithms, computer hardware and computer software, that facilitate the production of factor analyses. Mulaik (1987) notes that military work in nuclear physics, which developed sophisticated computational methods in linear algebra and matrix theory, resulted in development of computerized techniques to produce eigenvectors and eigenvalues as solutions for parameters for the factors in the models (see Goldstein, Murray, & von Neuman, 1959). Carroll (1953), Ferguson (1954), Kaiser (1958), and others, resolving factor or latent variable solutions in different ways, adopted these developments. Mulaik (1986, p. 27) notes wryly here: “[b]y the early 1960’s a number of computer programs were available for performing factor analyses from start to finish without the intervention of human hands, in fact, almost without the intervention of human minds.” Complex assumptions about the nature of the phenomena in question were assumed into the software programming for latent variable modelling, which did not necessarily lend themselves to scrutiny by the very researchers who were using the same software for psychometric analyses.

Whatever else is the case by this time, it was possible then for a researcher to perform a factor analysis without the careful consideration of what the factors or latent variables meant. Thurstone’s (1947) multiple factor tetrad solutions had demanded some reflection on what tests had been combined in the production of a factor or latent variable, as well as whether these tests could be considered ‘independent’ or not, in line with the assumption of conditional independence for the positive manifold. However, while the GLVM at this point remains unnamed as such, and modelling as a practice had yet to enter the picture of what a psychological researcher has as an instrument at their disposal, all the groundwork had been set and endorsed, by Thurstone and his followers. His generalisation of Spearman's model from a single factor to multiple factors potentially remains as the most important shift in the eventual emergence of the different versions of GLVM as described below. Meanwhile, let us quickly review the final contributions that have had a role in making the current versions of the GLVM possible.

3.4 D. N. Lawley (1915-2012)

Various procedures for something like what we describe today as exploratory factor analysis for continuous scale structure variables were used in psychometric practices through the 1930's, but Lawley (1940) was the first to formulate solutions for the factors or latent variables utilising Fisher's approach to maximum likelihood analysis (Jöreskog & Wold, 1982, p. 263). This was important, as there is a logical question for the GLVM regarding the meaning of any analysis if there was no way to resolve values or outcomes for the latent variable, for specific individuals. Lawley (1942) reversed the kinds of analyses done with factor analysis from the perspective of correlations between test outcomes or the positive manifold, and focused instead on a common factor model, with fixed latent variable or factor outcomes. Using this, he obtained conditions for a maximum of the likelihood function of these outcome values, hypothesising a numerical algorithm for finding the maximum point. However, Anderson and Rubin (1956) showed that in Lawley's problem, the likelihood function is unbounded. That is, Lawley had been finding unsuitable solutions for the latent variable outcomes (McDonald, 1979). We will explore some questions regarding the logic of maximum likelihood analysis in the next chapter, which addresses the nature of mathematical proof for parameters in the GLVM. What can be noted for now is that Lawley made a pathway towards efficient solutions to the problem of identifying parameters for the latent variable in terms of solutions or scores that individuals (or more likely, groups), which is still in use, today.

3.5 K.G. Jöreskog (1935-)

Jöreskog built on Lawley's proposal for likelihood functions by using statistical methods to find values for a model by developing a rapidly converging iterative procedure (Davidson-Fletcher-Powell method: Fletcher & Powell, 1963) in computer software (Jöreskog & Wold, 1982, p. 263). Up until the 1960's, Mulaik (1986) notes that latent variable analyses were conducted in the mode of exploratory factor analysis – bringing together a series of tests or item outcomes that seemed to exhibit the positive manifold, and utilising a version of factor analysis, to make a claim about a hypothesised construct as evidenced in some latent variable or underlying factor. Jöreskog's (1967) work is the foundation of the LISREL computer program which made possible a different approach called confirmatory factor analysis, as much as it allowed for new methods for resolving parameters in exploratory factor analysis (Mulaik, 1986). In confirmatory factor analysis, factor loadings and hypothesised correlations between manifest variables or test scores are specifiable a priori, and an overall fit of this hypothesised model is testable via likelihood ratio technique. At this historical moment, questions of what implications exist for different groups who undertook to complete a series of tests or items also became a concern. Simultaneous maximum likelihood factor analysis is further developed and outlined by Jöreskog (1971), which allows for analysis of different outcomes for the same set of latent and manifest variables across populations or groups, or the testing of what is now known as invariance (Jöreskog, 1971). More is said about invariance in Chapter 4. Jöreskog, as much as Lawley, has had a profound influence on what was possible in the practice of latent variable modelling. For instance, Mulaik (1986, p. 28) claims that "major problems of parameter estimation in both exploratory and confirmatory factor analysis were essentially solved", in his work.

Jöreskog however rationalises in a different way about the use of the GLVM, compared to the earlier approaches such as Lawley's. Whether in a confirmatory, exploratory, or structural equation modelling mode, Jöreskog (1979) suggests that one important reason to use a GLVM is that there are errors in measurement, for variables such as those used to represent psychological phenomena. In Jöreskog's (1979) work, we see for the first time the seedlings that become fully blown in the work of the psychometricians reviewed in the next section below. These seedlings are carried in the suggestion that the GLVM can be used to better reflect some underlying latent variable that cannot easily or directly be tracked with empirical methods. The seeds suggest researchers should use the GLVM as it provides a pathway approach to resolving which, out of the not-quite perfect manifest variables, should best represent the latent variable, while also having room for measurement error. Whatever else we might make of this shift in terms of thinking about measurement, modelling, and problems for the latent variable (e.g., the problem of logical circularity as discussed in Chapter 4), the generalisation of model types made possible by Jöreskog's innovations allowed for a shift from explanation to prediction, in logical structure, in the use of the GLVM. Because different models were testable for fit to manifest variable data values, it became possible to make an evaluation regarding which model better 'predicted' the actual manifest variable scores (overlooking the logical problem that the latent variable in the GLVM can only be *determined* by the manifest variable scores).

In psychology research, several other techniques that make use of concepts within the GLVM are subsequently historically adopted, including item response theory (Lord, 1977, 1980), structural equation modelling (Wright, 1918, 1934), Rasch (1960, 1966) theory, and latent growth curve modelling (Jöreskog, 1979), among others. Muthén (1983, 1984) also contributes methods that generalise the types of variables handled in the different types of models – whether they be of continuous, ordinal, or categorical data structure. Even with these blossoming developments, a core problem remains unresolved for the GLVM. This is the problem of factor indeterminacy connected to the assumption of conditional independence, described by fellow researchers since the earliest developments of Spearman (see Wilson, 1928), and logically foundational to the structure of the generalised GLVM. This is despite the work of great minds including Guttman (1955), Schonemann (1971), McDonald (1974), Steiger (1979), and Mulaik (2010). Maraun's (1996) concerns regarding factor indeterminacy are tracked in detail in Chapter 4. What can be noted at this point is that the developments beyond Spearman's initial proposal of common factor theory for cognitive ability have meant that the GLVM that we know today is used in situations and for purposes that are substantially different from those that were present in its initial emergence. In Chapter 3, we look in more detail at Spearman's proposal, and explore one technique that is used in the software for data analysis in GLVM to solve parameter values for the model, this being the maximum likelihood solutions approach.

4. Chapter Summary

In the historical review of the development of the GLVM, this chapter has considered the essential roles of conditional independence and the positive manifold as foundations to the model and found that logical tests were considered an important aspect of the early development of the model. A summary of the innovations presented in this chapter is in Table 2, below.

Table 2

Innovations leading to the generalised latent variable model (GLVM)

Author	**Proposal for latent variable modelling**	**What it generalises**
Spearman (1904)	Factor Theory	Matrix theory
Thomson (1916)	Bonds Model	Spearman's notion of a common factor
Thurstone (1934)	Multiple Factor Model	Matrix theory and Spearman's notion of a common factor
Lawley (1940)	Maximum Likelihood solutions for variables	Idea there can be 'solutions' for latent variables or factor parameters

Jöreskog (1968)	Exploratory and Confirmatory Factor Models	Model types
Muthen (1983)	Different kinds of variables in EFA/CFA/IRT/SEM	Data types

What can be appreciated across time is the eventual orientation to probabilistic analysis of data in models, distinct from analysis of questions of logic for the use of the GLVM. This, in part, seems connected to trends in science more generally, with a shift away from the search for unifying deterministic laws, and adoption of perspectives informed by the relativism that followed in the wake of Carnap (1950; see Michell, 2004b). The laws of thought have been introduced in this chapter in connection to logical questions, as relevant to research into psychological phenomena that makes use of the GLVM. Question has been raised about whether the goal of psychological research that utilises the GLVM can be considered to be limited to just scientific explanation. What followed was a view of how the generalized GLVM has unfolded through time, with a brief introduction to some of the rationales for a realist causal ontology for the GLVM proposed in the literature. This historical tour has brought to light some logical and methodological concerns to do with the use of the GLVM in psychological research practices. The next chapter begins by examining two accounts of the generalised latent variable model, and then opens the question of mathematical proof, in the early development of features that are foundational to the GLVM.

Chapter 3: Historical Moments in the Development of the Latent Variable Model and Mathematical Proof

1. Objectives

The objectives of this chapter are to present two generalised accounts of the latent variable model, looking to commonalities and differences between the accounts, and then secondly, to examine the role played by mathematical proof in the history of the latent variable model, with particular attention paid to Spearman's original factor theory and his claims for proof for the theory. We turn then to the more modern practices with maximum likelihood solutions for model parameters, with proof attempted for these by Ronald Fisher (1922), which were incorporated into latent variable modelling beginning with Lawley (1940). The previous chapter focused on the interpretation of latent variable modelling as a practice through time, in order to understand something of the domain over which practices are relevant. The first half of this chapter looks to two attempts to summarise the domain into presentations of definitions for the generalised latent variable model (GLVM). The second half of this chapter returns to questions regarding the mathematics underpinning the latent variable modelling technique. In this distinction, we follow Nunnally (1978, p. 328), separating out treatment of latent variable modelling as a method, and latent variable modelling as a mathematical practice. This is done prior to re-integrating these two approaches in the conceptual framework, in Chapter 6. As in the previous chapter, close attention is paid to the logical apparatus relevant to the kinds of thought processes involved in the development of scientific and psychological knowledge.

2. Development - the Generalised Latent Variable Model (GLVM)

In Chapters 1 and 2, it was noted that Spearman's factor theory was incorporated into computer software from the mid-20th century. As the computing power of software developed, different statistical methods were generated for conducting analyses of data in latent variable modelling forms. These distinct forms involve different relations between variables and variable types, different handling of error associated with a variable and/or an overall model, and different approaches for finding parameter solutions for the model. In the second half of this chapter, the viability of maximum likelihood parameter solutions as informed by the original work of Fisher are investigated in detail.

The review of Fisher's attempts at proof for maximum likelihood statistical solutions demonstrates a pattern like the one that will be discussed regarding Spearman's factor theory – proof was attempted, but never achieved by Fisher, for all statistical cases (Aldrich, 1997). In ways like the resistance to Spearman's factor theory as discussed in Chapter 2, some psychometricians had expressed disquiet about the maximum likelihood formulations of solutions for latent variables (Kaiser, 1976; see also Jackson & Chan, 1980). Other formulations for parameter solutions as estimates are possible, as, for example, with partial least squares (Bentler & Weeks, 1980), or diagonally weighted least squares (Li, 2016). It is nevertheless not easy to know in advance what solution method would serve best for the psychological phenomena in question (Breckler, 1990), and no proof exists for maximum likelihood, meaning we have no reason to preference it from a mathematical perspective.

Further, there are a number of methods to derive the latent variables, such as principal axes (Kelley, 1935), or the centroid method (Thurstone, 1947)[4]. The developments beyond Spearman's factor theory also feature different model structures as discussed in Chapter 2, such as the multiple factor theory of Thurstone (1947), where there was a rationale for more than one latent variable, in the model. Different types of variable structure now appear in latent variable modelling processes, so that we end up with

[4] Note the principle components method of Hotelling (1933) is excluded from this analysis, as it is not a method that truly involves a latent variable that is formulated under the assumption of conditional independence.

different combinations of continuous and categorical structured manifest or latent variables as seen summarized in the table below:

Table 3

Analysis types for continuous and categorical scale structure manifest and latent variables

Latent variables	Manifest variables	
	Continuous	Categorical
Continuous	Factor Theory (Spearman, 1904; Thurstone, 1935)	Item Response Theory (Lord, 1952; Rasch, 1960)
Categorical	Latent Profile Analysis (Gibson, 1959)	Latent Class Analysis (Lazarsfeld, 1959)

Thurstone (1947) introduced a decision point that was not present in Spearman's original theory. The researcher now had to choose how many variables best accounted for the correlations or columns in the matrix. Developments by Guttman (1954), Cattell (1966), Kaiser (1960), and others appeared in computer software that allowed for fast processing of substantial correlation matrices and associated statistical output in exploratory factor analysis (EFA; Fabrigar, Wegener, MacCallum & Strahan, 1999). These techniques however support different decisions, in terms of factors to retain, and different solutions for factor parameters (Thompson, 2004). Researchers now had to decide what correlation formulae to use, what cut-offs to apply for eigenvalues where they were utilised in factor retention decisions, and whether to use a scree plot, the residual correlation matrix, or parallel analysis, in factor retention decisions[5]. They had to decide whether to rotate the axes of analysis in order to make the factors understandable and choose some method for resolving factor scores (see Thompson, 2004, and Chapter 4 for more

[5] Note that often it is the case that software will provide a default regarding the decisions in application of assumptions – an inexperienced researcher may not realize that default values have been applied (see Kline, 2012).

description and evaluation of the processes used in these decisions).

Software facilitated the development of other techniques of examining relationships between manifest and potential latent variables. With the computer program LISREL, Jöreskog (1971) instituted structural equation modelling (SEM). SEM includes both a directed relation between variables, as well as a structural model for how the variables connect. Structural modelling is otherwise known as confirmatory factor analysis (CFA: Jöreskog, 1969). This involves postulating or hypothesising about relationships between variables, making choices about an error structure for the variables and the model, and testing these hypotheses considering a set of data, to evaluate the way that the data does or does not fit to the model. Both EFA and CFA have been conceptualised as falling under the rubric of the Generalised Linear Model (GLM: Nelder & Wedderburn, 1972; Thompson, 2004). The GLM was a development on the general linear model used, for example, in ANOVA and ordinary linear regression, which allowed for variation of the assumption of multivariate normal distributions for error terms (Nelder & Wedderburn, 1972). Such a model has a link function g, a probability distribution from the exponential family ($E(Y)$), and a linear predictor ($ax + b$), in the form $g(E(Y)) = ax + b$.

In generic form, as an extension of the GLM, the GLVM substitutes a latent variable for the x parameter. Distinct accounts of the mathematical structure of the GLVM appear in the literature, and close examination reveals that different assumptions are associated with each formulation. In what follows, two accounts of the GLVM developed over the last forty years are presented, looking to the commonalities and differences between the proposals. Specific focus is maintained on the assumptions and general guiding principles suggested by the developers of these approaches, with a view to assessing how these might affect reported research project outcomes, in psychology. This is completed with a view to characterising the logical structures for each of the generalised models, to understand something of the thinking of the developers with respect to the constraints expressed by the model.

2.1 Rod McDonald (1999) Test Theory

Roderick P. McDonald (1928-2012) completed his doctoral dissertation in 1967 and in this introduced an innovation to latent variable modelling: the use of nonlinear techniques to model the relations between variables, in a way that combined common factor theory and IRT (McDonald, 1967, 1987). In this work that depended on the technique of harmonic analysis[6], he explored the question of whether a categorical structure had been incorrectly assigned to latent variables when a continuous scale would be a better representation of the structure given a set of data (Wainer & Robinson, 2007). From the outset, McDonald's work demonstrated the integration of different model forms and variable structures, with focus devoted to IRT.

In his 1999 book *Test Theory,* McDonald builds his account of the unified GLVM by considering psychometric test theory as practice. The specific focus on test production means that he maintains focus on scale structure decisions, generalizability and reliability questions, as well as examination of the kinds of items that served as indicators for manifest variables. Specifically, McDonald considers the kinds of items usually combined to formulate some sort of psychometric assessment, and explores how these could measure a single attribute, drawing on psychometric reliability, validity, generalisability, and item discriminability practices, as they existed at the time under IRT. McDonald remains concerned throughout his body of work with the concept of *unidimensionality*. Unidimensionality expresses the idea that a latent variable must describe a single phenomenon, even though it may be assessed by aggregating responses across substantially different items in an assessment or questionnaire (McDonald, 1967, 1999; Ziegler & Hagemann, 2015).

[6] Note that either harmonic analysis or a link function can be used to express a generalized latent variable approach; see Maydeu-Olivares (2005).

2.1.1 McDonald - The essential definition and model

McDonald (1999) sets out the general model citing it as Spearman's original. On examination of it in comparison to the basic form of factor theory introduced in Chapter 1, McDonald's model is more complex than Spearman's. It involves IRT from the outset:

$$X_j = u_j + \lambda F + E_j \qquad j = 1, . . . , m.$$

Here, X_j is a random participant's score on the *j*th item, and *F* is some value or measure of the latent variable. *E* is the participant's value or measure for the specific item, and *j*. u_j is identified as item difficulty. E_j is described as the positive or negative movement of response for the item, given an expected response for the psychological phenomena in question[7] (see McDonald, 1999, p. 78). The λ loading is understood as the amount of difference in the item score, given a unit of difference in the latent variable, or the "discriminating power" (p. 78) of an item. Drawing on the assumption of unidimensionality, where there are several items that have high λ loadings, a homogenous test is defined as one where the item responses fit to a single factor model. In terms of the model presented above, we can note that the originating assumptions are those of IRT, and involve additional assumptions beyond those of Spearman's (1904) factor theory, including: i) a supplementary model of an item response function for each individual, and; ii) a probabilistic additive equation for item difficulty and person ability (see Michell, 2008). These take us beyond the simple patterns of correlation, of core interest, to Spearman.

2.1.2 McDonald assumptions, innovations, logic, and constraints

For McDonald (1999), the point of latent variable modelling is in aid of test construction. A concern carried through in his later work is the question of proscribing a

[7] Note that this is very different from Spearman's common factor analysis as we have seen in in Chapter 3, which by Spearman's own account implied that unique factors could only be positive.

domain for test items for a test of an attribute (see McDonald, 2003). This becomes important in solutions to the problem of factor indeterminacy discussed in Chapter 4 of this thesis. Briefly, McDonald follows in the tradition set by Guttman (1953, 1955) to consider test items as chosen for assessment of a psychological phenomenon, or in his terms, attribute, as coming from an infinite behaviour domain or universe of content (McDonald, 1999, p. 103). Such a conceptualization of test items provides a means to connect some psychological phenomenon to the actual items for which a score or indicator is calculated in representation of some aspect or property of the psychological phenomena in the manifest variables. This provides some rationale for adding items to a specific indicator or structure, of a test. McDonald sees much more than conceptual connection as warranted. In McDonald (2003), for example, a detailed account of assessing test items in virtue of the meaning of the item in relation to the domain of the phenomenon plays a role in the construction of the infinite behaviour domain. This would include examining items and deliberately including assessment items that both support and contradict the attribute structure, or evidence discriminability, of items. The conceptual framework developed later in this thesis is designed with reference to McDonald's proposed solution for factor indeterminacy, extending the idea of domain specification across elements in a network representation of the research project including for example theory, models, variables, data, and phenomena.

In *Test Theory* and in later work, McDonald problematises a causal interpretation of the latent variable model – the idea that a psychological phenomenon had a role in "causing" test scores (McDonald, 1999, pp. 368-371). Logically, for McDonald, a cause must be uniquely or independently identifiable in character. Evidence for the directionality of the influence of the cause must come from extra-statistical sources (p. 370), and McDonald notes that there remain philosophical concerns for how a latent variable constructed from an infinite behaviour domain serves in causal analyses. The philosophical conundrum arises because if it is possible to add more test items for a latent variable, then it is also impossible to say that the common factor or latent variable is uniquely identified. Unique identification is necessary, for attribution of a causal relationship between the latent and manifest variables.

One other key point is that McDonald (1999, p. 79) states that what are usually characterised as local independence assumptions necessary for the latent variable model are really “conditions” that are a part of the definition of the model. They are not assumptions *per se*. Local independence as a condition acts as a *constraint* on the model, and as a constraint, it has a role in the meaning of the latent variable. Specifically, the local independence assumption entails a kind of boundary setting for the mathematical latent variable:

> Latent traits/common factors are defined by the principle of local independence. That is, a functional definition of a scalar or vector latent trait is contained in the statement: Conditional on a fixed value of a m-vector of latent traits, the p components of a vector of empirical variables are a) uncorrelated (weak principle) or b) statistically independent (strong principle). The principle also defines the dimension m of the latent space. (McDonald, 1996, pp. 595-596)

There is a strong role suggested by McDonald for this dimension setting for latent variables in solving problems of factor indeterminacy. McDonald (1981) states there is no mathematical basis for a logically necessary connection between the mathematical conception of a unidimensional latent variable as defined under, say, a perfect ratio-structure scale, and an hypothesised relation to an empirical set of questionnaire items that notionally measure just one common property of examinees in a given population. Factor indeterminacy is discussed in detail in Chapter 4 of the present thesis.

Finally, McDonald (1999, p. 79) also addresses scale structure for latent variables. He states that we are "free to consider" (p. 79) the scale as continuous, consistent with a mean of zero and a variance of one in the population. From an empirical perspective, however, we may expect that there are good reasons to question whether scale structure for psychological phenomena could be free to vary, as McDonald believes. More is discussed in respect of this problem with relation to the definition of measurement in Chapter 6 section 2. What will be important in any regard is a clearly structured relational system that links the psychological phenomenon, to a concept, to a construct, and to a variable. This set of relationships and their role in the conceptual framework is discussed in Chapter 5.

2.2 Bartholomew, Knott, & Moustaki (2011) Latent Variable Models and Factor Analysis

Bartholomew, Knott, and Moustaki (2011) *Latent Variable Models and Factor Analysis: A Unified Approach* is a later edition of a text originally published by Bartholomew (1987). Much of what follows maintains attention to the work of Bartholomew, specifically. David J. Bartholomew (1931-2017) worked as a mathematical statistician rather than as a psychometrician, but his work is both cited in and draws on psychometric literature, with specific reference to the work of Spearman (Bartholomew, 1984; Bartholomew, 1995; Bartholomew, et al., 2009). Bartholomew advocates throughout his work for the GLVM as a statistically driven data reduction strategy, over and above anything else (Bartholomew, 1980, 1984, 1987; Bartholomew et al., 2011). Noting that the "theoretical foundations are somewhat obscure and subject to dispute" for the latent variable model (Bartholomew, 1984, p. 221), Bartholomew describes latent variables as random variables, or constructs of convenience that do not have real status, but rather facilitate our comprehension of the phenomena in question, because they reduce dimensionality. For Bartholomew, we do not need to have an advance hypothesis about a latent variable, but rather, we can use the model to let the data tell the story – a story about both existence of the latent variable, and scale structure suitable for it (see Bartholomew et al., 2011, p. 10).

Bartholomew evinces a perspective distinct from that of McDonald regarding dimensionality for the latent variable. For McDonald as described above, clear specification of the domain over which a latent variable operates is vital in overcoming the problems to do with factor indeterminacy discussed in Chapter 4. The closing chapter of Bartholomew et al. (2011) articulates an interpretation of latent variable modelling as a purely statistical process, where the statistical latent variable is treated as not real in the sense of being independently verifiable, “nor is it intended to be” (p. 245). Latent variables for Bartholomew et al. thus should be treated as random variables, which can be interpreted as variables where all observations are missing. Given that a random variable simply defined is a mapping from the sample space to real numbers (Easton & McColl, 1997), articulating the relationship conceived between the phenomenon of interest and the latent variable would be expected to be vital, but receives no comment in Bartholomew et al. (2011).

2.2.1 Bartholomew et al. (2011) - essential model and definition

For Bartholomew et al. (2011, p. 6), it is possible that latent variables may remain non-determined by a researcher prior to modelling, including both the number of them and the scale type relevant to the latent variable. The GLVM analysis then constitutes the evidence for some latent variable given to be expressed in the data. The conditional distribution of a latent variable or latent variables y given the data that has been obtained in manifest variables x is used as the basis for beginning to formulate some idea about the nature of the latent variable, as follows:

$$f(\mathrm{x}) = \int_{Ry} h\,(\mathrm{y})g\,(\mathrm{x}|\mathrm{y})dy$$

where h(y) is the prior distribution of y, g(x|y) is the conditional distribution of x given y and R_y is the range space of y. To use the information gathered in the manifest variables x1, x2, x3, etc., we examine the conditional density of y given x:

$h(y|x) = h(y)g(x|y)/f(x)$

What is needed then to analyse a latent variable is the assumption of conditional independence over the manifest variables. Bartholomew et al. (2011, pp. 7-8) note:

> "it is misleading to think of it as an assumption of the kind that could be tested empirically because there is no way in which y can be fixed and therefore no way in which the independence can be tested. It is better regarded as a definition of what we mean when we say that the set of latent variables is complete. In other words, that y is sufficient to explain the dependencies among the xs."

In other words, conditional independence is a foundational assumption for the latent variable model that explains the way that the manifest variables are related to each other.

The final model for Bartholomew et al. (2011) then takes the form:

$$f(x) = \int h(y) \prod_{i=1}^{p} g_i(x_i|y)dy$$

for *p, h, and* g_i, where *p* is the number of latent variables, *h* is the prior distribution and g_i is the conditional distribution. Bartholomew et al. (2011, pp. 8-11) note factor indeterminacy as a problem for the GLVM, in the sense that any attempt to change the latent variable value has an indeterminate effect on what changes in the prior distribution, for *h,* and the link function, or *g*. There is thus no way to determine the right way to change these functions given a change in the latent variable, just from the mathematical equations. Bartholomew et al. make use of the assumed freedom in decisions on, for example, scale structure for the latent variables: "the indeterminacy of *h* leave us free to adopt a metric for y such that *h* has some convenient form" (p. 9). The problem of factor indeterminacy is discussed in much more detail in the present thesis in the next chapter. It can be imagined that in the interests of reducing quality uncertainty associated with research outputs, that some rationale for adoption of a particular scale structure for a latent variable might increase the meaningfulness of the research outputs, regardless of the freedom that exists in running the data through the mathematical model. The conceptual framework articulated in Chapter 6-8 aims to support such clarification.

2.2.2 Bartholomew et al. (2011) assumptions, innovations, logic, and constraints

Bartholomew et al. (2011) note the problem of *reification* for latent variables. Reification is the idea that latent variables inherit the properties of real entities, distinct from remaining as constructs that appear in a model of properties or entities. Bartholomew and colleagues suggest that the methods do not demand that we decide one way or another regarding the status of the latent variable. This is like the proposal of McDonald (1996), who suggested we could postpone questions about what it is that underlies estimation and use the GLVM to arrive at estimates for property values for some hypothesised phenomenon.

One other vital aspect of Bartholomew et al.'s (2011) approach to the GLVM is the idea that individuals, such as study participants, can be positioned on the latent variable. Bartholomew et al. (2011) note that this relies on a posterior distribution. This means that the latent variable position for an individual is determined after the scores on the manifest variables are obtained. A further assumption is required for this, which is that the estimation about which latent category an individual belongs to occurs with minimal loss of information, given the use of maximum likelihood methods across a fixed number of latent classes. For Bartholomew et al. (2011), these fixed classes are determined by the data within the latent variable model. Circularity is thus evident, in this account. We will see more commentary on the limitations of a posterior distribution interpretation of the latent variable model in the section on factor indeterminacy, and more commentary on the problem of circularity, in Chapter 4.

Connecting back to the definition of the GLVM seen in Chapter 1 endorsed by Skrondal and Rabe-Hesketh (2004, p. 95) as consisting of a response model, structural model, and error model, we can note that Bartholomew et al.'s (2011) reliance on a posterior distribution interpretation of a latent variable position serves to act as a constraint consistent with the position expressed by Skrondal and Rabe-Hesketh (2004) on the GLVM. This is because the interpretation of the latent variable as a function of a prior distribution imposes a set of assumptions such as standardised normal distribution and errors with an average of zero, which act as constraints on the possible values that may be returned for the latent variable, given values of the manifest variables.

2.3 Interim Summary

So far, we have seen two distinct versions of the GLVM. McDonald begins with an integration of item response theory and common factor theory, with the principle that the GLVM expresses something about the properties of a test rather than a psychological process. Bartholomew, on the other hand, begins with ideas about statistical distribution and random variables, with the conviction that a latent variable model may tell us something about an individual's position or level of a latent variable, given some other data for some other variables. For McDonald (1999) the infinite behaviour domain will play a key role in resolving the problem of factor indeterminacy for the GLVM, discussed in detail in chapter 4. For an analysis following Bartholomew et al. (2011), what would be expected instead is some account of the data reduction achieved following GLVM application, with latent classes decided on a posthoc basis. The conceptual framework will need then to provide some explicit guidance as to how the latent variable can be understood to say something about a psychological phenomenon in broad enough terms to cover both potential circumstances. We have seen how both McDonald and Bartholomew characterise conditional independence as *defining* the GLVM, in common – it serves as a condition or constraint on what can be said, following analysis. This role of constraints is an important one for the conceptual framework and is discussed in more detail in Chapter 6.

3. Applied mathematics - turn of the 20th century

Research occurs within “patterns of disciplinary practice” (Danziger, 1990, p. 301), and is affected by cultural convention as much as it is informed by norms of scientific practice within specific fields. The patterns reflect geo-historically situated ideologies, concepts, and technologies (Teo, 2006). One advantage of geo-historically contextualising the research situation in a representation under a conceptual framework is that the relationship between the present project and its procedures, and those procedures and practices conducted in other contexts and times becomes amenable to further investigation. One example of a change in a pattern of disciplinary practice is that which occurs in the field of mathematics and the shifts in understanding about the constraints that apply to mathematical proof that occur at the turn of the 20th century, just at the time when Spearman develops his factor theory and a proof, for it.

3.1 Proof at the turn of the 20th Century

In formulating his factor theory, Spearman relied, at least in part, on mathematical conventions learned in his engineering training as a Sapper in the war (Lovie & Lovie, 2010). In the 1904 paper, he begins with correlational analyses with foundations in the work of Galton and Pearson as described in Chapter 2. Ultimately, what he relies on for his factor theory is what he describes as a “proof” based in linear matrix algebra, first published as Hart and Spearman (1912). The complete history of linear matrix algebra traces back to Leibniz (1693) and the use of determinants in mathematics in the late 17th century, as well as Cramer’s rule to solve linear equations using determinants, developed in 1750. Nearly 100 years later, in 1848, Sylvester introduces the term *matrix* to mean an array of numbers, and development of the theory of matrix multiplication follows in the work of Cayley (1855). Vector space analysis, particularly with the root of rank-one as Spearman has used in his setting out of proof, has clearest precedent in the textbook for mathematics and physics of Wilson and Gibbs (1901). Spearman's use of linear matrix algebra is described in the section on his tetrad equations method, below. In what follows, the conditions of mathematical proof are first set out, before examination is conducted of Spearman's use of linear matrix algebra, exploring the question of whether his use constitutes a mathematical proof.

3.2 Mathematical proof - its structure and nature

What *is* mathematical proof? Defined, a proof is "a sequence of statements such that every member of the sequence is either a basic a priori statement or a statement which follows from previous members of the sequence in accordance with some apriority-preserving rule of inference" (Kitcher, 1984, p. 38). Mathematical proof will typically entail reliance on: i) definitions, which explain "the meaning of a piece of terminology" (Krantz, 2007, p. 242), usually in concepts that are understood as self-evident, more fundamental, or likely to be already known; ii) axioms, which are mathematical statements formulated using terminology set in the definitions, and; iii) theorems, which are mathematical statements that can be proven, using, for example, links of logic (Krantz, 2007, p. 243). Mathematical proof has historically been identified with logical proof (Hamami, 2018). Formalism, or statement in a formalised notation, remains today as a goal or outcome sought for a mathematical proof, even though mathematical proof and logical proof are now treated as two distinct processes (Hamami, 2018), following the innovations of Frege discussed in Chapter 1.

Logical proof typically follows sequential steps, such that deductions about a set of premises falls directly out into a conclusion with a clear set of steps that follow each other in valid order. Mathematical proof, on the other hand, does not necessarily conform to these strictures (Hamami, 2018, p. 36). Mathematical proof typically eliminates self-evident steps, but these may be encountered as "proof gaps" by non-experts (Hamami, 2014; Fallis, 2003). Proof gaps occur when the construction of a given proof fails to conform to what is needed for the proof to continue to be counted as an actual proof under mathematical consideration, as construed by the mathematical community (Fallis, 2003, pp, 52-53). Fallis explores these as inferential, enthymematic, and untraversed gaps, while Hamami (2014) adds a fourth type of gap: a rigor gap. In what follows, the definitions of gaps as set out by Fallis (2003) and Hamami (2014) are clarified, and Spearman's output is then examined, in light of these considerations.

Inferential gaps occur when the sequence of propositions that the mathematician proffers as proof do not conform to what is needed to count as proof (Fallis, 2003, p. 51; Davis, 1972, pp. 260-262). Fallis identifies fallibilism at work here, in that mistakes occur in human reasoning about what is needed, to count as a full proof. Overall, when logically tested, the question of whether a proof with inferential gaps is a proof is answered as false, given discernible fallibilism in the reasoning process.

Enthymematic gaps occur where mathematicians intentionally leave out steps in the inferential process when they report their proof because there is some body of assumed knowledge for definitions, axioms, or theorems (Fallis, 2003, p. 53). While an inferential gap may be an enthymematic gap, they are not identical (Fallis, 2003, p. 54). Enthymematic gaps are characterised by their motivated nature. When tested for logical consistency with reference to the laws of thought, what happens here is that the outcome during some step of the reasoning returns a provisional false assessment regarding whether the proof counts as a proof, even though the proof may, with proper steps included, hold as true.

Untraversed gaps occur where the proof producer does not take necessary steps in the construction of the proof to ensure that there are grounds for the conclusion in the premises contained in the definitions, axioms, and theorems (Fallis, 2003, p. 57). Under a test of logical consistency, a proof has been offered, but it cannot be secured with true status, as a proof - this constitutes a mistake made in respect of the conclusion of the proof.

Fallis (2003, p. 60) notes a fourth type of gap which may arise, a gap that is filled in time via the development of mathematical knowledge. Hamami (2014, p. 22) labels this as a **rigorisation** gap. This type of gap occurs because of the shifts in historical knowledge within mathematics that emerge over time (see also Kleiner, 1991). An example of this is where Euclidean proof involved reading off diagrams (Hamami, 2014; Detlefsen, 2008), yet the discovery of non-Euclidean geometry with the work of Gauss at the beginning of the 19th century invalidated some of the old proofs - they were shown to no longer hold, when viewed from a different perspective. When tested for logical consistency, what is important to note here is that the proof was able to be counted as true at a certain historical moment but was later identified as false or questionable under the continued development of practices in the body of knowledge of mathematics.

One aspect of the production of proof that is important is the community of practitioners who have a role in evaluating the proof[8]. This body of practitioners provide the testing ground of the veracity of a proof (see Ferreriós, 2016), and may provide solutions for the proof gaps, when these are located. Fallis (2003) notes an example of enthymematic proofs in the work of Gauss, relevant to our quest for utilising statistical methods in psychological research. Fallis (2003) describes Gauss's assumption regarding the continuity of the algebraic curve (Struik, 1969) as an instance where a mathematician proposed that proof could be found, but who did not provide the proof. Ostrowski (1920) goes on some time later to fill in this proof. What is noticeable with such a pattern is that justificatory processes can be used within the mathematical community that allow for

[8] We return to the theme of the role of the research community in the production of research in Chapter 7.

practice to continue, which utilises some unproven result, with results becoming available for a proof in a non-linear fashion, with respect to time. It is perhaps this that Spearman had hoped for, with his proposal of factor theory, the subsequent production of the tetrad equations method, and the unfulfilled efforts that follow, to find a determinate proof for the factor theory method (see Spearman, 1927, 1948).

3.3 What about Spearman and proof?

In Spearman's time, logicism prevailed when it came to filling in the gaps for proofs (Crane, 2012; Demopoulous, 1994), with the expectation that slow deductive steps and first-order logic derivations were needed when it came to producing a proof. More recently, following the probabilistic revolution and the accompanying shift to relativist pluralism in science and philosophy of science noted in Chapters 1 and 2, there is evidence of a turn towards looking at processes such as explanation and social constructivism (see Kleiner, 1991) as relevant to mathematics and proofs. The pattern that is described above which resulted in proof for Gauss's work came to be recognised by practitioners and then incorporated into practices. The new proposals for contextualised proof are distinct from a requirement that proof conform to the old structures of first-order logic and has involved a re-articulation of mathematics as practiced with the understanding that it is fallible, rather than indefatigable (Kleiner, 1991; Ferreirós, 2016). Outcomes of this historical shift are that today, proof practices may better be described as local, or particular to the community of mathematicians, rather than global, as would be the aim of a logical proof which is recursive to, for example, first-order logic (see Hamami, 2017; Ferreirós, 2016). When we consider the nature of proof for mathematics, though, what is noticeable through time is recursion to first-order logic structures, and processes to secure consistency of proofs. As Detlefsen (2008, pp. 17-19) makes clear, even though the most recent outcomes in proof production show us that proof is much more a *performance* than a *conformance* to formal protocol, explanatory or logical consistency remains as a desired property for all mathematical proofs. The theme of local versus global approaches to science and research reappears in Chapters 5 and 7 of the present work, as does emphasis on explanatory consistency. In what follows we use the four types of proof gaps discussed above to analyse the proofs offered first by Spearman, for his factor theory, and second by Fisher, in his maximum likelihood methods.

4. Spearman

As noted in Chapter 2, at the time of publishing two 1904 papers that are described as “epoch making”, even by those who challenged Spearman’s ideas (see Thomson, 1947, p. 343), Spearman had no formal academic qualification. Spearman’s autobiography discusses dabbling in some Wundtian-style experimental research into what he describes as ideo-presentation (Spearman, 1930/1961, p. 310), yet subsequent researchers suggest Spearman’s true experiments were those that drew on mathematical engineering formulae as tools for analysing information gathered and manipulated in efforts to track psychological phenomena (Horn & McArdle, 2007). These mathematical experiments emerged as Spearman began to work towards clarification of the concept of general intelligence (Spearman, 1904, pp. 205-206).

The 1904 paper draws on correlational techniques developed by social biologists Galton and Pearson (Spearman, 1904, p. 225) applied to data from schoolchildren as well as one community sample. The extent of application that Spearman foresaw for his concept of general intelligence though was not limited to schooling. When writing with Hart, he envisaged that it was:

> not altogether chimaeric to look forward to the time when citizens, instead of choosing their career at almost blind hazard, will undertake just the professions really suited to their capacities. One can even conceive the establishment of a minimum index to quality for parliamentary vote, and

above all, for the right to have offspring. (Hart & Spearman, 1912, pp. 78-79).

In the 1904 paper, Spearman is attempting to use quantitative methods to secure a definition for the concept of intelligence. Spearman's paper does not make use of factor analytical techniques as we recognise them today. Rather, it relies on his conceptual analysis of the term intelligence as it had been investigated in earlier papers, and on quantitative evidence with respect to hierarchical arrangements of adjusted correlations between the varied tests, he adopted from review of those papers. Spearman's training in Wundt's laboratory seems to have lent to a faith that finding "Functional Uniformities" in experimental psychology could be followed by "conclusive proof" about the existence of the phenomenon (Spearman, 1904, p. 205). Following his correlational analyses on tests conducted across two groups of children at a Birkshire school, one group of boys at a Harrow preparatory school, and one group of adults, Spearman concludes he has provided evidence for the existence of general intelligence and what he labels as "General Sensory Discrimination" (p. 272):

> [w]e reach the profoundly important conclusion that there really exists a something that we may provisionally term "General Sensory Discrimination" and similarly a "General Intelligence"

However, he does not halt there:

> [a]nd further that the functional correspondence between these two is not appreciably less than absolute (Spearman, 1904, p. 272).

In his terms, this relationship constitutes a "proof", which should result in reproducibility of the relationship, "in all times, places, and manners" (p. 272). Spearman had made note of the deficiencies of earlier studies of intelligence in his work, stating the earlier studies often lacked a precise quantitative expression, for intelligence, and they did not handle in standard ways: i) probable error; ii) observational error, or iii) removal of irrelevant factors. For the last of these, Spearman did attempt to adjust his results for factors that may affect correlations such as age and practice effects, using experimental means to control for these. For a precise quantitative expression, he proposed that the method of attenuation he had developed in the other 1904 paper could be used to adjust correlational findings for both probable errors and observational errors such that "the total effect of all such errors can be measured en masse and mathematically eliminated" (Spearman, 1904, p. 226).

A good deal of Spearman's conclusions in the 1904 paper however have gone on to be challenged in time (see section 4.1, below). Rigorisation, to return to our types of proof discussed in the section on mathematical proof above, has served to challenge the fixity of Spearman's hierarchy of sensory discrimination and the correlation with intelligence (see Thorndike, Lay, & Dean, 1909). The notion of a single adjustment for error or attenuation that accounts for all types of error is also challenged by subsequent developments of latent variable models that make exact use of different types of error (see for example Weschler, 1950; Lawley & Maxwell, 1971). Spearman however expresses appreciation for the need to continue to produce more evidence for the existence of the common factor – in his terms, this must be quantitative evidence. We turn now to look at the subsequent attempts at achieving that, and its problems in terms of mathematical proof, in his work originally conducted with Hart and then Holzinger, from 1912, onwards.

4.1 The tetrad equations method

In the tetrad equations method, Spearman is making use of his background in engineering and a technique that was not known to psychologists at that time, but which may have been familiar to engineers, of linear matrix algebra (see Lovie & Lovie, 1996). The technique makes use of a convenient mathematical property of correlations arranged in hierarchy of value in a linear matrix, insofar as the cross-products of any two sets of sequential correlations in the hierarchy should be equal. This means that when cross-products are subtracted from each other, the net result should be zero (see Cowles, 2001; Hart & Spearman, 1912).

What Spearman did next was test data that he collected in this tetrad equations model, examining whether his resulting outcomes or tetrad differences were zero (see Spearman, 1927). For Spearman, in the case that the data fit the model with a set of cross-products that subtracted from each other equalled zero, he took this to indicate that he had found a common factor or latent variable. This common factor coincided with a rank order of one in the linear matrix of correlation outcomes. Such a test was determined by Spearman to constitute proof that a latent variable or common factor was distinct from any other factor that corresponded to the tests in question (Spearman, 1927).

Spearman gives his own account of the technique he uses in the 1927 book, *Abilities of Man* (pp. 74-75):

> whenever the tetrad equation holds throughout any table of correlations, and only when it does so, then every individual measurement of every ability (or of any other variable that enters the table) can be divided into two independent parts, which possess the following momentous properties. The one part has been called the ‘general factor’ and denoted by the letter g; it is so named because although varying freely it remains the same for any one individual in respect of all correlated abilities. The second part has been called ‘the specific factor’, and denoted by the letter s. It not only varies from

individual to individual, but even from any one individual ability to another. The proof of this all-important mathematical theorem has gradually evolved through successive stages of completeness and may now be regarded as complete.

4.2 Theorems proven?

Wolfe (1940) described a prevailing confusion in the literature over what the tetrad difference truly meant. Spearman claimed it meant that individual scores or measurements were divided into two parts - one part related to the common factor, another connected to the specific factor plus error (Hart & Spearman, 1912; Spearman & Holzinger, 1925). However, as Cowles (2001) notes, while it is the case that the first column of rank one ordering *can* provide evidence for a single general factor, there is nothing in the mathematical structure to say that it *should* or absolutely does provide evidence for a single general factor. Thurstone (1938, 1947), in fact, begins to build his argument for multiple factors by simply extending the number of minor determinants or rankings. With a two-factor theory, Spearman simply stopped at minor determinants of order two, but there is no natural limit for the number of orders considered. Thurstone makes use of other rankings, to provide an evidential base for a greater number of common factors.

Thurstone's writing always evidences careful respect for Spearman's work, even while it offers innovations based on it. Pearson and Moul (1925) on the other hand took a different tack, challenging Spearman's claims for a "Copernican revolution in psychology" (p. 291) and stating that the tetrad equations proof remained far from proven. They did this by highlighting firstly some problematic assumptions necessary for the tetrad method, such as reliance on normal distribution for each of the factors or latent variables, and linearity of relationship between the variables. Both must remain speculative at best, they say, because the latent variables are precisely latent, or unavailable for observation or confirmation.

Furthermore, the normal distribution assumption becomes a problem, as there is also no way to derive theoretical values for the tetrad distributions, which could be compared with the observed values and examined for probable error discrepancies. Finally, Pearson and Moul showed that although Spearman and Holzinger had taken the mean product values across all correlations for the scores to use in their tetrad equations, the tetrad method should only make use of a subset of the variables. What this meant was that the scores included in the tetrad equation correlations should only be for the specific variables matched in the tetrad matrix. Pearson and Moul demonstrated substantial divergences between the variable values that should have been included, and the subsequent analyses in the matrix.

With those problems, Pearson and Moul (1925) stated that it was impossible to determine whether there were problems: i) associated with choosing the wrong test or variables; ii) problems with the theory of two factors itself, or iii) some combination of both. Pearson later particularly takes issue with Spearman's mathematical claims for a "perfectly defined quantitative value '*g*'" (Pearson, 1927, p. 181). He points out that the linearity assumed by Spearman (1927) in his factor theory, that explicitly quotes Taylor's theorem, is well suited to physics experiments where the linearity can be tested, but completely unsuited to psychology where no such experimental confirmation is available.

Spearman's continued search for independent evidence for cognitive ability or *g* beyond the tetrad equation outcomes suggests that he took the critique from Pearson and Moul seriously. He goes on to make use of the neurological theory of the time to propose a theory of noegenetic laws, which described how he conceived cognition functioned as a series of relations between perceptions, and cognitions, via correlation and association. By 1930, Spearman is describing this theory as much more important than the factor theory (see his autobiography, 1930/1961).

Spearman's noegenetic laws however disappeared from psychology theory with an alacrity that speaks nothing of the longevity of factor theory (Norton, 1979). Following Pearson's analysis, however, it can be noted that the tetrad equations proof for factor theory exhibits *inferential* gaps - the steps of reasoning offered by Spearman are not those needed to say proof has been effected. In this way, the proof is also untraversed. The gaps in proof for factor theory however seem not to have effected its uptake as the basis of the GLVM, or the prevalence of the GLVM in psychology research today. Subsequent developments in the GLVM evidence different concerns with respect to mathematical proof. Solutions for the latent variable parameters in the GLVM may be found through maximum likelihood solutions. These have foundation in Fisher's maximum likelihood estimation techniques, and these are presented, in the next section.

5. Fisher and maximum likelihood statistics

The method of maximum likelihood is typically cited in the literature as addressing the problem of estimating parameters from some set of data (see Hinkley, 1980; Thompson, 2004, p. 127), such as the sort gathered for manifest variables in the GLVM. It is the default for parameter estimation across many GLVM software packages, today (Thompson, 2004). Maxwell (1959) cites Lawley (1940) as the first publication that sought to bring the statistical analysis of maximum likelihood methods to the mathematical structure of Spearman's factor theory. Young (1941) summarises the notions that were present in the early work of Lawley (1940) in looking to techniques that would provide solutions for latent variables. Historically, latent variables or factors now were considered to form "true scores" for a population who had the same latent variable outcome (see Lawley, 1940; Thomson, 1934). With this assumption in place, where there were data for the manifest variables, the next step was to estimate values for the latent variable, based on the assumption of normal distribution for the errors in the population.

What this effected, in the beginning, was an integration of Spearman's two earliest papers. Factor theory as discussed above (Spearman, 1904a) broke down a test score into a common factor, and a specific factor connected to the specific test administrated. The second paper made use of a similar idea about bifurcation, but here splitting the test score into two components, a true value, and a value representing random error. The error value here was intended to account for the difference between the observed outcome and some outcome in an infinite population (see Chapter 4 for further discussion). This second paper introduced the idea that some true score value could be calculated, for the latent variable, given a set of data for the manifest variables. One statistical technique for estimating these parameter values is the maximum likelihood method.

Maximum likelihood was developed first by Ronald Fisher (1912, 1922, 1934a) who through his body of work provides two distinct methods (Edwards, 1974), and three justifications for maximum likelihood procedures (Aldrich, 1997). Fisher sought not only to provide techniques for estimation of parameters making use of statistical assumptions, but also a way to think about criteria for testing hypotheses about what the parameter solutions were. The maximum likelihood method in its original form was computationally prohibitive, but uptake was facilitated by the introduction of software and computing technology to handle the onerous calculations (Thompson, 2004). We will first review Fisher's background, and then have a look at the development of the maximum likelihood method for estimates, focusing on the question of mathematical proof, and then look to some of the specific problems that emerged in the application of maximum likelihood to factor analysis or latent variable modelling.

5.1 Fisher's background

Fisher studied mathematics at Cambridge, abandoning a keen interest in biology at a historical time where genetics had just become a subject credited with academic recognition (Box, 1978, p. 22). Fisher's interest in biology remains with him throughout his career, however, publishing in 1930 *The Genetical Theory of Natural Selection,* which even today remains as a major influence on the way that biology is conducted (Edwards, 2000). Fisher was a strong advocate for Mendelian and Darwinian approaches to genetics. He united them in a vision both biometric and eugenic, creating the Cambridge Eugenics Society in 1911. He advocates directly for the use of biometrics to improve the genetic lot, of the human race in a way at once both nationalistic and grand. Charles Darwin's son, Major Leonard Darwin, sponsored Fisher's eugenic interests.

Fisher's fascination for the process of evolution and implications for mutation in otherwise 'as is' situations provides the foundation for the statistical innovations that he develops including analysis of variance methods, and techniques for discerning sufficient, efficient, and consistent statistics (Box, 1978). Fisher (1912, 1922, 1934a) also develops the methods of maximum likelihood estimation which make use of the idea that data, as gathered, represent a certain distribution amongst variables. With assumptions such as normal distribution and constancy of errors in infinite populations, Fisher proposed that something could be stated in shorthand form, which summarises all the information contained, in the data. His thinking about maximum likelihood progresses in stages, and to these, we now turn in detail, including some review of his ideas about the infinite population, from which we could draw, statistics.

5.2 History of maximum likelihood

From Hinkley (1980, p. 2), we can understand that up to the year 1912 several statistical innovations had occurred which become relevant to psychology research. These included least squares estimation technique and the theory of errors (Gauss, 1809; see Stigler, 1981); chi-square goodness of fit technique (Pearson, 1900); breaking down variance into components (Edgeworth, 1908); asymptotic normal likelihood technique (Edgeworth, 1908); and probable error handling (Edgeworth, 1908; see Pratt, 1976). Gauss's method of least squares fits a regression line through a set of data points in order to best account for the variance in the data, and is the precursor to the maximum likelihood method, as Fisher was dissatisfied with the technique. Fisher was critical of the method of least squares because “an arbitrariness arises in the scaling of the abscissa line” (Fisher, 1912, p. 156). What this meant was that he noticed the least squares method failed to consider any variations in the distributions for data-points, assuming a uniform prior or unchanging distribution for the data.

Fisher’s (1912) paper is influenced by his study under astronomer F.J.M. Stratton at Cambridge University, who lectured on the theory of errors (Aldrich, 1997; Stigler, 2007). In this paper, Fisher is working with the notion that variables need to be rescaled for every movement around a distribution curve. His argument applies both for approximating one curve by another, and for fitting a curve to data, noting that modelling based on real-world data introduced a further source of variance beyond just that in the data, because the researcher could arbitrarily choose where they located their starting point. He uses the idea of testing whether an observation occurs within a probability range as the way to maximise the likelihood, of a curve of best fit, for the data. What he had done, however, within his mathematical modelling, was assume a uniform prior for the variables – that the errors stayed constant, even as the data-points for a variable may change (Aldrich, 1997).

Eventually Pearson challenges Fisher in such a way that he moves beyond data-points to consideration of probability density or dispersion of data across the data-points and the associated variances. In 1918, Fisher worked on establishing sufficiency in estimation of normal standard deviation (Fisher, 1920). A sufficient statistic, as defined by Fisher (1922), is one calculated such that no other calculation could provide more information about the parameter of interest. This is work drawn into a mathematical argument for maximum likelihood, using two estimates for a given parameter, each assumed approximately normal in large samples.

The mathematical proof for maximum likelihood follows in the 1922 paper. This paper takes two potential solutions as estimates for any parameter θ for analysis. It also relies on an assumption that these potential solutions are sufficient, themselves, having bivariate normal distributions and an expected solution of θ. Fisher then uses the assumed sufficient estimates and combines them into an overall estimate, that he also labels as sufficient, using expected mean squared errors as a criterion. These would be sufficient, however, only because they are assumed as sufficient, in the model. What this means is that sufficiency is assumed for the parameters, in order to find, sufficient parameters. Fisher's reasoning here is circular.

5.3 Sufficiency in statistics

Fisher did not see the problem with assumed sufficiency initially, claiming sufficiency implied optimality, when combined with normal distribution and consistent estimates. Fisher's claim, following his analysis, was that "maximising the likelihood always led to an estimate that was a function of a sufficient statistic" (Stigler, 2007, p. 601). Stigler (2007) notes that as the paper went to publication, Fisher expressed doubts about whether he had offered a proof, but also, he gives reasoning for still releasing the paper, in an expectation that applications would lead to improvements in the mathematics of his technique.

In this first articulation of maximum likelihood, situations where a sufficient statistic was not available were not addressed. Fisher historically from this point steps away from grand claims about the sufficiency of maximum likelihood estimates - even while pointing out some valid concerns about Pearson's production of point estimates (see Stigler, 2007, p. 602). What follows is a subtle shift in Fisher's work, where he de-emphasises sufficiency, and begins to focus on efficiency, in statistics. An efficient statistic is one that "requires that the fixed value to which the variance of a statistic (of the class of which we are speaking) multiplied by *n*, tends, shall be as small as possible" (Fisher, 1925, p. 703). For this new work, Fisher was interested in minimising potential loss of information in the estimation process given a set of data for the hypothesis. For Fisher (1922), the task of data reduction given a data set was best achieved by:

> constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample. The law of distribution of this hypothetical population is specified by relatively few parameters, which are sufficient to describe it exhaustively in respect of all qualities under discussion. Any information given by the sample, which is of use in estimating the values of these parameters, is relevant information. Since the number of independent facts supplied in the data is usually far greater than the number of facts sought, much of the information supplied by any actual sample is irrelevant. (p. 312)

Here we see the introduction of ideas about random sampling from infinite populations, and a close alignment of ideas discussed above that are relevant to the work of Bartholomew et al. (2011), with the notion of summarising data by representing it with few variables. Note this is distinct from what is described above as Spearman's purpose for common factor analysis, which was to provide a precise quantitative expression, for cognitive ability or intelligence. Returning to Fisher, he then developed a measure for efficient estimates for maximum likelihood, calculated for situations where estimates moved in any direction away from the original data. Later researchers labelled this as second-order efficiency (see Rao, 1961). This meant checking calculations as a test of how far the model was from the data; using ideas of "approximate" normality (see Stigler, 2007). Maximum likelihood now, instead of producing sufficient statistics, in Fisher's account, was a technique constructed to produce efficient ones, given large samples, normal distributions, and small standard deviations, in approximate normality, for all variables.

Fisher goes on to develop a working relationship with Harold Hotelling, who in

1930 published on the *consistency* of maximum likelihood estimates. Consistency describes the idea that a parameter can take values as a random variable not for a sample but for a population, and that estimates approach a true value for the parameter, in the limit of an infinite population (Rao, 1992). Hotelling (1930) had demonstrated the impossibility of minimising variance across a set of data that maximises likelihood for all or general cases using geometrical methods. In response to this Fisher produces a third proof in 1930. This proof did not demand the old assumptions of normal distributions and large samples, but instead demanded homogenous functions of degree zero as functions of the relative frequencies for the variables – the advantage being here that the estimates did not rely on sample size, but rather, presumed a limiting, but infinite, population. What this allowed was for a smooth differentiation over the point estimates given that they would converge on true values in the limit, and for reliance on known covariances, for multinomial expressions.

In summary, Fisher's first proof assumed sufficient statistics were always available for a set of data and the initial parameters, from which he claimed that maximum likelihood solutions were themselves always sufficient. Arriving at an understanding that this was not always a good initial assumption, Fisher produced a second proof, which relied on assuming approximately normal distributions for parameter solutions, only. It made use of the assumption that the variances for the estimates were constant, and that likelihood itself was regular enough for calculation and evaluation of integral functions. This proof still failed to be relevant to all situations considered by Fisher. The final proof assumed multinomial normal distribution and smoothly differentiable functions for relative frequencies over an infinite population, which were even more restrictive assumptions than were used for the first two proofs. Cramer (1946) noted the lack of generality for this final proof, however, and went on to demonstrate cases where the maximum likelihood estimates were inconsistent. This was followed by work from others demonstrating inconsistent estimates from maximum likelihood methods (see Hodges, 1951; LeCam, 1953).

Stigler (2007) poses a relevant question - given the evidence for inconsistency of maximum likelihood estimates, and the lack of finality regarding the proofs for the method, how is it the case that it is so widely used, as a solution, today? He further notes that the tautology in the proofs. Assuming all the conditions that are necessary for each proof, each proof holds, but, “any statement is true if all the conditions required for its truth are assumed” (Stigler, 2007, p. 613). Fisher had several implicit assumptions not clearly stated in the first instances of the proof, but which become apparent in his correspondence with sceptics. Problems subsequently pointed out by his detractors included increasing parameter numbers, local improvements in superefficient estimates over efficient estimates, and unbounded likelihood functions - problems that are relevant with the size of data pools that we have at our disposal, today. Stigler (2007, p. 614) nevertheless notes that usefulness of the method outstrips any reasonable efforts aimed at proof. What happens then is there is risk of applying the method where it cannot reasonably be said that it is relevant. Stigler (2007) rounds off with a final warning in his last line:

> Maximum likelihood remains a truly beautiful theory, even though tragedy may lurk around a corner. (Stigler, 2007, p. 614)

In terms of our tests for mathematical proof, we can note the outcomes are like those we saw with Spearman’s production of proof. Early inferential gaps of fallibility result in aspects of the proofs remaining untraversed. Rigorisation, subsequently pursued through history provides us with evidence of the earlier existence of the other two sorts of gaps, without the need for us to assume, necessarily, any enthymematic gaps in the proofs so offered. No lasting proof exists for an accurate solution for parameters, using maximum likelihood solutions. The community of practitioners around both Spearman and Fisher played vital roles in checking claims and furthering the methods, but unlike was the case for Gauss with Ostrow (1920), no final proof has ever been found, for either factor theory or maximum likelihood methods, following.

5.4 Maximum likelihood in latent variable modelling

The demonstrations of inconsistent estimates were published around the time that maximum likelihood techniques began to be employed in latent variable modelling methods (see Stigler, 2007; Hodges, 1951; Lawley & Maxwell, 1962). Concerns expressed in the literature about the process of estimation focused on the limits of calculation methods given inclusion of maximum likelihood algorithms in the software, and typically did not address situations where the onerous a priori assumptions did not hold (see for example Lawley and Maxwell, 1963). Not all psychometricians were as enthused about the possibilities for maximum likelihood estimation combined with factor analysis as the early advocates such as Lawley, Maxwell, and later, Jöreskog. Kaiser (1976) points out that while Lawley and Maxwell lauded Jöreskog (1967) as a seminal piece of work, difficulties remained for psychometricians in practice, with a series of numerical problems such as Heywood cases, where the maximum likelihood estimate found comes out below the acceptable lower bound, for the parameter in question. Prior to this, Savage (1954) had commented on the idea that a maximum likelihood estimate can have some role in a hypothesis test, noting that the parameter so estimated relied primarily on sample size, some arbitrary level of significance, and the power of the test. These concerns orient our attention to questions regarding application of the maximum likelihood method and arise in part because inferential gaps persisted in the mathematical proofs offered by Fisher through the three successions of attempts at the same. Kaiser (1976) suggests that what is truly at stake in latent variable modelling that makes use of maximum likelihood solutions is an infinite universe of possible variates, distinct from what may or may not hold true for an infinite population of individuals. Kaiser (1976) here is speaking to the problem of factor indeterminacy, one connected to Spearman's failure to produce a mathematical proof for his tetrad equations method. Factor indeterminacy is a topic of core focus in our next chapter that looks to the problems associated with the use of the latent variable model today, distinct from these problems to do with the development of the approach in the past.

6. Summary

In the first half of the chapter, two distinct definitions and models of the GLVM were presented, and it was noted that researchers would be required to produce distinct information in support of their research claims, at the conclusion of a GLVM analysis for a research project. From the account of Spearman's attempt at proof for factor theory, it has been stated that because Spearman's analysis cannot be counted as proof of the split between the latent and manifest variables in the model, independent evidence is needed for the construct or phenomenon behind the latent variable. This is because Spearman's aim of providing a precise quantitative definition for the common factor was not met. For the use of maximum likelihood methods used in resolving parameters for GLVM, presented in the final section of the second half of the chapter, it was noted that it is possible to find superefficient estimates (Hodges, 1951), which are estimates that have smaller asymptotic variances than the maximum likelihood estimate. These would notionally therefore be a better estimate, but these estimates are always inconsistent. All maximum likelihood solutions themselves are only estimates, and therefore their generalisability to broader contexts where other variables may play distinct roles must remain at question. Development of maximum likelihood techniques were rationalised in terms of data reduction and remain constrained by the condition of usefulness, rather than truth.

Where does this leave us? There several gaps for the proofs offered for methods associated with the GLVM. While these are not fatal to the practice, they make it impossible to completely extricate and attribute distinct sources of error associated with the modelling practices. What they indicate is a need for careful delineation of assumptions associated with GLVM techniques. The need for a robust standardised framework where these gaps can be localised and addressed by theory/empirical finding/independent evidence is indicated.

Chapter 4: The GLVM, Assumptions, Constraints, Problems

1. Objectives

The prime objective of this chapter is to examine developments for the GLVM modelling through the twentieth century, with attention to uptake of techniques in psychometric practices and research projects. This is done with a view to logical questions about the problems associated with the use of the GLVM, the answers to which inform the structure of the conceptual framework proposed in this thesis. The way that the logical questions inform the structure of the conceptual framework is via looking to the impact on quality uncertainty, with respect to research outputs and claims (see Vazire, 2017, 2018). Quality uncertainty, as it is relevant to the GLVM, is described in more detail in the section immediately below.

In the last chapter, the early history of the development of the mathematical model for the GLVM was explored. Limits were acknowledged regarding the application of mathematical proof both for the model itself with its origins in factor theory, and for one of the primary methods for finding solutions for the parameters in the model, the maximum likelihood method, of Fisher. In this chapter, subsequent developments for the GLVM are addressed, with respect particularly to statistical inferences. Even without conclusive proof for the mathematical model, which would locate some determinate fault line between the manifest and latent variables, researchers still make use of the GLVM today, in for example latent profile analysis of the big five personality factors (Fisher & Robie, 2019) or Rasch analysis of a depression inventory (Christensen, Oernboel, Nielsen, & Bech, 2019).

In making inferences about latent variables, no matter the context, researchers rely on correspondence between model structures and real-world phenomena, as well as practices that are accepted within the psychological community as those that facilitate robust scientific inquiry. In what follows, a series of concerns for the present-day use of the GLVM are presented, and common strategies that are employed with the GLVM are discussed. Particular attention throughout the presentation is paid to logical implications of the practices, in order to best inform the structure of the conceptual framework that follows.

2. Quality uncertainty and the GLVM

Vazire (2017) recently has introduced the term quality uncertainty into psychology literature. In "*Quality Uncertainty Erodes Trust in Science"*, Vazire notes the term quality uncertainty itself originates in the work of Nobel-prize winning economist, George Akerlof (1970), in an original analysis of how trust influences market activity. Akerlof describes quality uncertainty with specific reference to "lemons" and the effects of information asymmetry in the used car market. Information asymmetry applies because the seller has more information than the buyer does. Aklerlof proposes three key claims regarding the impact of quality uncertainty in situations of information asymmetry (Cooper, 2007). Firstly, where quality uncertainty may exist among items in a market, overall consumer confidence falls in that market. Secondly, the quality uncertainty has a direct impact on recipients of goods in the market but it does not really influence the individuals making market goods available – the vendors – in the market. Thirdly, where quality uncertainty exists, secondary markets grow, that aim to provide assurances about quality. These include, for example, guarantees, warranties, and development of certification standards and bodies.

The effects of market uncertainty are recognised in other fields beyond psychology, which are not traditionally understood as constituting markets. These include, as examples, higher education (Cooper, 2007), medication effectiveness (Ching & Ishihara, 2010), and environmental planning (Jansson & Waxell, 2011). Vazire (2017, 2018) conceptualizes the communication of scientific knowledge by psychology researchers as a market, and proposes that quality uncertainty, as it operates in psychological science, impacts on the ability of the research community and the public to have confidence in the research outputs that follow research projects. Vazire (2017) specifically targets failures in transparency as a primary cause of quality uncertainty and makes some specific suggestions regarding transparency that are taken up as a theme in Chapter 6 of the present thesis.

One other aspect of quality uncertainty of relevance when considering the use of the GLVM in psychology research is quality management. Quality management in economic literature is usually described as total quality management (TQM: Wankhade & Dabade, 2006). There is no universally accepted TQM model or definition (Sila & Ebrahimpour, 2002), but early influences in the TQM movement focused on maximising the economic value of production outputs while maintaining economic efficiency and effectiveness in processes in organisational contexts (Black & Porter, 1996). Recent focus in TQM has shifted away from quantitative to qualitative inputs in processes, such as assessment of stakeholder involvement in quality practices and focus on leadership and communication in organisations (Motwani, 2001). Following the recent inclusions of qualitative and quantitative elements in TQM processes, Wankhade and Dabade (2006) propose that quality uncertainty itself is a function of a lack of TQM and information asymmetry.

For the purposes of psychology research, improving scientific transparency directly addresses quality uncertainty caused by information asymmetry, but at best, transparency indirectly addresses the lack of quality management regarding research outputs. Indirect address would take place where researchers perform activities that provide quality assurances about their research outputs, because of the threat presented to the credibility of their findings without such disclosures[9]. As stated above, the ways that the conceptual framework can facilitate transparency will be described in Chapters 6 & 7. Regarding

[9] Vazire (2018) notes as transparency increases, it is likely that increased uniformity of practices given research community acceptance will also follow.

quality management for psychological research, over time, several practices have been developed in research communities that address the quality of outputs of GLVM practices. These, ultimately, can be included as elements reported in a framework which takes as a focus the maximisation of transparency. The practices are addressed in the section on applications, below. They include reliability and validity checks as relevant to the GLVM variant that is employed in a given research project (see Borsboom, Mellenbergh, & van Heerden, 2004; Slaney & Maraun, 2008; Slaney, 2017; Maul, 2017).

It is a proposal of this thesis that certain problems that are inherent for the GLVM, no matter the variant used in practice, which render use of the model directly susceptible to quality uncertainty. To address quality uncertainty for implementations involving the GLVM in research projects, what is required is not only transparency, but also conduct of quality assurance processes relevant to the application of the model. It further demands accountability for researcher decisions and explicit declaration of the research domains in a standardised conceptual framework. This involves both quality assurance practices relevant to GLVM analysis and Bayesian-style assessment of entries into the conceptual framework, in a self-audit process by the researcher. This is described in detail in Chapter 6. In the meantime, to clarify the nature of this foundational threat to quality uncertainty for research claims following from the use of the GLVM, the next section addresses three problems that are present no matter the specific form that is adopted regarding the model. These are the conceptual, mathematical, and logical problems of conditional independence, factor indeterminacy, and circularity, respectively.

3. Problems in present day use

In Chapter 2, two different versions of the GLVM were presented, from McDonald (1999) and Bartholomew et al. (2011). Both versions of the GLVM are general enough to express an array of latent variable model types, and it was shown that each account asks researchers to perform distinct practices on their way to making a claim from GLVM analysis. It is not the case, however, that each model that falls under the rubric of any specific account of the GLVM operates in the same way or is utilised in the same set of circumstances. A research project may employ, for example, EFA to create a questionnaire with a number of items purporting to assess one or more latent variables. CFA may be employed at a second step, in examination of, say, the degree to which the latent variables and their relations hold as a model for a second sample of individuals (see Thompson, 2004; Finch & French, 2015). SEM allows for testing a directionality of relations between variables, as well as the relations themselves, in processes described as causal modelling (Jöreskog, Sorbom, & Magidson, 1979). These distinct processes bring with them distinct decisions related to the application of the model in a research situation that the researcher must address in utilising some variant under the GLVM.

In what follows, the consideration of problems and concerns regarding decisions for the GLVM fall into two categories: i) theoretical or conceptual problems, which have relevance no matter the type of model employed; and ii) application problems and decisions, which are associated with the use of particular forms of the GLVM. The problems and decisions covered here are not exhaustive for the GLVM precisely because of the broad array of forms the latent variable model may take, under the generalised approach. They are set out here, however, since they are more common in the practices that are included in the GLVM rubric and show what sorts of considerations we can expect the conceptual framework to at a minimum cover. We begin with the theoretical or conceptual concerns.

3.1 Problems - theoretical

The GLVM, as stated in previous chapters, has been defined by its reliance on the assumption of conditional independence, with both McDonald (1999) and Bartholomew et al. (2011) agreeing at least on the degree to which conditional independence is not an assumption, but a constraint which defines the model.[10] This constraint distinguishes the GLVM from other types of statistical models. As seen in what follows, to apply the theoretical model for the GLVM in practice, two other key features accompany any application of the GLVM. These include the mathematical problem of factor indeterminacy, and the logical problem of circularity.

3.1.1 Conditional independence

Conditional independence is the mathematical foundation of the latent variable model (Lord, Novick, & Birnbaum, 1968, p. 538; Mulaik, 2010)[11]. As noted in McDonald and Bartholomew's accounts of the GLVM in Chapter 2, it is a condition for the latent variable, rather than an assumption, because it cannot be tested as a hypothesis. As a condition, conditional independence works as a constraint on what the possible outcomes from the latent variable analysis may be. Lord, Novick, and Birnbaum (1968) provide some breakdown of characteristics relevant to the conditional independence constraint. Psychological phenomena for these authors can be accounted for as a function of "certain consistent and stable human characteristics, or *traits*" (p. 537, italics in original). The idea, according to these authors, is that it is possible to collect "values" (p. 537) for a person on these traits, and thus predict or describe aspects of the person's behaviour in certain situations considering these scores.

Lord et al. (1968) note typically there is a theoretical formulation of the kinds of traits that belong together in any kind of an assessment. Spearman (1904), for example, in his account of general intelligence drew on conceptual analyses of earlier theory and some empirical studies to formulate his account of factor theory. Lord et al. (1968) go on to state, however, that the kind of theoretical knowledge that is advanced does not necessarily

[10] Note this is often labeled as local independence in IRT (see Rosenbaum, 1984).
11 This feature distinguishes a latent variable model from principle components analysis or other forms of what have been called formative models (Edwards & Bagozzi, 2000).

have to be about a trait that exists:

> … in any physical or physiological sense. It is sufficient for most purposes in psychology that a person behave as if he were in possession of a certain amount of each of a number of relevant traits and that he behave as if his value on these traits substantially determined his behavior. (p. 537)

The other key aspect of the adopted definition of a latent trait for these authors is that it is:

> the only important factor and, once a person's value on the trait is determined, the behavior is random, in the sense of statistical independence. (p. 538)

This emphasis on statistical independence and randomness, over and above some physical non-random basis for a trait allows psychometricians to assume that one event is completely independent of any other that may occur within the same domain for that trait. It also allows for inferences or estimates over an infinite population for that same domain for the trait (see Fisher, 1956, p. 139ff). Question is raised by Lord et al. (1968) regarding the ways in which the concepts of randomness can be said to apply to human participants, who typically are not amenable to the brainwashing that is demanded for true randomness to apply (see also Borsboom, 2005). These issues will be further discussed in the distinctions made between concepts, constructs, and variables following Markus (2008), in Chapter 5.

The assumption of conditional independence goes a step beyond statistical independence, regarding constraints however, as described by Lord et al. (1968). Conditional independence is defined where, for any population or subpopulation *P* (be it finite or infinite), for a fixed θ, the joint distribution of variables $X_1, X_2, \ldots, X_n$ factors into a product of the marginal distribution functions for *F*:

$$F(x_1, x_2, x_3, \ldots, x_n \mid \theta) = F(x_1 \mid \theta)\, F(x_2 \mid \theta)\, F(x_3 \mid \theta), \ldots, F(x_n \mid \theta)$$

Lord et al. (1968) quote Anderson, regarding this assumption (1959, p. 11):

> Apart from any mathematical reason for such an assumption there are psychological or substantive reasons. The proposition is that the latent quantities are the only important factors and that once these are determined behavior is random (in the sense of statistical independence). In another terminology, the set of individuals with specified latent characteristics are "homogenous".

Homogeneity here means all conditional error distributions are assumed equal for that set of individuals, for that specific latent variable. Lord et al. (1968) note this idea of behaviour being random given a value for individuals on the latent variable is a strong assumption. They also connect to the discussion in Lazarsfeld & Henry (1968) on experimental independence. Experimental independence includes these assumptions: i) each measurement or data point is distinct from any other data point; ii) the actual score or true score has a constant value (or logical identity); as well as, iii) error terms remain statistically independent (See Lord et al., (pp. 44-45). Here, in connection to latent traits, conditional independence is defined as assuming both experimental independence, and statistical independence for the model error terms (beyond just simply the variable error terms, which formulate a part of experimental independence). Lord et al. (1968) make clear their position that these assumptions remain theoretical, in their formulation.

Conditional independence not only involves further assumptions about experimental and statistical independence with homogeneity of conditional error distributions for the manifest variables, it has also been demonstrated for multivariate GLVMs to rely on assumptions of monotonicity. Monotonicity refers to the assumption that the function between the variables is entirely singular in its direction – for the GLVM, this is an assumed positive function. Van Rijn and Rijmen (2015) describe the problem of 'explaining away' in multivariate modelling, where random latent variables are used. While the original assumption may be that the latent variables are statistically independent, the problem is that it may be observed that increase in the correlation value for one latent variable may reduce the correlation for another latent variable in the model. Using the framework of graphical modelling, van Rijn and Rijmen (2015), following Pearl (1988), explore to what extent the assumption of conditional independence holds in real situations. Following problems examined earlier by Suppes and Zanotti (1981), where it had been noted that conditional independence could always be obtained in a joint distribution, van Rijn and Rijmen suggest that what is needed is an additional assumption of monotonicity for the variables. Essentially, this amounts to a requirement that there is a reliable second-order positive relation between the manifest and latent variables (see Karlin & Rinnott, 1980 for a full discussion).

Summing up, use of the GLVM inevitably involves reliance on the assumption of conditional independence. Conditional independence also involves further assumptions about a positive relationship between manifest and latent variables, which may or may not be relevant in certain contexts. Statistical independence may be involved in the use of the GLVM, particularly where modern statistical software is utilised in finding solutions for parameters, but this is not the same thing as conditional independence. In this case, homogeneity of conditional error distributions is also assumed. Moreover, if there are multiple latent variables, the assumption of monotonicity applies whenever conditional independence is employed, to help resolve problems of dependence relations between random latent variables. A psychological researcher then may make use of the conditional independence structure without any understanding of the vital nature of the assumptions that underlie its effectiveness in a mathematical sense. A conceptual framework must be malleable enough that a researcher is drawn to consider the kinds of assumptions like these that are critical to understanding what it is that can be said about the analysis, following the completion of the research project. There is another concern however for the mathematical latent variable model when reflecting on the implications of the GLVM for the conceptual framework: this is the problem of factor indeterminacy, and it is to do with the structure of the mathematical model, itself.

3.1.2 Factor indeterminacy

Factor indeterminacy is a mathematico-grammatical concern for the GLVM because any act of obtaining a solution for a latent variable mathematically logically applies for an infinite number of any other possible latent variables (Guttman, 1955; Rozeboom, 1988; Maraun, 1996). It is only because factor indeterminacy is present within the mathematical model that more can be said about the latent variable than could otherwise be said from either the evidence garnered from the manifest variables, or the fit of the hypothesised model to the data (Mulaik, 2010). Factor indeterminacy also has a long history. When reviewing Spearman's (1927) *The Abilities of Man*, Wilson (1928, 1929) noted a problem for Spearman's concept of *g*, or general intelligence, which involved more than the problem of error. It involved the impossibility of deriving a unique solution for *g*, a problem he had indicated earlier when analysing the independence of tetrad difference equations (Lovie & Lovie, 1995, p. 241; Mulaik, 2005, 2010). Using vector space analysis, Wilson (1928) later demonstrated that latent variables lie partly outside the space described by the linear combinations of the manifest variables, and thus, cannot be *uniquely* determined. Part of the latent variable is estimable from the manifest variables, but part is also not estimable (Wilson, 1928; Maraun, 1996; Mulaik, 2010).

Spearman (1929, 1933) responded to Wilson (1928) suggesting that indeterminacy could be solved with the addition of a variable that was exactly correlated with *g* to the already included manifest variables (Mulaik, 2010, p. 380). This attempt at a solution addressed *exactness*, and could lead to a suggestion that perhaps the new variable could be used as a direct measure, rather than relying on all the analysis relevant to a latent variable. It still did not address *uniqueness*, however. The addition of a variable does not overcome the innumerability of solutions for *g*, and Mulaik and McDonald (1978) set out a proof demonstrating that increasing the number of variables included *to infinity* does not eliminate the problem of factor indeterminacy.

Looking at this problem considering questions of logic, we can note that indeterminacy leaves us with no implicit logical structure for the variable against which the indeterminacy problem can be resolved – it is a problem to do with the *uniqueness* of the solution for a latent variable. A *unique* solution for the latent variable would give us a means by which a logical relation could be established between the construct, and the latent variable in the mathematical model. This would support inductive inference, with a greater degree of certainty then available about the inferences than is the case, otherwise (Rozeboom, 1988; Maraun, 1996; Mulaik, 2010). Indeterminacy can be understood as impeding the possibility of securing identification for the latent variable – where identification would indicate the relation of the latent variable in the model to the psychological construct we have formulated in our research project, which ultimately connects back to the psychological phenomena in question.

Various solutions have been proposed for the problem of factor indeterminacy since Spearman's fated proposal of adding another variable that perfectly correlated with the latent variable. A leading article by Maraun (1996) in the journal *Multivariate Behavioral Research* characterised McDonald and Bartholomew as offering contrasting perspectives on potential solutions to the factor indeterminacy problem. The alternative solutions position eventually adopted by McDonald[12] acknowledges that there remains an infinite array of other possible solutions for the latent variable. What this entails is a reliance on extra-statistical assumptions regarding the meaning of the analysis and its connection to properties of psychological phenomena (Maraun, 1996, p. 520). These include specification of the infinite behaviour domain as was described in Chapter 3, as one response. The posterior moments position of Bartholomew, on the other hand, interprets the latent variable as a random variable with non-point posterior distribution given the parameter values in the data for the manifest variables. With this position, Bartholomew (1996) proposes that it is possible to quantify the degree of knowledge or determination one has of the latent variable (Maraun, 1996, p. 523). Maraun (1996) notes these are opposite positions, but for him the logic of latent variable modelling still suffers indeterminacy insofar as there is no external identity criterion available for the latent variable, just given the mathematical model.

In terms of what will be needed in the conceptual framework to address the problem factor indeterminacy, it can be seen that the philosophical stance of the researcher towards the GLVM and what it is and does has an influence on the constraints that apply to the claims that follow from analyses using the GLVM. The philosophical commitments as illuminated in the positions of Bartholomew versus McDonald make a difference to what the researcher is asked to report in their final analysis. A posterior moments interpretation

[12] Note Maraun (1996) characterises McDonald as taking a perspective aligned with posterior moments. McDonald (1996) clarifies a number of turns taken in his own position on the meaning of the latent variable, and by the time he publishes McDonald (2003) on the infinite behaviour domain, it is clear he falls in favour of the alternative solutions position.

of the GLVM following Bartholomew means that a coefficient of determination should be reported that indicates something of the confidence or amount of knowledge that is assumed to be addressed via the latent variable analysis (Bartholomew, 1981). It also implies an assumption that there is only one latent variable (Maraun, 1996, p. 525), thus constraining what we would expect to see in terms of the latent variable analysis to a single rather than multiple latent variable model. An alternative solutions position following McDonald on the other hand demands that the researcher account clearly for the expected relationship between the phenomena of interest, the concept that defines it, the psychological construct used to assess it, and the variables that represent it in the model. More is specified regarding clarification of these relationships in set theoretical structures in Chapter 5 drawing on techniques first set out in Markus (2008).

For Maraun (1996), the factor indeterminacy problem for latent variable modelling under a posterior moments position presents a paradox where a good fitting model is found for the data. This is insofar as we had to assume that a single latent variable was present to use the model, yet the solution itself implies an infinity of latent variables. Where an alternative solutions position is adopted, what is needed is clear specification of the conceptual terrain for the phenomenon thought to underlie the latent variable. The tools incorporated into the conceptual framework for this will be further discussed in Chapter 6 and 7, looking to set-theoretical resources endorsed in Markus (2008) discussed in Chapter 5, and formalised in Suppes (2002). There is one final logical concern to address for any use of the GLVM in psychological research; this is the problem of circularity.

3.1.3 Circularity

Circularity was discussed briefly in Chapter 2. It is present in the process of latent variable modelling where the existence of intelligence, for example, is asserted via evidence garnered in inductive statistical analyses of test-score outcomes where intelligence was already assumed as the feature of interest, in the test items (Boring, 1923).

When test items are prepared, they typically express some abductive guess about the kinds of items that would best suit the tracking of the psychological phenomena in question (see also Chapter 5 section on Haig for advocacy for a role for latent variable modelling in abduction in psychology research). Circularity occurs because the conditional independence assumption in the model entails a further consequence – all that can be said about the latent variable, comes from information garnered via analyses of the manifest variables (Howe, 1990; Boag, 2011). The problem of circularity, then, is that researcher want to use the latent variable to explain the data, but the evidence for the latent variable is the data. In terms of reification, a pattern of data (a set of relations) is used to 'concretise' an underlying variable (also explaining the data). The problem of circularity has also been labelled as a fallacy or false inference in logic, which may also be called the fallacy of begging the question, or *petitio principii* (Walton, 1994; Wood & Walton, 1978, 1982). "Verbal magic" is one term employed in the psychology literature, to describe what may happen in the scenario of psychological research (Maze, 1954; Boag, 2011). For example, because psychological phenomena remain non-observable, and because psychological phenomena always occur within a network of relations as an event, it may be the case that relations are mistakenly treated as properties of individuals or things with their own independent characteristics (Boag, 2011). This entails the further problem of reification, which is distinct from the problem of circularity, but has the same result. Something "new" is said to come into being by virtue of analysis of the relation, yet relations can only occur between two independently existing entities or properties (Boag, 2007; Maze, 1983; Passmore, 1935). Ultimately, what would best serve GLVM analysis of a latent variable is independent evidence of the existence of the phenomenon behind the latent variable, distinct from merely relying on the outcomes of GLVM analysis, to make claims about the existence of some entity.

The suggestion of domain setting by McDonald (1999, 2003) following Guttman

(1954, 1955) gives some address to the concern of circularity. This works by setting a conceptual domain for the test items, providing some confirmation of analyses considering some background theory or research work (see also Slaney & Maraun, 2008). In mathematics, McDonald (2003, p. 213) notes that domains express the set of elements on which a variable is defined. In terms of a domain for the latent variable or for manifest variables, formalisation of the domain is important in order to be able to coherently add or remove test items, given latent variable analysis. Circularity in these terms is addressed by clarifying or grounding a conceptual domain for the psychological phenomena. Once the conceptual domain is clarified as a ground, a construct can be defined, and some estimate of invariance for a set of items that pertain to the concept can be assessed. A theory of expected relations for the concept can be asserted, before the GLVM is employed. More will be said about the construction of such domains in the next Chapter. In the next section below, questions regarding statistical assumptions for the GLVM are explored, looking to where concerns may perhaps increase uncertainty, in respect of logical questions about application of the model.

3.2 Problems - applications

The concerns that are described above relate to theory for the GLVM; a series of distinct concerns arise with respect to the use of the model in empirical research circumstances. The computer revolution following the work of Lawley (1940), Jöreskog (1969), and Muthen (1984), also introduced new decisions and assumptions for a researcher to make about the relationships between and structure of manifest and latent variables, and their errors (Cliff, 1992). The mathematical concerns highlighted above point to the application of constraints for use of the model that apply in every circumstance where the GLVM is used. Concerns to do with statistical applications of the GLVM, however, typically crossover between the substantive context, or the study situation, to different properties of the statistical models, themselves. What will be needed here is specification of these research elements within the conceptual framework, for studies where there is relevance. Concerns identified here include between versus within subjects relevance of the latent variable, causal inferences, infinite populations, the problem of equivalent models and decisions regarding variable structure.

3.2.1 Between versus within subjects

Molenaar (2004) describes a problem connecting the latent variable, which is constructed from between-subjects data, to any singular aspect of within-subjects individual differences. This is because there is no logical relation that can be established between data collected at a particular point in time from a number of individuals, which is then analysed in latent variable models, with any kind of variation (say, in intelligence) as a latent variable for any one individual or intra-individual variation. Molenaar (2004) notes it is possible to add the assumption of ergodicity to facilitate any inferences made for any individual given the outcomes of between-subjects analysis. Ergodicity assumes that any individual is representative of the population. Ergodicity in this sense would mean that the population probability distribution and the individual probability distribution for the latent variable are assumed identical in the limit (Molenaar, 2004). Molenaar notes that there are good reasons to suspect that ergodicity cannot hold for psychological phenomena, because we know, for example, that psychological development implies non-stationarity for psychological phenomena, through time. This implies that even for the same individual, variable values should change in time. Following this line of thought, in responding to a claim by Weinberger (2015) that no between-subjects variation could occur without within-subjects variation of some sort, Borsboom (2015, p. 363) states that we would need to be able to specify some individual-specific outcome for the parameters of the latent variable for this to hold. This is simply not possible when conducting between-subjects analysis:

> Positions of individuals on the latent variable are not parameters in the model that can be estimated or meaningfully specified, even in the most restrictive models (e.g., Rasch, 1960); all that one can specify, estimate, or determine are differences between positions of individuals (at best, one can estimate the metric distances between these positions, but typically an ordering is more realistic). Thus, latent variables provide ordinal or at best interval representations of the individual differences under study (Ellis & Junker, 1997). They are not absolute or ratio scales, for which the actual values of a numerical representation have a direct empirical interpretation independent of the individual differences between measured entities (Krantz, Luce, Suppes, & Tversky, 1971), nor are such scales assumed in the background.

A relevant example from the patient-centred care literature in psychology research which is described in detail in Chapter 8 in a retro-fitted conceptual framework is the development of the Mood and Anxiety Symptoms Questionnaire over time (MASQ: Clark & Watson, 1991). A shortened version of the original 90-item MASQ, developed as the MASQ-D30 (Wardenaar et al., 2010) has been presented in a number of studies as connected to biologically-based markers for anxiety and depression. These include for example differential patterns in HPA-axis functioning (Wardenaar et al., 2011), and co-occurrence of metabolic syndrome diagnoses (Luppino et al., 2011). An important question which does not get addressed in either of these two studies is how the proposed within-person biomarker is connected to the cross-sectional or between-person analysis of the data. Such a connection is important, as other studies which have attempted to replicate full MASQ outcomes in clinical samples have not successfully reproduced the factor outcomes (see Boschen & Oei, 2006; Boschen & Oei, 2007). The risk then is that incorrect treatment protocols may be developed based on inferences that have no theoretical nor logical foundation.

The implications for such concerns in respect of developing the rationale for a conceptual framework is that a structure needs to be in place whereby the researcher can account for the decisions that they have made regarding the connection between the type of analysis that they are performing with the GLVM, and the way in which psychological events or phenomena occurs in individuals[13]. Much of the literature that addresses the between versus within subjects distinction speaks to concerns regarding whether ranking differences between individuals can be said to constitute an act of psychological measurement (see Borsboom, 2005, 2015). More will be said regarding the question of measurement for GLVM analysis in Chapter 6. Meanwhile, it is noted that in the conceptual framework, room for declaration of whether the study is within or between subjects in design stands to make completely clear whether a relationship between the data collected and some secondary model of the within-subject psychological process, should be said to apply at the substantive level.

3.2.2 Infinite Populations

One important aspect of the shift to statistical solutions for GLVM parameters is the reliance on assumptions specific to statistical inferences, distinct from directly mathematical assumptions. In the previous chapter, for example, it was clear how the shifts in Fisher's accounts of maximum likelihood solutions in terms of an increasing set of constraints for statistical efficiency, sufficiency, and consistency played out across the different versions of maximum likelihood solutions. Violations of normal distribution, to take up an example, are not uncommon in GLVM practices, and alternative techniques that

[13] Note that group-level latent variable modelling which does not attempt to connect to properties of individuals is possible and has been used in other fields beyond psychology – see for example Blei (2014).

do not rely on the assumption of normal distribution are available, such as weighted least squares methods (see for example Beauducel & Herzberg, 2006). Typically, in applications there is reliance on a statistical assumption in respect of normal population distribution. Normal distribution in a population as an assumption itself relies on an ideal of an infinite population, with random sampling in the limit (see Fisher, 1935).

Statistical inference generally is noted to involve statements about populations made from a set of observations that are usually construed as a random sample of observations, in such a way where some estimate of uncertainty is included in respect of the statements (Fisher, 1935; Cox, 1956). This distinguishes statistical inference from scientific inference, where the process of inference follows a description of certain facts regarding the population, to some further aspect of the phenomena in question, in a way that seeks to generalise details contained in the description to other instances (Trusted, 1979). Scientific inference is also called ampliative inference. It typically relies on ideals concerning operation of laws of nature (Trusted, 1979). Laws of nature once were considered in science to be inherently objective, such that any specific law could always be considered as observed to apply and in all places, in the same way. Such an attitude appears to be clear in the work of Spearman (1927, 1948) with his idea of noegenesis, discussed in Chapters 2 and 3. By the 1920s, it was beginning to be understood that the perceptual field within which observations occurred operated as a domain, constraining any operation of what were taken to be laws to a particular set of circumstances or event sequence(s) (see Hobson, 1923).

Domains are important when we consider statistical inferences. For example, Fisher (1935) attempts to connect statistical and scientific inference using principles of deductive or mathematical logic. To achieve the connection, and sustain the certainty associated with deductive closure, he notes that we must assume that the population from which sampling takes place in the statistical analysis is a fully *known* population. If the population is fully known, then likelihood is a function of specific solution values, for the parameters. Such a known population can be referred to as the domain of reference, for a likelihood or parameter solution outcome. What is noticeable for psychology research is that we neither have the certainty of the operation of natural law in specific contexts, nor fully known features for the population parameters for our statistical inferences. The domains of reference for our statistical inferences in respect of law remain largely, even today, unknown.

To pick up a relevant example in respect of the key nature of population characteristics for statistical inferences from the patient-centred care literature referred to above, administration of the MASQ items across distinct sample groups such as clinical and non-clinical populations has been demonstrated to produce different outcomes in terms of both the correlations relied on for inter-rater reliability, and the model fit information that follows from latent variable modelling analysis (see Boschen & Oei, 2006; Boschen & Oei, 2007). Reidy and Keogh (1997) had attempted replication of the factor structure of the MASQ with a sample of British students, and had found outcomes consistent with the earlier studies of Watson et al. (1995) for these items. This earlier study had involved three groups of students, as well as a non-clinical and clinical sample of participants. The clinical sample however was a group of individuals in a substance use clinic. The later study of Boschen and Oei (2007) examined the MASQ factor analysis outcomes for individuals who had a specific clinical diagnosis of anxiety or depression - presumably the individuals of most relevance, for this particular scale. Boschen and Oei (2007) found that none of their hypotheses were supported, which were all aimed at confirmation of the tripartite model of depression and anxiety that was proposed in the original MASQ literature (see Watson & Clark, 1991).

Questions regarding the domain of reference for statistical inferences in psychology research generally are well-represented in the early psychometric literature. These however have largely disappeared, today. For example, Lord, Novick, and Birnbaum (1968) noted that the assumption of random sampling with replacement for an infinite population involves assumptions about brainwashing an individual and putting them back into the infinite population ‘pool’ for each sample selection in respect of the psychological properties to which the psychometric analysis refers, as discussed above (see also Borsboom, 2005). Infinite population statistics also have useful mathematical properties such as stochastic independence and normal distribution - properties that facilitate further statistical inferences. However, it is not the case that psychological phenomena typically occur within fully known populations, nor fully articulated parameter-sets. This then raises questions concerning the statistical inferences in psychometric practices. One learning that emerges from the development of mathematical set theory that occurs around the same historical moment that Spearman develops his factor theory are that there are different forms of infinity, such as countable and uncountable infinity (see Cantor 1882/1932; Ferreriós, 1995). These different types of infinity imply the use of different kinds of mathematics or statistics, as a new kind of number had come into existence that for example may have a helical or circular rather than linear construction (Ferreriós, 1995).

In the evolution of statistical methods for psychology research, use of the GLVM demands for example that we specify a countable infinity for the population, where individuals are considered as *events* (Ellis & Junker, 1997). Conceptualising the population for whom a research finding is relevant as countably infinite, and then conceptualising individuals in a population for whom a research finding may be relevant as events plays a vital role in the interpretability of findings made following GLVM analysis. These practices are not explicitly supported in the use of the GLVM as it occurs in psychology research, today. Themes introduced in Markus (2008) emphasise the importance of clearly specifying the population domain over which the researcher expects inferences from processes such as those utilised with the GLVM will apply. These will be discussed in more detail in Chapter 5.

Connecting to ideas about handling infinity in psychometric practices, McDonald (1999) introduced the idea of an infinite behaviour domain connected to the construct behind the test items, rather than the population for which the construct is relevant. He endorses this approach of setting a countably infinite domain to articulate the conditions belonging to the phenomenon of interest, which we attempt to track with the latent variable model. In this way, a rationale is provided for adding items to a questionnaire, for example, used in the construction of a test that notionally tracks the phenomenon of interest. Domain setting in a conceptual framework to help provide clear conceptual linkages between the domain of the finite, the domain of the infinite, and the assumed type of infinity will be an important feature to be supported in the set-theoretical structures laid out in subsequent chapters.

3.2.3 Causality

We see advocacy for causal interpretations of the GLVM model, particularly from the mid-twentieth century in the psychometric literature. Causation enters considerations for the GLVM insofar as researchers posit an explanatory role for the latent variable that is also causal, in what has been labelled as the reflective latent variable model (Edwards & Bagozzi, 2000; Borsboom, 2005). This interpretation of the GLVM is consistent with the idea that the latent variable causes the test score outcomes seen in the manifest variable, such that, for example, intelligence is considered to cause test scores, similar to the idea in medical science, that smoking causes lung cancer (McDonald, 1985; Markus & Borsboom, 2013).

Causation and explanation are unified in the causal accounts of the GLVM, in a way consistent with views popularised in philosophy of science literature, particularly by Popper (1935/1992). The popularity of such an interpretation follows the acknowledge failures of logical positivism regarding the status of causation and explanation, in scientific practice (c.f. Borsboom, 2005; Langer, 2015). Hartley (1954) notes in the mid-twentieth century advocates for factor analysis, including Burt (1941) and Kelley (1940), were relatively disparaging regarding the possibility for causal inference from latent variable modelling techniques. For them, the process of latent variable modelling consisted in description of a phenomenon, in a systematic way, using the techniques to interpret correlation patterns. This closely follows Spearman's (1904) stated objective in searching for some pattern in test scores that helped to constrain or detail a definition of intelligence or general ability.

We have seen already in our section on McDonald in Chapter 3 an argument that a common factor cannot be a cause, as it must be possible for us to imagine an infinite set of additional items for the factor or latent variable, in the infinite domain interpretation. The conditions for independent and unique recognition of cause, and effect, would thus be violated, in McDonald's (1985, 1999) perspective. Others have noted that any causal inference relies on extra-statistical assumptions, whether SEM is used, or not (see Breckler, 1990; Freedman, 1987). Statistical modelling itself does not provide enough evidence for a causal relation, for these authors, precisely because psychological phenomena are unavailable for independent confirmation, with respect to the operation of causal mechanisms. Freedman (1987) further notes that it is not only the case that error is unknown (as is the case where the model is interpreted as carrying a random variable for error), but also, the latent variable itself is unknown, or unable to be directly connected to the phenomena in question for the researcher. In this way, it is far from the case that data confers error-free observations, as proposed in Borsboom (2008) in latent variable modelling. Rather, data may obscure several errors, not all of which may be able to be tracked or estimated by a researcher. With a conceptual framework, some headway can be potentially made on separating and refining the kinds of information that is associated with error, whether this be model error (wrong model for theory), variable error (wrong variable or mix of variables of import for the psychological phenomena in question), or data error (error associated with the actual event of the research study). Distinguishing this kind of information is relevant to establishing independent evidence for the independent existence of the psychological phenomenon represented by the latent variable, beyond what is discerned from application of the GLVM. Extra-statistical evidence such as this that separates the cause from the effect can serve a crucial role in supporting causal claims (Boag, 2007, 2011; Maze, 1983, Passmore, 1935). Such evidence can be presented in a research element relevant to the research project, the details of which are described, in Chapter 6.

3.2.4 Equivalent models

Model equivalence (ME) is the notion that two or more different models can just as effectively describe the same phenomenon, or factor, at the level of the mathematical or statistical model (see Lee & Hershberger, 1990; Bentler & Chou, 1987; Raykov & Penev, 1999). The problem of equivalent models is detailed in earlier literature as occurring in two guises (see Borsboom, Mellenbergh & van Heerden, 2003). The first is a choice of model from those available under the common framework. Raykov and Marcoulides (2001) demonstrated, for example, that in SEM an infinite series of models that make substantively different assumptions about the nature of the phenomena can be constructed (see also Stelzl, 1986), yet exhibit the same goodness of fit statistics. In these models, the nature and number of latent and manifest variables, their error terms, and their relations can be varied according to researcher-specified rules. A logical problem exists for equivalent models, however, insofar as an acceptable model is one that the data fails to disconfirm (Cliff, 1983; Breckler, 1990). We are thus left in a situation of uncertainty regarding whether the latent variable model in question is the best model, or whether a different model would represent a more accurate account of the phenomena and its relations.

The second kind of equivalence can also occur where identical fit statistics may be found for different parameter restrictions applied to the same structural model (Raykov & Penev, 1999; Breckler, 1990). In this case, parameter estimates imply identical covariance matrices with the same goodness-of-fit statistics, for different formulations of relations between variables, and variable error terms (Pearl, 2000; Hershberger, 1994). Where model equivalence is found in empirical outcomes in any form, it is argued by Hershberger (1994) that solutions regarding model choice are found in assessing substantive knowledge regarding the psychological attribute, where the greatest benefit "lies in the ensuing obligation imposed on the researcher to gather further evidence" (p. 104) to justify the researcher's selection of model. This accords with suggestions by Raykov and Marcoulides (2001, 2007), whereby information of the non-statistical type is the only way to distinguish between equivalent models, since statistical analysis will not provide the justification necessary for the decision to adopt one over the other.

The problem of equivalent models has been compared to the problem of underdetermination of theory by data (Borsboom et al., 2003). Underdetermination of theory by data occurs where two theories appear to "be logically incompatible and empirically equivalent" (Newton-Smith, 1978, p. 71). Quine (1970) suggested that *all* theories must be underdetermined by data, insofar as several theories can fit a specified data set equally well. Newton-Smith however concludes that the suggestion of equivalence of theories or models is likely misconstrued. For Newton-Smith (1978, p. 78), it is not ever the case that two theories are equal as a set of deductive postulates, because there are different underlying conditions and assumptions, which were assumed as truth conditions in the construction of each model. These underlying conditions and assumptions should be evaluated in each instance with respect to the proposed context of application, of the theory (Newton-Smith, 1978). Substantive considerations relevant to the context are proposed as best guide regarding the relevance of specific theories in specific circumstances. Psillos

(2009), with a different perspective, proposes balancing first-order evidence about substantive model plausibility with second-order evidence that references soundness of methodological conditions and assumptions. No matter which proposal a researcher follows to justify their choice of model or theory, a conceptual framework for use with the GLVM analysis in psychology research projects will need to offer infrastructure that supports attention to detail in respect of the context of the research. Details in respect of such a conceptual framework are described in the present work, in Chapter 6.

3.2.5 Variable structure

As seen in Chapter 2, one aspect in common between the quite distinct proposals for the unified GLVM from McDonald (1999) and Bartholomew et al. (2011) was the idea that the modelling process can be employed to help to determine what sort of structure is relevant for the latent variable. There is historical persistence through to more recent times with the idea of determining the latent variable structure by referencing patterns in the data. For example, Lubke and Neale (2006) propose that the GLVM can be used to help determine the nature of the underlying variable with respect to scale type. They note there is one commonality that extends across all different types of scale structure for the manifest variables. This is insofar as there remains an interest in explaining the pattern of covariances between the manifest variables. They further note that the question of whether clinical diagnoses differ quantitatively or qualitatively remains unresolved. While there is a clear theoretical or conceptual distinction to be made between qualitative and quantitative data points, Lubke and Neale argue that at the level of statistics, the difference is approximately trivial (see also Bauer & Curran, 2004), meaning that the distinction receives little attention, in present day research that makes use of GLVM processes. Such a distinction is a crucial one for psychological phenomena; however, as it may for example indicate different treatment protocols for psychological diagnoses (see for example McClimans, Brown, & Cano, 2017). McClimans et al. (2017, p. 72) note that there remains a paucity of attention in psychological research to what they describe as “attribute theory” – robust theoretical formulations for psychological attributes. Such formulations themselves would provide guidance on the structure that is appropriate for the latent variable, and reference points for how values on the latent variable can be connected to the research phenomenon in question. In carrying through to the conceptual framework, it is noted that flexibility will be a key characteristic that is demanded when it comes to accounting for variable structure. This is because different processes are likely to involve different decisions when it comes to the latent variable structure There will presumably be some reference back to earlier research literature, given the conclusions about the lack of attention to the development of attribute theory in psychological research that exists, today (McClimans et al., 2017; Slaney, 2017; Borsboom, 2006).

3.3 Problems as quality control – instrumental certainty and measurement validity

In *Test Theory*, McDonald (1999) devotes four out of the eighteen substantive chapters to analysis of reliability and validity techniques for the purposes of test development, and another two chapters to techniques that support assessment of the model by testing it for fit, invariance, equivalence, and unidimensionality. One way to understand these processes when undertaken in the conduct of a research project that utilises the GLVM is in effort to manage and address the quality uncertainty of research outcomes in respect of the instruments used in the reporting of the research outcomes, which are often questionnaire or self-report instruments. Historically such practices are described in the literature as measurement validity assessments (Borsboom & Mellenbergh, 2004; Markus & Borsboom, 2013; Slaney, 2017). In describing the techniques that are employed in psychometric practices that make use of the GLVM in what follows, what is sought is understanding of how the conceptual framework can support transparency regarding researcher decisions and practices in respect of quality control with regards to research claims and outputs.

3.3.1 Unidimensionality index

First described in the literature as a concern with the homogeneity of tests (Loevinger, 1948), the problem of unidimensionality addresses the question of whether a latent variable is structured in such a way as to adequately describe a single dimension for psychological phenomena such as intelligence or neuroticism (Gerbing & Anderson, 1988, p. 186). Unidimensionality has two tests. It is attained statistically when a single latent variable accounts for the statistical dependence of all items in a set, and when there is evidence that the items maintain statistical independence for any homogenous subpopulation (Hattie, 1985; Heene, Kyngdon, & Sckopke, 2016, p. 1). Any assessment aiming to measure individual differences on a single attribute demands unidimensionality and violation of the unidimensionality assumption will lead to wrong conclusions about either the nature of the underlying construct for the latent variable, or about persons with scores on the latent variable (McDonald, 1981). Lumsden (1961) further notes that unidimensionality applies where all items can be said to be assessing the same phenomenon. It is also possible to consider what sort of answer pattern may be expected in a series of items for unidimensionality to hold. For example, a Guttman scale involves a set of items of increasing difficulty, which may for example assess a single ability such as mathematical ability (Stouffer et al., 1950). It may not be necessary however for a candidate to get each item in the assessment correct, to be assessing mathematical ability. The assessment of unidimensionality then is usually conducted in probabilistic terms, looking to the probability of correctly answering an item, given an assumption of a specific, singular phenomenon that the item assesses (Hattie, 1985). The question regarding how unidimensionality is attributed to a phenomenon via assessment outcomes or scores is an important one, because as noted by McDonald (1981), unidimensionality cannot be said to logically hold in degrees. The term itself implies a strict criterion of true or false.

Consistent with our points about circularity and conditional independence above, Lumsden (1961) notes circularity is evident in practices that attempt to use latent variable modelling to identify manifest variables or test items as formulating a singular dimension or construct. This is because some initial thought goes into items that would belong together as a construct, and then under assumptions of conditional independence given the latent variable, we confirm unidimensionality to say together they describe a singular psychological phenomenon.

Heene et al. (2016) note, however, that it is possible to have more than one psychological or physiological phenomenon underlying a unidimensional outcome, whenever local independence is assumed for a set of manifest variable outcomes conditional on some latent variable (see Suppes & Zanotti, 1981). That more than one phenomenon may be involved in some assessment of a latent variable may or may not be clear in the model structure that is chosen by the researcher. For example, shared error variance in a confirmatory factor analysis may represent some phenomenon or psychological construct that is known to the researcher, but not declared in the model. This may be, for example, where a confirmatory factor analysis assesses intelligence test outcomes, but shared error variance amongst the different types of abilities represented in the manifest variables such as mathematics, verbal, or spatial skills scores may be known to represent participant’s level of motivation regarding the test.

Following the sections above, there is no simply discernible relation between the psychological phenomenon, and the assumed structure of the latent variable. The conceptual framework should have within it structures that support clarification of item relations that are considered to hold, prior to assessment of model fit to any data collation. The exact address of this in the framework will be made clear in Chapter 6.

3.3.2 Model fit

Model fitting is a concern relevant to techniques falling under the SEM family, insofar as a particular set of relationships is specified between manifest and latent variables with their respective error terms, and then a test is performed to see whether the data supports or does not support this particular model formulation (Thompson, 2004; Ropovik, 2015). There nevertheless remains no consensus on the best indicators of good model fit in the field (Ropovik, 2015; Myung & Pitt, 2002; Thompson, 2004). The most commonly used test to check global model fit is the χ 2 test (Cochran, 1952), but this test is dependent on the sample size: it rejects reasonable models if the sample is large, and it fails to reject poor models if the sample is rather small. There are three other types of fit indices typically used to assess fit of a model. First, the comparative indices that compare the fit of the model under consideration with fit of baseline-model, such as, for example, the Tucker-Lewis Index (TLI: Bentler & Bonnett, 1980), and comparative fit index (CFI: Bentler, 1990). Fit is considered adequate if the CFI and TLI values are > 0.90, and better if they are >.95. The TLI attempts to correct for complexity of the model but is somewhat sensitive to a small sample size. In addition, it can become > 1.0, which can be interpreted as an indication of over-fitting, and making the model more complex than needed. Second, there are absolute indices that examine closeness of fit, such as, for example, the root mean square error of approximation (RMSEA: Steiger & Lind, 1980). The cut-off value is RMSEA < 0.08, and is better when <.05. The RMSEA is insensitive to sample size, however, but sensitive to model complexity. Third, there are information-theoretic indices, for example the Akake information criterion (AIC: Akaike, 1973) and Bayesian information criterion (BIC: Schwarz, 1978). Both can be used to compare competing models and can make a trade-off between model fit (i.e., -2*log likelihood value) and model complexity (i.e., a computation of the number of parameters). A lower IC value indicates a better trade-off between fit and complexity. There is no rule of thumb, as the values depend on the actual dataset and the function simply chooses the model with the lowest IC value (see Kline, 1998).

It is entirely possible that reliance on model fit statistics will result in what is called over-fitting or selecting too many latent variables as fitting for the data set (Pitt & Myung, 2002). The reason this comes about is that model fit heuristics, as typically employed, reward increased model complexity. Such a phenomenon sits contrary to the ideal of parsimony often expressed in connection to the GLVM, where we saw for Bartholomew et al. (2011) that parsimony or reduction in the number of variables formulated an entire rationale for the use of the GLVM. What the conceptual framework will need to provide is space for articulating a rationale for model choice that extends beyond mere statistical criteria – given that the criteria are heuristics and involve elements of subjective judgement by necessity. Where a researcher can provide their extra-statistical qualitative reasoning, future researchers are better placed to adopt, adapt, or ignore earlier research efforts, as fits the goals of their own research programs.

3.3.3 Measurement Invariance

Measurement invariance (MI) addresses the question of whether one latent variable assessed in a research project is the same as a latent variable assessed in a different context such as a different cultural group or at a different time-point (Milfont & Fischer, 2010; Floyd & Widaman, 1995). MI implies that the same factor structure is relevant for different subgroups within a population and/or longitudinally, given a set of manifest variates and latent variable(s) (Widaman & Reise, 1997). Meredith (1993) works through a series of theorems and proofs that hold in the general case for latent variable approaches. These theorems state that where conditional independence is in principle assumed to hold across a sample space both for a set of items and a population, MI is demonstrated for continuous and discrete variable types. This is when the latent variable and manifest variables can be interchanged in equality between two or more different groupings or subsamples taken from an overall population (see Lord, 1980). Meredith (1993) is careful to note here, however, that in terms of theorems and proofs, the notion of measurement invariance is an idealisation, because of the conditional independence assumption. By the time of publication of Meredith and Teresi (2006), we see recommendations in the literature that any decisions following a GLVM analysis, such as, for example, resource allocation, should be made with reference to manifest variable outcomes rather than latent variable outcomes. This is simply because the latent variate solution is formulated from an underlying variable and, as such, implications of its influence are simply not available to be observed. Meredith and Teresi (2006) do observe that some analysis of different patterns in the manifest variable values may add to scientific understanding. This may be where we may expect group differences in specific or manifest variable values that contribute to the existence of some illness such as hypertension, which is understood to underlie values for the latent variable.

Different types of MI have been further specified in Meredith (1993). For example, weak or pattern invariance is where there exists an invariant latent variable loading pattern across groups. Without this as a basic requirement, group comparisons regarding performance cannot be made (Meredith & Teresi, 2006). Strong invariance, on the other hand, adds an additional constraint whereby the latent variable means must be identical across groups, while strict invariance implies that there are also no group differences. If strict or strong invariance holds, it is possible to say for the groups tested that their combination into a single group will yield the same pattern matrix for the items and manifest variate combinations. No matter the type of invariance, however, MI always remains a function of an empirical test, and the assumptions behind the models relevant to the formulation of the test (see Meredith & Teresi, 2006). In this way, the conceptual framework must clarify these assumptions, for us to see into the relevance or otherwise, of MI analysis outcomes.

Principle component analysis, exploratory or multi-group confirmatory factor analysis (MGCFA), or multidimensional scaling can all be used to assist with evaluating invariance, but these techniques are not interchangeable, and the conceptual framework should provide the opportunity to note which technique was used, and with what reasoning.

Several other types of invariance are noted in the literature. While the notion of factorial invariance, the idea that the same set of factors should be found for different manifest variables that are indicators for the same phenomena, is historically attributed to Thurstone (1947; see Mulaik, 2010), the concept of MI was introduced by Byrne, Shavelson and Muthen (1989). These authors observe that different types of MI can be assessed. For example, using statistical methods, metric invariance is confirmed when only the factor loadings are equal across groups, but the intercepts can differ between groups. This tests whether respondents across groups attribute the same meaning to the latent construct under study (see Vandenberg & Lance, 2000). Scalar invariance on the other hand involves holding both factor loadings and intercepts constant, with scores on the latent variable then compared (Milfont & Fischer, 2010). Other types of measurement invariance demand more qualitative means for analysis. For example, functional equivalence asks whether the psychological construct in question exists in all groups in the same way, while structural equivalence evaluates whether indicators are meaningfully related to the psychological construct in question (Fontaine, 2005). Functional equivalence is a basic requirement for any other type of invariance to be present yet can only be accounted for in qualitative terms. An important aspect of the conceptual framework then will be the support of the maintenance of qualitative information that supports any quantitative analysis as relevant to either the model parameters, or evaluation of the model performance, with respect to parameters.

3.3.4 Reliability and validity

Techniques such as calculating an indicator of internal consistency via a Cronbach's alpha quotient to assess reliability for a questionnaire or set of test items, or testing correlation coefficients to assess convergent or divergent validity in GLVM processes and in psychological research generally are well established in the psychometric literature (Davis et al., 2018). Typically, these techniques are cited in the context of measurement of psychology phenomena (see for example, Fried & Flake, 2018; Slaney, 2017). It will be shown in Chapter 6 that there is good reason to pay close attention to the question of whether use of the GLVM in empirical contexts can constitute an act of measurement. Beyond this question, the practices that constitute psychometric measurement validity techniques are presently under scrutiny from within the field itself. Critique includes everything from researchers misrepresenting reliability checks as validity checks (Flake, Pek & Hehman, 2017), to deconstruction of the traditional techniques used in survey validation (Maul, 2017), to a call for an end to the ideal of construct validity given the seeming historical failure of it through time to actually produce replicable construct validity (Borsboom, Cramer, Kievit, Scholten, & Franic, 2009).

The original aim of these reliability and validity practices was to provide assurance about the quality of the research outcomes as reported, usually for a questionnaire or battery of test items. These may directly exist in relation to the latent variable, such as in EFA, or which when accumulated may function as an indicator value for a manifest variable, in say CFA. Generalisability theory aims in a similar way to assert the replicability and reliability of research findings given specific aspects of research contexts (see Cronbach, Nageswari, & Gleser, 1963). A shift is noted in the psychometric literature, where earlier forms of specifically validity practices were aimed at interpretation of test outcomes and epistemology, or how we can know about psychological phenomena (Messick, 1989). Later practices however are proposed as focused on instrument assessment and ontology or questions on being and truth for psychological phenomena (Borsboom, Mellenbergh, & van Heerden, 2004; Hood, 2009). Wilson (2013) notes that it is possible that psychometric practices embrace either or both positions. Interpretation may be important where there is little substantive theory developed for the phenomenon behind the latent variable, questions pertaining to instruments are much more important when theory is established and psychometric testing for the phenomenon has a well-developed track record (see also Borsboom, 2006). For the conceptual framework, flexibility will again be demanded, to facilitate the inclusion of techniques under either of these rationales. The outcomes of any statistical analyses employed in support of research claims are less important than the account of the researcher's decisions about what should be adopted in this capacity, for the purpose of transparency regarding researcher decisions in respect of reliability and validity practices. Some examples of these techniques reported in the framework will be provided in Chapters 6-8.

4. Summary

A summary table for all the concerns presented in this chapter is included below (see Table 4). In the first part of this Chapter, it has been noted that three key concerns are relevant to any use of the GLVM. Firstly, the conditional independence constraint defines the model, but it is always possible to find a common factor given the conditional independence constraint for a set of variables, following Suppes and Zanotti (1981). More evidence is needed for the application of conditional independence for data that is relevant to a latent variable, which is extra-statistical, and which is beyond the GLVM model, itself. The second problem discussed was the mathematical problem of factor indeterminacy, where we are unable to connect the psychological phenomena in question to the mathematical model just through use of the model itself. Again, extra-statistical reasoning or evidence is required in this regard. Thirdly, the logical problem of circularity prevails for the GLVM, meaning that nothing more can be said about the latent variable than was involved in the original model assumptions. In common among these three concerns is the reliance on extra-statistical information to support any scientific inference. That theme has been repeated as we have visited with concerns regarding applications of the GLVM and as quality control practices have been assessed. Combined with the themes from Chapter 3, what is indicated for the conceptual framework is not only a tool that facilitates disclosure, but also a tool that facilitates the tracking of effects that follow from decisions made in the process of conducting and representing the phenomenon of interest in the research project. These decisions are of influence not only in solving the kinds of problems that are involved in empirical research, they have implications for the reporting of research outcomes, and evaluation of these as practices. In the next chapter, differences that philosophical perspectives bring to the production of GLVM outputs are pursued with similar questions in mind – do these perspectives provide some resolve, for any or all the problems, we have encountered, so far.

Table 4

Problems for GLVM in research applications

Type	**Problem**	**Specific issue**
Theoretical	Conditional Independence	It is always possible to find a relationship between variables, whenever the assumption is made.
	Factor Indeterminacy	Cannot connect the construct of interest to the latent variable with determinacy.
	Circularity	All that can be said about the latent variable is a function of relations with the manifest variable.
Application	Between vs within subjects	GLVM calculations performed on cross-sectional data cannot be said to apply directly to within-subject processes.
	Causation	Must rely on extra-statistical assumptions to assert the phenomenon which the latent variable represents as an independently existing phenomenon which can function as a cause.
	Equivalent models	No pre-specifiable criterion within the GLVM itself establishes which model best fits the substantive situation.

	Variable structure	The latent variable is precisely, latent - it is not available for confirmation regarding the most appropriate structure, in respect of the data.
Quality control	Unidimensionality index	Unidimensionality strictly speaking is a true/false characteristic for some construct - there is a logical confound in presenting an index.
	Model fit	Assessing model fit between a prespecified model and the data always involves use of heuristics that are outside of the GLVM itself.
	Measurement invariance	Even though indices are produced in evidence of invariance, practices must be supported by assertions that can only be stated in qualitative terms.
	Reliability and validity	Researchers can choose what, if, and how they perform reliablity and validity checks; the psychometric field itself now is scrutinising these practices for efficacy.

Chapter 5: Realism, Set Theory, and the GLVM

1. Objectives

It is noted in the literature that psychological researchers generally are realist in their research activities, insofar as they treat psychological phenomena as existing in the world, independent of the researcher's thoughts about the phenomena (Hibberd, 2014; Michell, 1999). In this chapter, there is a return to the theme of realism for the GLVM, with two objectives. Firstly, realism for the latent variable model is endorsed across several recent accounts that attempt to provide a philosophical basis for its use (see Borsboom, 2005; Haig, 2014; Maul, 2013; Hood, 2013). This chapter initially explores realism as endorsed particularly by Borsboom and Haig, two recent proponents for integrating philosophy of science with methodological considerations about the use of the GLVM. Each author proposes distinct accounts of realism, which are shown to have different consequences for the kinds of activities involved with GLVM analysis. The consequences of the specific realism endorsed by each author is pursued for its implications in a representation of the research outcomes in a project making use of the GLVM.

Following this, realism is taken up as a theme, in order to arrive at a view of the structure that is demanded in the conceptual framework when we take the situational context of the research as a real foundation for the representation of the research project and the GLVM analysis, no matter the philosophical perspective chosen by the researcher. This perspective is adopted because all research projects, no matter the philosophical underpinnings, occur in situations in space and time. They are unified in this regard (see Petocz & Mackay, 2013). The situation for a research project marks a domain of invariance (see Suppes, 2002), the representation of which marks an initial constraint that is relevant to setting the context for the research project, and the articulation of research claims.

Markus (2008) provides a touch-ground for the reasons that contextualisation is vital for use of the GLVM in research, insofar as researchers wish to remain coherent in claims following the conduct of their inquiry. Contextualisation in a geo-historical setting is explored in the set-theoretical structure proposed by Markus (2008), prior to closing the chapter with a brief historical overview of the development of set theory itself.

Methodological and philosophical arguments for realist interpretations of latent variable modelling are presented in two prominent texts in the psychometric literature, these being *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics* by Denny Borsboom (2005) and *Investigating the Psychological World* by Brian Haig (2014). In the work of both Borsboom and Haig, there is a contrasting of their preferred philosophical realist stance with other variations on ideology relevant to psychometric practices, such as instrumentalism, operationalism, pragmatism, social constructionism, and constructivism. These distinct philosophical stances are pursued for their consequences for what can be said following a GLVM analysis. In Borsboom (2005, p. 58), for example, constructivism is characterised as a position that “regards the latent variable as a construction of the human mind, which need not be ascribed existence independent of measurement”. This would mean that psychological constructs are merely “fictions” (p. 60), presumably leaving any GLVM analysis with little scientific import.

The realist proposals from Borsboom and Haig as described below address a metatheoretical perspective on psychology research practices (see Borsboom, 2008), extending beyond mere application of psychometric models, to address psychology research and psychometric methodology. Metatheory is defined following Witherington et al. (2018) as the background theory that is assumed on a first-principles basis. It is the conceptual background against which ontology, or claims about what is, and epistemology, or how we can know about what is, is set, for research[14]. Metatheory works to

> establish what does and does not make sense to even consider or investigate in the observations that we make and the theories that we construct. All of our scientific work, therefore, necessarily presupposes, and is preconditioned by, the background concepts of metatheory (Witherington et al., 2018, p. 183).

Metatheory for psychometric practices, such as the scientific entity realism of Borsboom (2005), the constructivist realism of Markus and Borsboom (2013), or the naturalised methodological realism of Haig (2014), each may offer researcher resources that they consider best supports their psychometric practices. While the realisms listed may serve in this respect, there is no point presently at which during a research project that the researcher systematically describes the chosen perspective and traces the consequences that follow from the perspective through to the claims made from GLVM analysis. In what follows, it can be observed that when it comes to a philosophical perspective on the use of the GLVM, different stances demand that researchers utilise different processes, interpretations, and reporting structures, in the production of research outputs. The constraints so set by adoption of a philosophical perspective deserve, then, full articulation for their role in the research project. This is the point of the systematic conceptual framework, described in Chapter 6.

2. Present day philosophical and methodological argumentation for LVM

In the wake of the computer revolution and software advances that took place in the mid-20th century as discussed in earlier chapters; exponential growth occurred in the

[14] Witherington et al. (2018, p. 185) state that that metatheory is “pre-empirical” and “pre-theoretical” (p. 185). It is the position of the present author that there are good reasons to consider that any conceptual grounding has some empirical and theoretical input (see Jenkins, 2008). A commitment to an ideal of pre-empirical or pre-theoretical first principles is not necessary for the adoption of the conceptual framework described in this paper.

uptake of latent variable modelling as the favoured tool for statistical analysis in psychology research projects. What followed at the turn of the 21st century was the emergence of a series of rationales for philosophy of science-informed approaches, relevant to the GLVM. One of these is the scientific entity realism of Denny Borsboom, described in his 2005 book, *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics.*

2.1 Borsboom's scientific realism

2.1.1 Summary

In his 2005 book, Borsboom devotes one chapter to the latent variable model, advocating for it as a measurement model for psychological phenomena, which represents an advance over classical test theory (CTT) for psychometrics (p. 49). This is because in his interpretation, latent variable modelling overcomes the limits of operationalism inherent in CTT. Operationalism as a variant of constructivism under CTT implies that a completely new construct is investigated every time the CTT model is utilised in a research project. For Borsboom (2005), the latent variable model overcomes the limits associated with conceptualising new constructs for every study, by presuming a kind of constancy in virtue of which data can be assessed via latent variable modelling (see Borsboom, 2008, p. 49). The power of the latent variable model for Borsboom consists then in the inclusion of the latent variable in a function that can serve in an explanatory capacity for psychological research (see 2005, p. 51), with the idea that it is the same latent variable that features across different studies. Specifically, Borsboom (2005) connects explanation directly to causal reasoning based on measurements in psychology research, such that the latent variable is considered to play a causal and explanatory role for the manifest variable outcomes, in the reflective model following Edwards and Bagozzi (2000). In this way, for example, intelligence is considered to cause test scores.[15] For Borsboom (2005, p. 144), in a latent variable model "measurement is a causal relation between the latent variable and its indicators". The implications for the GLVM as a theory of measurement is explored later in this thesis, in Chapter 6.

Borsboom (2005) goes on to argue that because psychometricians treat latent variables as real (p. 52), a specific variant of scientific realism - entity realism (see Hacking, 1983) - is an appropriate philosophical position relevant to the use of the latent variable model. Entity realism is also called selective realism (Psillos, 1999), and in the version described in Hacking (1983) involves eschewing judgement about realism for theories, but accepting realism about entities via which scientists can create effects. For Borsboom (2005), the latent variable has a role in creating effects seen in the values of the manifest variables. Borsboom also sets out arguments against the perspectives of constructivism and operationalism in connection to the psychometric use of the model. Given the use that is made of the latent variable model in practice, Borsboom (2005) argues for a causal, realist ontology for the latent variable (p. 68). The account of scientific realism given in this book is cited as that of Devitt (1991) and Hacking (1983). There are three core claims: (i) that theories are either dichotomously true, or false; (ii) that theoretical entities exist, and (iii) that these entities can be causal in nature. Ultimately, Borsboom (2005) suggests that theoretical realism should not be adopted for the latent variable model, but that entity realism indeed should be, based on how researchers treat the latent variable in their research projects.

It is not sought here to defend nor attack any of these three claims in reference to the latent variable model. It is interesting to consider, however, what sort of counter-claim

[15] It should be noted that it is not the case that Borsboom (2005) treats this causal interpretation as unproblematic. For example robust coverage of the problem of interpreting between-subjects research outcomes as relevant to within-subjects psychological processes is presented in this work, with further coverage in Borsboom (2015).

may be levelled against the position, in such a way as to know what will be necessary to consider when constructing the conceptual framework. For i), for example, as we have seen in Chapter 3, asserting truth or falsity for the latent variable model remains challenging because Spearman's factor theory remains even today, without proof. For ii) and iii), Borsboom assumes that because of the way that practitioners employ the GLVM in psychometric techniques, practitioners must believe that the latent variable exists, and that the latent variable has causal efficacy. Entity realism is adopted here insofar as Borsboom (2005) surmises that the researcher must ascribe to the principle that the latent variable exists, independently of measurement, in order to make use of the model in the way that they do.

What is also noticeable in this account of realism is a tendency toward conflation between the underlying psychological attribute, and the latent variable itself, into a singular and notionally, real, entity (see Borsboom, 2005, p. 134). Other researchers have identified such conflation as illegitimate concept equating (see Maraun & Gabriel, 2013). What this means is that the psychological phenomena and the mathematical latent variable are treated or represented as the same real thing. Such a feature has implications for generalisation of study findings where the GLVM has been employed in analysis. Following Maraun and Gabriel (2013), these problems come about in part because the psychological construct does not mark a natural kind but has a number of contextualised features bound into the structure of the construct, which include researcher's ideas about how to measure the psychological phenomena that is of interest. This is distinct from the latent variable, which takes its existence in formulation of a mathematical model based on the constraint of conditional independence, as discussed in Chapter 4 of the present thesis. In the section addressing the work of Markus (2008) below in the present chapter, some resolutions are explored for the problems that follow from conflating the construct and the latent variable.

Another immediate concern arises for entity realism with the conflation of the attribute and the variable in the model, where ontological significance is assigned to the latent variable. Because of the way that psychology researchers seem to use the latent variable model, and given the limitations of constructivism and operationalism, Borsboom (2005) concludes that a question of ontology, or of "what is" regarding the latent variable is answered best with reference to entity realism. However, there is a problem with referencing psychometric practice in this way. The concern is, with such a claim for ontology for the latent variable, there is no room to question the methods used in the production of latent variable modelling outputs. If ontology is assigned to the latent variable in the model based on how the model is used, there is no meaningful way to connect the latent variable to the psychological phenomena with reference to the same philosophical position or perspective. A broader framework is needed which makes possible some assertions with independent evidence for the psychological phenomena, as has been described as needed to solve several the problems described in Chapter 4 of the present thesis.

With reference to this point, it is possible to progress use of the GLVM without deigning any ontological or entity significance to the latent variable at all. It is possible, for example, to use the GLVM in a social constructionist paradigm, without reference to ontology or existence of a particular entity that has effects in the world. This would be the case where the meaning of the latent variable may be designated as entirely a function of semantic agreements about, say, the nature of political persuasion indicated via item endorsement for a group in a cross-sectional study (see Feldman & Johnson, 2014). A researcher then is not making any ontological assumptions about any real-world existing entity that is under scrutiny. Rather, the researcher may be exploring how meaning might be situated within a series of structures and variables in formulation of the network of responses to items that make up a variable representing political ideology. A broader framework that facilitates integration of distinct philosophical stances into the GLVM process and analysis as an element of a psychology research project is indicated as supporting any claims that may be made about the meaning of research outcomes, following completion of the project.

2.1.2 Recent accounts from Borsboom

Borsboom appears to abandon his position of scientific realism with specific attention to entity realism by 2013, with Borsboom's publication with Keith Markus of *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. In this text, there is advocacy for Messick's (1981, 1989) approach to constructivist realism, which itself followed Loevinger (1957). This is a position that recognises roles for researchers and scientific communities in the discernment of entities that are constructed in scientific practices (Markus & Borsboom, 2013, p. 11; Messick, 1989). The discussion in Markus and Borsboom (2013) on realism begins with acknowledgement of the possibility of independent existence of the kinds of psychological attributes that are of interest to psychology researchers, including, for example, anxiety and extraversion. Here, we see a clear distinction made between the positing of an attribute with elements that are independently existing, and what is in part moulded by researchers in their creation of assessments that cohere particular terms and practices in order to trace the existence of the attribute via the latent variable in the mathematical model (p. 11; see also Markus, 2008).

The importance of the constructivist perspective on psychometric methodology is that it acknowledges that the formulation of a psychological construct contains the researcher's ideas about what the phenomenon is, and what sort of items will represent the phenomenon, as a set that can be used in a research project (Messick, 1981, 1989; Wilson, 2013a). Messick also notes that the constructivist-realist perspective does not entail commitment to any real trait or entity occurring in all people in all places (Messick, 1981, p. 583). His later work acknowledges a broad array of possible philosophical or ontological perspectives on psychological constructs, including realist, instrumentalist, and hybrid positions (Messick, 1989; Slaney, 2017). Messick (1989) does acknowledge that the philosophical standpoint of the individual will have distinct implications for research programs (Slaney, 2017). More is said about this below. A conceptual framework thus must maintain space for declaration of relevant standpoints and further needs to be flexible enough to support pluralism within a single project, or when combining or comparing different projects. Also discussed in Markus and Borsboom (2013) is a role for the GLVM in the process of abductive inference to the best explanation for psychological phenomena. Such an attitude to psychological research is exemplified in the work of Haig (2005), an author to whom we now turn.

2.2 Haig's naturalised methodological realism

2.2.1 Summary

In *Investigating the Psychological World,* Haig (2014) sets out a methodology for the conduct of psychological practices, adopting a realist perspective, and framing methodological realism as a core element of scientific realism that is described as the majority position in philosophy of science (p. 17). The basic commitments common across all versions of scientific realism as presented by Haig are that there is a real world, and that both observable and unobservable features of that world can be known via the scientific method (pp. 17-18). Haig (2014) claims for his proposals a naturalistic realism as a variant of scientific realism, drawing on Quine (1969). This position advocates for reversible relations between philosophy and science, such that philosophy constrains science, and science constrains philosophy. For Haig (2014, p. 19), the outcome is that the resulting methodology "gives us our best methods from which to choose and encourages us to constrain our theorizing in light of reliable scientific knowledge". In this regard, a conceptual framework for a research project may provide some means by which the researcher can declare their conceived relation between philosophy and science.

One other note of importance is that Haig (2014) distinguishes between local and global philosophies of science, proposing that what is needed for psychology is a local philosophy of science that "is realistic about the sciences to which it speaks" (p. 18). What this entails in Haig's view, is formulating realist perspectives consistent with the "particular natures and achievements" (p. 20) of the science in question, which for the purposes of this thesis is the field of psychology. Following Mäki (2005), Haig formulates five principles that together are relevant to any science. These principles include fallibilism, cognitive dependences between terms and practices, correspondence truth between phenomena and claims, object or entity status for observables and non-observables, and contextualising aims, for science. This view to find, support, and maintain local perspectives on practices that cohere with the best of developed knowledge in a field is consistent with several accounts across philosophical and mathematical literature that describe local realism, as discussed in Chapter 1 (see Henderson, 2018; Kincaid, 2000a). Local realism is exactly what the conceptual framework formulated in the next chapter and grounded in research situations aims to endorse, and is discussed in detail in section 5, below.

Haig (2014, pp. 14-15) notes a paucity of attention to methodology in scientific practice that pervades in present day research, and suggests that this may stem from ideals adopted in the mid-20th century regarding probabilistic theorising for inferences and the notion that one could begin gathering data and information in a theory-free manner, by just observing the natural world. It is noted here that one other possible reason for the lack of attention to the need for rigorous methodological practice is the absence of a standardising framework that allows for comparison of different approaches and different research outcomes, such as is proposed in this thesis. For Haig (2014), adequate methodology sets out a description of methods which are appropriate to the research project, facilitates evaluation of them against each other, and provides advice on which would be most suitable, according to the constraints considered. Such an approach to methodology is supported when systematisation of application of methods in research projects through time is supported in a consistent and logical manner. The conceptual framework developed in this thesis aims to provide such a structure.

2.2.2 Recent accounts from Haig (and Borsboom)

In a 2012 paper co-authored by Haig and Borsboom, we see explication of a rationale for a role for truth and theories of truth in psychology research. Specifically, correspondence truth is construed as mapping the relation or correspondence between a statement and reality (Haig & Borsboom, 2012; Haig, 2014, p. 12). The importance of a correspondence concept of truth for psychology research for Haig and Borsboom as distinct from coherence, pragmatist, or deflationary accounts is that it demarcates something that is different from falsity, regarding propositions that address states of affairs in the world. Such a role for truth is described as facilitating a factual interpretation of real-world phenomena, which itself is described as supporting scrutiny of scientific assumptions. This is because falsifiable circumstances are generated in the statement of a truth claim (p. 279). What this means is that in making a claim for truth, the opposite of it, its falsehood, is rendered directly available for scrutiny against the real state of affairs, in Haig and Borsboom's account.

This position is not far removed from the situational realist perspective that is further described in the last section of this present chapter. Situational realism takes the infinite complexity of any real situation as logically constrained under conditions of discourse (Petocz & Mackay, 2013). Logic and the situation coincide, in this perspective. In such a scenario, falsity is the direct opposite of truth, following the principles of the laws of thought described in Chapter 2.

What can be imagined as necessary for the correspondence theory of Haig and Borsboom is a detailed and systematic account of the context or situation within which the research takes place. This is because correspondence can only be assessed in the context of the conduct of the research project that makes use of the GLVM. As noted by Haig and Borsboom (2012), the mapping relation for correspondence between reality and truth is not uniform across different aspects of psychology research projects. This is particularly true for the latent variable in the GLVM, which, as stated in earlier chapters, is unavailable for direct scrutiny in a state of affairs. Because of this lack of availability, this thesis explores the potential of coherence accounts in scientific practices, to examine in what way such an approach may play an important role in justifying knowledge claims that follow from the use of the GLVM, in psychological research.

Original versions of coherence theory made use of the idea that truth was a property of a consistent set of propositions, where consistent means without contradiction (Benjamin, 1962). Consistency for sets is further described in detail in the section on set theory below, and the relevance of this consistency for the conceptual framework is pursued in Chapter 6. It is possible to note at this point that because latent variables are precisely latent, it remains challenging to make a firm claim regarding the correspondence of a latent variable with a real-world state of affairs. Further, the GLVM alone cannot easily be said to support existence claims regarding the latent variable, given the nature of the problems discussed for the model in Chapter 4. The key concerns of conditional independence, factor indeterminacy, and circularity render any claim regarding correspondence truth garnered through latent variable modelling practices open to question. An aim of the conceptual framework proposed in this thesis is that it can be used to demonstrate systematic links between researcher decisions and processes. A demonstrably systematic conceptual framework that spans the whole of a research project stands to provide exactly the information needed to justify a knowledge claim where the research includes use of the GLVM.

2.3 Comparison and summary

In considering the implications of these distinct accounts of realism for what would be presented in a conceptual framework for GLVM analysis in a research project, it is noted that Haig (2014) advocates for a role for exploratory factor analysis in his abductive theory of method for psychological phenomena under local realism, following Mäki (2005). Mäki (2005) approaches the idea of realism for science by fixing some aspects of realism for a specific scientific project, then adjusting the account of realism, as the project continues. To do this well, the scientific community is best supported by clarification of the researcher's original set of commitments in the context of the research situation which occurs in space and time, as well as any other relevant aspects for the adopted realist perspective in a conceptual framework.

Borsboom's (2005) account of scientific realism, on the other hand, maintains a distinction between theory and entity realism, and suggests that for latent variable modelling that theory realism be dispensed but entity realism be retained. This represents significant commitments about excluding theory from formulations of the psychological construct represented in the latent variable, representing a distinct set of constraints from those adopted under Haig (2014). While both Haig and Borsboom declare realist positions, each form of realism has distinct implications for the constraints that would apply to a researcher's project. For Haig (2014) it appears that theoretical adjustment may take place throughout the project, maintaining realism for all elements; for Borsboom (2005) however, theory should be excluded specifically from a realist formulation of the variable structure. The different realisms do not have the same implications for how the final latent variable model results should be conceptualised. Future researchers will need to be able to understand exactly how the philosophical commitments made by the present researcher impact on the formulation of not only variables, but all elements of a research project. It is completely possible to have different philosophical commitments in the articulation of theory, versus, models, versus variables. The conceptual framework aims to support transparency, where researchers adopt such stances.

Researchers of course may choose to adopt perspectives consistent with Haig (2014), Borsboom (2005), or possibly indeed, some other perhaps anti-realist framework. Because all research occurs in geo-historically situated contexts, this thesis proposes that all projects exhibit minimally, a realism consistent with the situation of research. In the next section, distinct formulations of realism are investigated to arrive at a perspective on the kind of philosophical and methodological positioning which is commensurate with understanding the centrality of the geo-historical situation, as contextualising, research claims.

3. Realism – thin, local, situational

To consider realism as relevant for geo-historical situations of research projects is one thing; what is also needed is a realism adequate to the construction of the GLVM and analysis of its practices. An important question at this point may be, how are the notions of a real situation that contextualises the research project, and the construction of the GLVM as a representational model, able to be integrated in coherent ways commensurate with a conceptual framework in which research protocols may be recorded? The answer to this question lies in a philosophical perspective that facilitates integration of both. In this thesis, all psychology research is supposed, in at least the first instance, as realist, to the extent that research occurs in real geo-historical situations. These situations are taken as constituting the domain or universe in which the research takes place. The GLVM is specifically, a constructed model. It is further proposed, that we can be nothing less than realist constructivists, in our use of the GLVM, in psychological research.

As noted above, in the existing psychometric literature, realism is endorsed as a stance relevant for the users of the GLVM. Borsboom (2005) advocated for a scientific realism with ontological import ascribed to a latent variable, which in his view psychology researchers treat as real, in practice. In section 2.1.1 above, this stance towards the practical use of the model by researchers was noted as falling short of what is needed when conceptualising the use of the GLVM in the whole of a research project. A broader ontology must be set for researchers, to be able to logically test and account for the phenomena behind the latent variable, in a research situation. It was noted that Borsboom later seems to abandon scientific realism for a constructivist realism, in Markus and Borsboom (2013). Constructivist realism is compatible with the position adopted in this thesis, with one important distinction – realism is given primacy over constructivism, for our purposes here. The ordering is taken to be important. Logically, what is needed is realism first, in respect of occurrences in the real world that feature invariance that can be tracked in correlational patterns. The real world situation of the research project is recognised prior to representing relationships between the real world circumstances and the representations of the situation, the phenomena, the variables, the models, and the theory that have roles in connecting the real world circumstances to the analysis conducted via use of the GLVM.

Haig's (2014) naturalised methodological realism is also consistent with the perspective endorsed in this thesis. An important aspect of this is realism for methodology is the recognition of constraints that are provided by reliable scientific knowledge. While Haig and Borsboom together go on to endorse a correspondence theory for psychometric practices with recognition of the importance of the role of truth in statements regarding states of affairs, this thesis takes a step that can perhaps be understood as a step further. It is proposed that the intrinsically constructivist process of representing psychological phenomena in the GLVM relies on practices consistent with a coherence account for psychological phenomena. Correspondence theory and coherence theory work together. Correspondence is needed so that the truth or facts about the research situation provide the means to set a series of constraints, for knowledge. Coherence is needed in representations that are made in respect of the use of the GLVM in the research project, as the number of problems that are involved with the use of the GLVM as highlighted in Chapter 4 introduce uncertainty into the claims that follow from a research project.

As confirmed in Chapters 1 and 2, logical testing plays a key role for the conceptual framework presented in this thesis for use with the GLVM. This is in the role of a check for truth or falsity in the process of self-audit, on behalf of the researcher, and community audit, where others within the scientific community scrutinise the representations articulated within the conceptual framework. More is described about the management of quality uncertainty via quality control, in Chapters 6 and 7.

Logical analysis, distinct from testing, also plays an important role in the conduct of psychological research generally, as shown in the following example. Logical or conceptual analysis has been cited as an activity that is rightfully prior to any observational activities in a research project (Petocz & Newbery, 2010; Boag, 2011; Petocz & Mackay, 2013). This is because logical analysis of phenomena, concepts, constructs, models, or theories may reveal implicit assumptions, contradictions, or inconsistencies in the research project that should be resolved before beginning any observational analysis, let alone statistical analysis in a latent variable model. Logically, conceptual analysis in the structure of a conceptual framework "has the power to preclude observational inquiry, whereas observational analysis can never reveal that conceptual analysis is inappropriate" (Petocz & Newbery, 2010, p. 131). Petocz and Newbery note that given finite resources for research purposes, there is no point in undertaking research protocols that, for example, assume quantitative structure for psychological phenomena, where conceptual analysis may reveal no logical ground for quantitative structure for psychological phenomena.

Such attention to logical ordering is consistent with an overarching approach to psychological research grounded in what has been described as situational realism (see Petocz & Mackay, 2013). Developed from the empirical philosophy of Australian John Anderson (1893-1962), situational realism "defines psychology as the study of those organism-environment relations and interactions (dynamical systems) that involve the psychological categories cognition, motivation and emotion." (Petocz & Mackay, 2013, p. 216). These authors note that a typical psychology researcher is realist to the extent that they take psychological phenomena generally to be something amenable to objective, if fallible, scientific scrutiny.

Situational realism however provides more resources than this. The core features of situational realism include that reality remains as a set of infinitely complex situations, where situations are process-based spatio-temporal events that are contextually embedded (Petocz & Mackay, 2013). No relation is constitutive in a situational realist account, and relations must involve at least two terms, so relations cannot be reduced to either one of the terms alone. This means that, for example, that it cannot be the case for latent variable modelling that intelligence is created by the relation between the modelled construct and the questions asked in the test items, for that construct. Importantly though, too, concerning logic, situational realism holds that the conditions of discourse have a role in the general form of situations and the logical implications of situations and their relations. Situations themselves ground logic, not the laws of thought, or abstract logical systems.

Nevertheless, situational realism remains silent on the use of modelling techniques in scientific practices. In terms of integrating set theoretical structures in the representation of real relations for a research situation, as is explored in the next chapter, one variety of realism consistent with logical testing for sets without the metaphysical overlay of specific categories as is the case for situational realism is thin realism. This is described in Maddy's (2011) *Defending the Axioms: On the Philosophical Foundations of Set Theory.* Maddy's thin realism as it is presented in this book is formulated within philosophy of mathematics, and is not far removed in its ideals from the local realism of Mäki (2005) that is discussed by Haig (2014) in his characterisation of naturalistic methodological realism, and which is further addressed below. Maddy (2011) asks questions about realism for set theory, attempting to answer questions such as whether sets are just fictions that inhabit mathematician's minds. Maddy contrasts thin realism with robust realism in formulating her answer. A robust realism would have a prescribed metaphysics, replete with fully articulated relations between logic and metaphysics, against which claims about the truth or falsity for existence of sets could be judged (Maddy, 2011, p. 63). Maddy notes, however, that this kind of reconciliation does not conform to how mathematicians do their work (pp. 66-7). Rather, mathematicians take a problem of interest, and articulate sets of axioms relevant to the problem, and test outcomes for these axiom sets. The backdrop of the real world and what science is taken to tell us about the real world, is taken as a given, in these mathematical practices, and is functional to the performance, of these mathematical practices. This distinguishes the thin realist account from the notion of a purely linguistic framework as was the target of Carnap (1950, see Maddy, 2011, p. 67-8), where all that matters are the within-framework relations. Thin realism advocates for a position of just-enough realism, in order that the method of the practice may be rigorously pursued to what she calls claims of "mathematical depth" (p. 60). The concept of mathematical depth is however noted as under defined by reviewers (Roland, 2014; Kennedy, 2014), although there is some reference in Maddy (2011) to fruitfulness and usefulness of set-theoretical analysis when she discusses mathematical depth. One problem with reference to claims of usefulness is relativism insofar as, what is useful for one person or in one scientific context may not be useful in some other context or in reference to a different philosophical perspective. (see Thomas, 2014; Roland, 2014). Thin realism by itself then does not provide all of the resources needed for a realist account of the research situation relevant to the GLVM. This is insofar as we do not seem to be better placed than any of the other realisms considered so far, when it comes to being able to say something about reality and its true relation, or otherwise, to a latent variable.

On the other hand, consonant with many of the intuitions that are present in Maddy's (2011) thin realism is the perspective of local realism as endorsed by Mäki

(2005), in the philosophy of another social science, economics. Local realism accepts the different practices that fall under different domains of science, and it attends to the distinct ways that realist perspectives may be formulated for the different sciences, noting simultaneously that there are some aspects of what is real “that are not up for grabs” (Mäki, 2005, p. 235). This comports with Maddy’s (2011) approach of accepting the best of what other sciences say as true, in carrying on work in a specific field. Maddy’s (2011, p. 71) thin realism serves primarily in an epistemological role, however, while Mäki (2005, p. 238) conceptualises primarily an ontological role for scientific local realism. Mäki argues here that a position on realism does not have to make direct claim about actual existence of entities. For Mäki, it is enough that entities *might* exist that explain phenomena, for researchers to be able to be realist about the circumstances and events that involve them. Fallibilism has a role in this local realism, as the notions about existence for a specific entity may be proven not to hold in all circumstances. Mäki (2005, p. 239) leaves open the possibility for determining true existence for an entity via stronger realist claims.

Mäki (2005) poses a specific kind of realist positioning in attending to the nature of social sciences phenomena. He notes that it is less likely that the entities of potential interest to social scientists such as psychologists are mind-independent. This is specifically because as seen below in the set-theoretical distinctions of Markus (2008) between concepts and constructs, researcher intuitions have a vital role to play in the existence of the evidence that provides insights into the nature of the phenomena of interest. Mäki (2005, p. 246) proposes that it is not mind-independence that is of greatest significance, but rather, science-independence. What is needed here is a position that recognises the distinctions between science-world relations, mind-world relations, and phenomenon-world relations. With such information, a researcher can make a justified choice about realism or antirealist attitudes towards the phenomenon in question. Local realism facilitates awareness of the distinctions between these.

The shift beyond the position of Mäki (2005) proposed in the local realism relevant to this thesis is that the geo-historical situatedness of the research project is taken as constituting real and foundational science-world relations. These are the departure point for any further claims made in the conduct of the research project, and specifically for our purposes, projects which involve use of the GLVM. Local realism grounded in an understanding of the geo-historical context of the research situation as the initial domain of interest in a conceptual framework has a further advantage. This is insofar as it will provide the ground for a risk assessment conducted by the author in a process of self-audit, and by fellow or future researchers who have an interest in the outputs of the research work. To describe more about risk assessment and its role in reducing quality uncertainty, what is needed is explication of the conceptual framework. This follows in the next chapter, drawing on the set-theoretical hierarchy of models set out in Suppes's (2002) informal view of science, described in detail in the next chapter. In the next section, an introduction to set theory is presented and an application of set theory that supports distinction between psychological phenomena, concepts, constructs, and latent variables building on the work of Markus (2008) immediately follows.

4. Set theory for the GLVM

As a broad characterisation of an initial position on set theory, it is of interest for the formulation of the conceptual framework presented in this thesis in several ways:

1) It stands to unify mathematical, statistical and psychology theory goals in one framework suitable for scientific study, as clarified in Chapter 1[16];
2) The history of development of set theory grafts a trajectory that can be followed to secure the rigorous development of a concept, as needed for projects making use of the GLVM, following the problems associated with the conditional independence assumption, circularity, and factor indeterminacy, as seen in chapter 4;
3) The philosophy of set theory invites reflection on the notion of psychological phenomena as empirical *events* or occurrences (see Suppes, 2011), which are analysed using methods that rely on a non-empirical concept of *infinity.* Given the tension between the empirical event and non-tangible infinity, rationale is founded for the reporting of our research projects as a series of *decisions*, where it is not yet the case that clear frameworks exist for the latter yet in psychology research and psychometric practice, methodologically speaking;
4) In the model-based systematic conceptual framework, sets provide a founding form for the relational structure that is adopted for research elements, described in detail in the work of Suppes in Chapter 6.

Careful inspection of psychometric literature reveals that a good deal of analysis already conducted in the field is conducted using set-theoretical terms (see Markus, 2008; Wilson, 2013b; Maraun, 2017). The language of set theory has crept into psychometrics as a field, consistent with the adoption of set theoretical formulation in the fields of applied mathematics (see Maddy, 2011). What has not yet appeared in the psychometric literature is a detailed account of how set theory may be directly relevant to and applicable in

[16] A key feature of mathematical set theory as described in Chapter 1 is the flexibility with which it can be used to describe theory for the the whole of mathematics (see Grattan-Guinness, 2000).

psychology research. This thesis aims to address such an account.

4.1 Consistency as criteria for set membership - ontology

Sets, as originally conceptualised in Cantor's work and further described in detail below, are defined entirely by the members or elements that make up the set (Stoll, 1961, p. 2). The simple relation of belonging to the set is what makes a set a set, and there is no structure outside of the elements of the set, themselves. In an example relevant to what is described above regarding the importance of clearly defining a research situation, this means that definition of a situation in set-theoretical terms is possible by defining the set of elements that belong, to the situation. These elements are taken as consistent, to the extent that they belong together, as a set.

A key feature of set-theoretical practice is the subtle distinction between belonging and inclusion, for elements of a set. Much use is made in later development of axioms in set theory of the idea of distinction between the simple relation of belonging together, and the notion of being included under some pre-existing set (see Fraenkel, Bar-Hillel, & Levy, 1973; Bourbaki, 1968). Such a distinction is introduced in the history of development of set theory in order to help resolve paradoxes that are described in detail below. Another important aspect of the development of set theory are tests for logical consistency. Grattan-Guinness (2000, p. 120) notes that it is most likely the case that Cantor intended his mathematics to meet at least the logical requirement of consistency. Logical consistency is obtained when there is no contradiction in the propositions that make up a claim or argument (Tarski, 1946). Cantor explicitly states for example, in respect of a well-defined set:

> On the ground of its definition and as a result of the law of the excluded third, it must be seen as internally determined as whether any object belonging to any same sphere of concept belongs to the considered

manifold, or not (Cantor, 1932/1882, p. 150).

In this thesis, consistency is taken to be an attribute of the set of elements that describe the research situation, which belong together as an account or description of the research situation. Consistency is assumed as a beginning point for all research elements that are described in more detail as model-based relational structures in Chapter 6. The criteria of consistency can be understood as a primary or initial constraint of any situation, under the conditions of discourse. What this means is that a real geo-historical situation is also, for the purposes of representation in the conceptual framework for the use of the GLVM, a consistent one. Initial consistency aligns the circumstances of the situation with the articulation of its representation, but also makes room for the way that thought and reality can come apart. This would occur for example where some logical test of truth or falsity demonstrated some inconsistency in virtue of a set of premises or claims. What that means is that when a researcher creates an ordered tuple via the method described in Chapter 6, which is a finite list of ordered sub-elements to represent a particular research project element in a model-based representation, the representation itself operates as a constraint that renders the terms of the element as consistent, for that element. When an element for a research project is represented in set-theoretical terms, each of the components of the element should be able to be understood as existing, without contradicting each other, in respect of that element. When something is taken to consist as part of a set that defines an element, it is understood to fall within a situation, as a countable and ordered component, of that situation (see Jacquette, 2014).

Further to the topic of consistency for representations, psychology research that utilises statistical analysis in the GLVM is likely to deal in uninstantiated properties. This would be the case when, for example, a range for a continuous latent variable is employed in parameter solutions, and there remain some values within that range where for which no relation can be said to exist to values for manifest variables. This would occur when the data set does not contain these values for the manifest variables. Uninstantiated properties not only lack ‘thingness’, or entity status, they also lack foundation for real-world ontology, in terms of the notion of correspondence truth for the representation, and some real-world existent property (Franklin, 2015). It is of course challenging to claim correspondence truth, for an uninstantiated property. Yet, the uninstantiated property as an uninstantiated value may have a vital role for the GLVM, in terms for example of being able to perform statistical and probabilistic analyses that make use of latent or manifest variable value ranges which could be applied over a specific data set. In this value range may be values that are not represented in the empirical data collected for a variable, but these variable values still have a role in identifying meaningful data values, for our data set.

Another important but slightly different uninstantiated property relevant for the GLVM is connected to the role of infinity (Franklin, 2015) in a Gaussian or normal distribution. Normal distribution makes use of the central limit theorem, and the notion that an infinite series of events or infinite population would produce a certain shape, to the data frequencies. Here it is not so much that the uninstantiated property takes form as a specific data value, but rather, that it has a role in generating data distributions that have certain frequency assumptions built into them. At the same time, uninstantiated properties in either the missing data role or the Gaussian distribution form play important roles in GLVM processes. Their correspondence with a state of affairs in reality has no truth status, meaning they have no value under a correspondence theory of truth. Where the uninstantiated properties are formulated as belonging to a set, coherence theory facilitates assessment of consistency of the set and any related propositions. Taking set theoretical formulation into the structure of the conceptual framework for the GLVM provides the means to record researcher commitments regarding these uninstantiated properties, in a way that supports the assessment of meaningfulness for an element and for the research project in terms of quality uncertainty, described in detail in Chapter 6.

There are a series of postulates, or axioms, typically adopted for applications of set theory, but no constants associated with the theory of sets itself (Bourbaki, 1968). It is only the belonging relation that characterises a set as a set. Axioms are statements taken to be self-evident, while postulates obtain truth in virtue of the consequences derived from them (see Kline, 1972, p. 52). Postulates were proposed for set theory to resolve problems associated with the application of set theory to real-world circumstances, or to resolve logical paradoxes, and thus maintain consistency. These are explored below. Before this, to learn something of how these principles come into being, next we turn to Cantor's original development of sets. Cantor revolutionises the concept of infinity (Dauben, 1992), and subsequent work by other mathematicians and logicians addresses the occurrences of antinomies (paradoxes) in a way that leads to the explication of axioms by later researchers, to maintain the consistency of set-theory and overcome the paradoxes (see Maddy, 2011; Ferreiros, 2016; Grattan-Guinness, 2000). The history is of interest to the degree that the steps involved in producing solutions for intractable dilemmas resembles in some ways the problems for the GLVM of factor indeterminacy, conditional independence, and circularity.

4.2 Cantor's innovation

Georg Cantor (1845-1918) was a German mathematician, working roughly at the time of the development of latent variable modelling, just prior to the turn of the 20th century. In a series of research efforts directed to trigonometric series, Cantor recognised a need to distinguish between different orders or sizes, of infinity (Stoll, 1961; Dauben, 1990). This included a distinction between the sizes of infinite and finite sets of numbers. What Cantor saw in connection to his work with infinitesimals was that he could use a technique of testing matching numbers in one-to-one relationships (Dauben, 1990). Such a technique for Cantor involved an attempted one-to-one relational matching of sets of numbers, so that two sets having the same power defined the situation where all the members of one set could be put into one-to-one relations with all the members of another set.

In 1883, Cantor publishes his *Foundations of a General Theory of Manifolds*, which extended the concept of the infinite beyond the interpretation which had existed in mathematics up until this point, which was as a limit to the finite numbers. What Cantor found and proved with his methods was that there were many more real numbers than there were natural numbers (Cantor, 1883; Dauben, 1990; Grattan-Guinness, 1978). These sets of numbers could not be put into one-to-one relations with each other, proving that there were different orders of infinity, relevant to each (Grattan-Guinness, 1978; Stoll, 1961; Dauben, 1990). These transfinite numbers had identifiable and determinate number-theoretic properties and meant that mathematics could no longer simply be conducted with the perspective that an infinity coincided with a single limit (see for example, Gauss, 1860). There were, from Cantor's results, many and successive, infinities, in which numbers may behave differently (Dauben, 1990). A further distinction between these actual infinities and potential infinity allowed Cantor to distinguish between consistent and inconsistent multiplicities, or sets.

As mathematicians applied Cantor's set theory, paradoxes or contradictions became discernible, regarding sets. For example, Russell's (1908) paradox addresses the problem of the specification of sets where all elements are sets that do not contain themselves[17]. This statement produces a contradiction. For a set that is not an element of itself, the definition states that it must contain itself. If it contains itself, it contradicts the definition

---

[17] Note that Zermelo (1908) lays claim to a simultaneous but independent discovery of the paradox. Typically when accounted for in the history of mathematics, it is cited as "Russell's paradox or Russell's antinomy" (see for example van Heijenoort (1967).

for it, as the set of all sets that are not members of themselves. Zermelo's (1908) response arguably proposes the most famous resolution for Russell's paradox, which is founded in an axiomatic technique. In the solution, the concept of a power set and the inclusion relation is utilised, distinct from the belonging relation, to nominate the elements that are included under a set. This axiomatic technique has a number of equivalent formulations (see for example Russell, 1908), and was subsequently further developed by Fraenkel and Skolem (see Fraenkel et al., 1973), into what is presently called the ZFC axiom set. The formulation under ZFC approaches typically make use of weak existence axioms, and include introduction then of axioms such as that for the power set. This would specify what is included in a set as a set, and for the empty set, as included in all sets whether there are other elements to the set, or not (Fraenkel et al., 1973).

Importantly, when considering the relevance of axiomatic set theory for the GLVM, the ZFC system does not assume that for every property that a set of all entities or things satisfying that property exists. More, it asserts *definability* - given any set, a subset of this set is definable, in logical, consistent, format. Most set theoretical approaches work with principles of iteration that take their form from this principle of definability, working in first-order and second-order logic to eliminate circularity, and to create different structures, such as a cumulative definable hierarchy, where lower-order elements inherit all properties from higher-order elements (see Stoll, 1961).

Set theory then has enough structure to facilitate definitions of research elements in the conceptual framework, but makes no commitments about existence for properties or objects, or even sets (Chihara, 1990; Maddy, 2011, p. 52). A realist commitment about the geo-historical situation of the research project is taken as a grounding of existence statements for use of the GLVM in an applied context, throughout this thesis. What this means is that consistency for research elements in model-based relational structures is assessed with reference in the first instance to the geo-historical situation in which the research occurs. The mechanics of this are described in Chapter 6. In the next part of the present chapter a brief example of a set-theoretical approach to the clarification of the distinction between research phenomena, concepts, constructs, and latent variables, based on principles adapted from Markus (2008) is presented.

5. Set theory, conceptual and construct clarification in Markus (2008)

In *Constructs, Concepts and the Worlds of Possibility, Connecting the Measurement, Manipulation, and Meaning of Variables,* Markus (2008) distinguishes meaningful differences between psychological concepts, and research constructs using set-theoretical formulation. He is explicit in employing set-theoretical structures as “building blocks” (p. 62) without making any philosophical commitments regarding the existence of sets (see Chihara, 1990). One key feature of the set theoretical distinction that Markus does draw on is the difference between an extensional and intensional definition for sets. Extensional definitions apply the relation of belonging to a set over all actual cases or real objects or properties. Intensional definitions on the other hand apply the relation of belonging over all possible objects, or properties.

The distinction provides guidance on the difference between a concept and a construct. For a research construct, Markus follows Cronbach and Meehl (1955) to state that the meaning of a construct is formulated in a nomological network of variables. In this way, individual differences on the construct come about because of individual differences in variate values for the variables that make up the construct. A construct for depression for example may be made up of variables for mood, interpersonal interaction level, and activities of daily living, on each of which individuals may exhibit different levels of scores in a self-report questionnaire. The interaction of mood, interpersonal interaction, and activity level together is assessed, when the construct labelled “depression” is assessed.

One important aspect of a construct for Markus (2008) following Cronbach and Meehl (1955) is that a construct is defined relative to the pattern of covariation in a specific population – for Cronbach and Meehl, this was in the relations established in a nomological network, for a specific term. Markus (2008) gives constructs then the structure of a set with an ordered pair, so that for each member of the population there is a specific matching construct score. The relation is a homomorphic one such that for a construct depression labelled as D, we have a set of all ordered pairs <x, y>, where there may be many x's for a single given y, such that each person has one depression level, as depression is defined for that population.

Concepts on the other hand as defined by Markus (2008) extend across both actual and possible real-world cases. The set looks the same as an ordered relation, but the x term applies over a different domain, or series of values for the first expression y in the ordered pair, which includes actual and possible individuals. An important aspect of this is that it allows for reasoning to be extended for the research concept over hypothetical individuals and situations, for uninstantiated properties, as discussed in section 4, above. For a construct however a change in structural relations such as new emerging values of y for some part of the population x involves an entirely new construct, following Cronbach and Meehl (1955), when the nomological network is interpreted in a strict way. In terms of assessment of construct validity for the GLVM, benefit may be obtained by testing the matching of ordered pairs <x,y> on different but similar-enough constructs, defined over the same population group. The testability of constructs is improved, by maintaining this distinction, between concepts and constructs, and by defining clearly the population domain over which the construct applies. In as much as the domain for the test items was shown to be vital for reasoning about GLVM analysis outcomes in Chapter 4, so too here does clarification of the population domain serve to support coherent reasoning about the psychological construct and phenomenon of interest, following GLVM analysis.

An example from the patient-centred research literature presented in further detail in Chapter 8 and which was discussed in examples in Chapter 4 demonstrates the usefulness of distinguishing between concepts and constructs with respect to use of the GLVM. The Mood and Anxiety Symptom Questionnaire (MASQ: Clark & Watson, 1991) was developed originally as a 5-level Likert scale self-report 90-item assessment which aimed to measure a tripartite model of depression and anxiety. One striking aspect of the MASQ is the frequency with which replications of the initial results have been attempted; across a broad array of participant groups, with distinct characteristics. For some of these replications, a replication of the tripartite model has been secured, although not necessarily for all factors. For other replications, no indication of the same factor structure has emerged. There are aspects of methodology that may have some bearing on the differences in outcomes observed across the different studies. For example, Lin et al. (2014) attempted a replication of the tripartite structure for the 90-item MASQ on a small sample of 147 help-seeking young people between 15-24 years old in Australia, and report confirmation of the three-factor model (X^2 = 193.5, df = 83, CFI = .92, RMSEA = .10). Boschen and Oei (2006) had attempted a replication with 470 outpatients at a university psychology clinic in Australia, aged between 17-70 years. These authors report that their attempt at replication of the empirical tripartite model was not successful (X^2 = 9042.55, df = 2697, CFI = .72, RMSEA = .07). Note that neither study just listed realises the goodness of fit criterion described by Hu and Bentler (1999) described in Chapter 4.

It should be noted that one of the reasons for this difference in outcomes from the original studies of Clark and Watson (1991) may have to do with applications of different analytical methods, across the studies. Lin et al. (2014) for example construct their analyses using weighted least squares (WLSMV) and polychoric correlations, citing their reasoning as the model indicators (presumably the subscale totals) being categorical. This appears to be different from the validation studies performed by Watson et al. (1995a, 1995b) which used principal components analysis and Pearson's correlations to construct analyses across groups which included university student, community adult, and substance use clinical participants. Interestingly, Watson et al. (1995b) generate one and two factor models as well as the three factor model, and state that these models replicate across groups. Replication of the two and three factor model variations are attempted by Boschen and Oei (2006) without successful outcomes, from their report. These authors also attempt item level analyses of the MASQ tool. Both this item-level analysis by Boschen and Oei (2006), and the WLSMV analysis of Lin et al. 2014) are rightfully described as analyses of different constructs from the analysis of Watson et al. (1995b). This is in part because different calculation methods are used, and in part because different population characteristics can be used to describe the participant samples. In this scenario, to maintain integrity to what was originally described as the tripartite model, the tripartite model could beneficially be treated as a concept, rather than a construct. This approach to the tripartite model would mean that a different construct is implied for each different type of population group, model and method combination that is analysed.

Specification of the research situation as individual domains further provides a grounding point for testing correspondence claims. Domains are noted by Borsboom (2005, p. 47) as foundational to assessment of the kind of variation that is of interest to psychological researchers. Variation over the domain of time, the domain of individuals, or the domain of a specific situation each cannot be directly generalised to any of the other domains. Set theory provides the structural apparatus most amenable to clear specification of domains and does it in a way that facilitates clarification of networks of relationships in inferential processes. In the next chapter, the full picture of the conceptual framework is described in detail, and the import of linkages between domains in representing a research project speaks, for itself.

6. Summary

In this section, distinct realist perspectives on psychology research methods have been explored with attention to psychometric practices, looking at entity realism, naturalised methodological realism, and constructivist realist perspectives. It was noted that these typically offer a metatheoretical lens on research practices, and that a researcher may be interested in adopting any of them and perhaps some other perspectives, in their research project. Philosophical perspectives have been demonstrated to have consequences for what is of import in a research project, influencing the principles adopted, assumptions checked, and practices undertaken. The GLVM has any kind of interpretation open to it; what is less important it seems than offering a determinant view on the philosophy behind the GLVM is offering to the present-day researcher a way to integrate their qualitative commitments with their quantitative or empirical, practices. A realist position situates the research project in a geo-historically determined world, setting ground that operates as an initial constraint regarding what occurs, in GLVM analysis, and what is said following GLVM analysis. In the next chapter the conceptual framework structure is introduced, building on principles of, representation, and invariance as exemplified in Suppes (2002), and grown from a realist founding in the situation, for the whole of the research project.

Chapter 6: Conceptual framework - defined

1. Objectives

In this chapter, the conceptual framework for the use of the GLVM in psychological research is formulated and presented. Key considerations that the conceptual framework must address were developed in the previous chapters. Chapters 2-4 traversed the historical development of the GLVM, including an overview of the lack of completed proof for elements of the model. Different formulations and definitions of the GLVM have been reviewed from McDonald (1999) and Bartholomew et al. (2011), and it has been shown that each of these involve distinct sets of decisions and practices, for use of the model. A series of common problems and decisions related to application of the GLVM in research situations was presented in Chapter 4, with specific attention to the key concerns of conditional independence, factor indeterminacy, and circularity. These three key concerns demand attention from researchers, in terms of deliberation and resolution on choices about how to overcome the gap between the construct that represents the psychological phenomena, and the quantitative model. In Chapter 5, realism for GLVM practices was explored. Chapter 5 demonstrated that diverse philosophical perspectives on the GLVM, even ones similarly named as realist from Borsboom (2005) and Haig (2014) have diverse consequences for statistical analysis and for the reporting of research findings. Distinct sets of constraints, working as interactions between the assumptions necessary to use the GLVM and philosophical principles follow from adoption of distinct philosophical points of view. The last part of Chapter 5 pieced together an argument for a local realist account of the research situation. Local realism does not attempt to set an account of metaphysics for the whole of reality, rather, it adopts the best of what science tells the researcher about the nature of reality in their specific geo-historical moment. It attends then to both the circumstances and practices within a scientific paradigm to formulate an understanding of the kinds of constraints applied in that paradigm in developing its specific relationships between truth and knowledge. This philosophical perspective, grounded in the geo-historical context of a research situation, must be understood as realist constructivist to the degree that we acknowledge we are constructing models, when we use the GLVM in research. More about the constructivist part of realist constructivism will be discussed in Chapter 7.

1.1 Key role of the research context as a domain

Chapter 5 also demonstrated the way that set theory can be employed to distinguish concepts, and constructs, following the example set by Markus (2008). It has been noted that set-theoretical formulation facilitates clarification of the domain or universe over which a concept or construct is defined. The domain marks something invariant about the research project, in which researchers have an interest. A domain or universe that has already been marked as indispensable for the conceptual framework for the GLVM is the geo-historical situation of the research project. This originating domain for a research project provides a first framing for correspondence between the representations made regarding the phenomena of interest, and the invariances that are of scientific or research interest that are tracked in the models used in psychometric practices.

Formal acknowledgement of research environments is a typical recommendation in publications endorsing research integrity in science generally (see National Research Council, 2002; OECD, 2010). The definitions of research environment in this literature extend beyond the parameters of the geo-historical situation already discussed here as the realist context for the GLVM, to extend to a broader domain which may include for example other institutions that may be running similar research projects (National Research Council, 2002) . The reasons for clarifying the research environment from the perspective of scientific integrity align with the reasons for disclosure of the research situation, in that such a process facilitates a reconciliation for the choices made by researchers in their research practices to a set of circumstances that are relevant to those choices (see National Research Council, 2002; Mayer & Steneck, 2012). In defining a research situation for the use of the GLVM, as in the research integrity literature about environments, it is worth considering the recording of information such as the political, institutional, and geographical circumstances, of the research (National Research Council, 2002). Where time-stamping is becoming a recognised imprimatur in terms of fulfilment of pre-registration processes for psychology research (Nelson, Simmons, and Simonsohn, 2017; Davis et al. 2018), a time-stamp for the research context also serves to give a beginning point for the declaration of a priori commitments for the project in question.

In what format, then, should this information about the geo-historical situation as a research domain be recorded, as a representation in a conceptual framework? In a 2002 book named *Representation and Invariance of Scientific Structures*, Suppes summarises his body of work across several empirical sciences including physics and cognitive science, to present set-theoretically informed approaches to formalising scientific knowledge about empirical situations in models. This has been paradoxically labelled as Suppes's "informal-structural view" (Muller, 2011, p. 93). Suppes's approach to formalising scientific processes in hierarchical model form will be further elaborated in Section 3, below. Suppes is perhaps best known for his formalisation in measurement theory (see Vessonen, 2017; Boumans, 2016), and some of the examples considered below touch on his measurement work. Before moving to the informal structural view of Suppes, given what we have seen in Chapter 4 and 5, it may be helpful to address questions such as: in what way can the GLVM itself be understood as a measurement technique?

2. Measurement and the GLVM

In the previous chapter, it was noted that Borsboom (2005) describes the GLVM as a measurement model. Claims that the GLVM serves as a measurement tool are pervasive generally in the literature from the time of Spearman. Amongst early theorists, Burt (1940) took factor analysis as a technique of diagnostic importance for "measuring" traits (see p. 115). Thomson (1939) does not mention measurement in connection with factor analysis but does focus on the status of 'metrics' or scores for the manifest variables (see p. 329). More recently Kline (1998, p. 65) states that both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) function as measurement models. The term measurement is also maintained in the Structural Equation Modelling (SEM) literature to distinguish a structural model, which describes hypothesised causal relationships, from the measurement model that simply describes covariances between manifest and latent variables (see Kline, 1998, p. 57; McDonald & Ho, 2002).

However, what can measurement mean, when the GLVM is founded on the assumption of conditional independence represented by correlations of other variables? The Joint Committee for Guides in Metrology (JCGM: BiPM, 2012) defines measurement across all sciences as the "process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity" (p. 16). Three notes follow the definition. The first states that measurement is not applicable to nominal properties, which would mean that latent class analysis under the GLVM, which deals with categories or nominal classes for the latent variable, is unlikely to be cast as a measurement activity. The second states that measurement implies the comparison of or counting of entities. The GLVM is not a technique of comparing or counting variables or phenomena behind the variables, as much as a link function that connects the latent and manifest variables, given a pattern in the data for the manifest variables (Skrondal & Rabe-Hesketh, 2004). From the perspective of measurement involving comparison or counting of entities, the GLVM process does not seem to fit, as a measurement process. The last note on measurement in the JCGM states measurement relies on a description of the quantity that aligns with the intended use of the measurement result. De Bievre (2012) comments that this last point is distinct from earlier versions of the definition of measurement, but also that a "consequence of this definition is that the analyst must think clearly about what (s)he is going to measure (and announce that in the beginning of any subsequent publication of the ensuing measurement result)" (p. 232).

There are two main points to note, in reference to these notes on measurement in the JCGM. One is that there are circumstances where practices that make use of the GLVM would not count as measurement practices. The second is that, to the extent a researcher wishes to pursue measurement, the standards recommend the inclusion of qualitative description, to support the quantitative analysis. In conclusion, to the extent that a researcher prefers to accumulate arguments for a measurement interpretation of the GLVM, the conceptual framework will need to provide a structure that can accommodate the specification needed to clarify researcher intentions regarding the measurement, as described in the JGCM (BiPM, 2012). Any preference for casting the GLVM in a role of measurement can be understood as involving several steps beyond use of the GLVM, itself. Some brief examples of how to account for these extra-statistical steps are set out in what follows, below.

In turning to the relation between factor indeterminacy and measurement, what is needed is an understanding of the act of measurement, and what it aims to address. Suppes and Zinnes have described the problem of measurement (1963, p. 7) in terms:

> of showing that any empirical relational system that purports to measure (by a simple number) a given property of the elements in the domain of the system is isomorphic (or possibly homomorphic) to an appropriately chosen numerical

system.

Factor indeterminacy, as discussed in Chapter 4, has implications for the status of the GLVM as a measurement theory, considering the above quote, in a realist context. It means that isomorphism, or the structures that preserve one-to-one relations between empirical outcomes and variable values that could otherwise be demonstrated as outcomes from statistical analysis suitable for measurement (Tal, 2012) lapse into many-to-many both knowable and unknowable structures (Mulaik, 2010). It is not possible to ascertain a unique relationship between the latent variable, and the construct in question, precisely because of what is entailed by the mathematical functions under conditional independence. Measurement as described by Suppes and Zines entails confirmation of both existence and uniqueness (see Bacelli, 2018). The latent variable, as discussed in Chapter 4, is not unique, because of factor indeterminacy (Mulaik, 2010). The latent variable is also latent, or unavailable for direct confirmation. In the interpretation of Cliff (1983), this means that its existence remains unconfirmable, because there can be little or no agreement about what exactly it is, that constitutes the latent variable. It is not available for objective, scientific scrutiny, regarding its nature.

Markus and Borsboom (2013a) give extended treatment to the question of measurement theory for psychometric practices in their book *Frontiers of Test Validity Theory.* These authors note that confirming uniqueness for an axiomatic approach to measurement remains challenging in the context of researching psychological phenomena. This is because violations of transitivity, a key feature of measurement of quantity, are frequent for psychological phenomena, given the complexity of variables and interactions that are involved for psychological instances. They further note that it is hard to track the cause of the violations, making it impossible to meet the precision required for axiomatic measurement specification. Markus and Borsboom (2013a, p. 40) suggest that the latent variable model offers an alternative to the axiomatic approach to measurement, because it offers a common cause interpretation of the latent variable. Chapters 4 and 5 have already described problems with a common cause interpretation of the GLVM. As previously stated, a causal interpretation of the GLVM relies on extra-statistical assumptions (these are noted by Markus & Borsboom, 2013a, p. 40). Yet Markus and Borsboom suggest that in virtue of relying on conditional independence for the GLVM within the context of causal accounts of the model, "the evaluation of measurement models is scientifically accessible by statistical means" (p. 40). Without grounding in the context of the research situation and clear qualitative framing of the extra-statistical principles assumed by the researcher for causal reasoning, causal reasoning about the GLVM as a measurement technique is subject to circularity of claims. This is a condition described as relevant to the GLVM in Chapter 4 of the present thesis.

In responding to concerns about the foundations for psychometric measurement practices which rely on the GLVM, McDonald (2003) follows Guttman (1953, 1955), Lord and Novick (1968), and Cronbach, Gleser, Nanda, and Rajaratnam (1972), to consider the consequences of setting of conceptual domains for psychological phenomena. Mathematical domains are defined as a set of elements on which a logical variable can be defined (Stewart, 1995). McDonald (2003) proposes that each item that makes up a battery or questionnaire in psychological research has its own domain, which is the set of possible responses that may be recorded as data for that item. A conceptual domain for each item, with a relation between the concept and some property value for the concept (e.g. level, presence or absence, correct or incorrect response) can be modelled for a sample of individuals in a population[18]. The question then becomes how do we rationalise a

[18] Note that manifest variables rarely are constituted as a single item or concept, item bundling, or batteries of items for a single construct as 'mathematical ability' feature frequently in GLVM processes. McDonald (2003) is followed for focus on the item domain, with more specificity and more limitation as he discusses (p. 213).

connection between a latent variable or common factor, and each of the item conceptual domains, when factor indeterminacy prevails for the GLVM? Factor indeterminacy implies that no single conceptual domain for the latent variable can be connected to the multiple item domains, for the manifest variables. The problem is in some ways compounded by the practices that psychometricians already engage in, to lengthen or shorten the number of items included in some evaluation of the concept that the latent variable represents. If a series of items, say, was originally conceptualised as assessing a definition of anxiety, how is it possible to add or subtract an item, and say that it is the same anxiety that is still being measured?

In methodological support of such practices, McDonald (2003) proposed setting an infinite behaviour domain, in order to articulate some position from where uniqueness or existence claims could be made for a specific concept. This yields a universal set of possible items that has countably infinite length. From this set, a finite number of items are realised in practice. For McDonald (2003, p. 219):

> The countably infinite set of item stems in the domain substantively gives a unique identity to that attribute, and a function of the scores in that set determines a measure of the attribute uniquely, as the quantity to be estimated from any finite subset.

McDonald (2003) notes that Guttman's (1955) solutions to the problem of factor indeterminacy asked researchers to choose maximally discriminant alternative mathematical solutions for model equations. These mathematical solutions however cannot be connected directly to responses of examinees on test items chosen from an infinite set of such. There remains a gap between the chosen mathematical model, and the psychological phenomenon or attribute under examination for which measurement is intended, which can only be closed by choice of a qualitatively specified rationale. McDonald (2003) explores two possible rationales. The first is the omitted cause conception of the latent variable. This expresses the idea that the latent variable would be directly measured as the common cause of all item responses, if researchers could just observe it. McDonald (2003) notes that there is nothing in the mathematical model that connects easily nor necessarily to a causal interpretation of the latent variable (c.f. Markus & Borsboom, 2013b). Most importantly however for McDonald (2003, p. 222) “investigators do not operate a common-cause notion”, in GLVM use.

McDonald (2003) goes on to state that a common *properties* conception of the latent variable makes much more sense to researcher practices, in this regard, following Thomson (1934). A common properties approach, distinct from a common cause approach, conceptualises the items behind the manifest variable variate values as containing some common element, for the items. While the detail of his arguments is not recounted here[19], the result is a recommendation that researchers account for their decisions to combine items using “marker variables” (McDonald, 2003, p. 223). These marker variables represent an a priori choice to connect the concept representing the phenomena of interest, to a series of items that can be tested as a psychological construct. McDonald (2003) concludes that this still does not give an account of psychometric activities that is likely to satisfy requirements for measurement, leaving his “unsatisfying conclusion” (p. 226) as a challenge for future researchers.

In conclusion, specifying an infinite behaviour domain sets a context of relevance for the set of items chosen by the researcher, such that a rationale is available for addition

[19] The value of the common properties approach specifically is that items can then be grouped for alternative item domains, with correlations examined, for the domains (see Thomson, 1934). The effect on the correlations of adding and subtracting different items from different domains can be compared, to ensure the most relevant items are included under each domain, and then, under each manifest variable McDonald (2003) in conclusion recommends banking such tests to create ultimately a large database of item functioning. The conceptual framework can support data banking, in a generic format suited to any study or situation.

of items under a common properties approach (c.f. Markus & Borsboom, 2013b). Everything turns on the choices that the researcher makes, and the trail of evidence they accrue between the phenomenon of interest, and the steps taken to represent something about the phenomenon in ways that make the information available to the community of practitioners interested in psychology research outcomes.

2.1 Relevance to the conceptual framework

One important aspect of examining the GLVM with respect to the question of measurement is noticing that research making use of the GLVM involves not only quantitative, but also qualitative elements. To set a domain for a concept, what is demanded is qualitative specification of the concept in question, prior to any assessment of quantitative properties for the same concept. With the problem of factor indeterminacy for the GLVM and the conditional independence assumption, under a realist interpretation of the research situation, qualitative and quantitative information is inextricably interwoven into any claims that follow from the completion of a project. The link or relation between the concept and any variate values for the latent variable can only be described in a qualitative statement. The conceptual framework to be presented here is a tool for holding these qualitative and quantitative elements together as a representation of the psychological phenomenon, in the context of a research project.

At the end of Chapter 5, a local realist position was adopted regarding the situatedness, of research projects. This perspective is grounded[20], in the sense that all research projects are contextualised in a geo-historically specific way. A core proposal for this thesis is that all psychology research projects, no matter the further philosophical commitments of the researcher, are locally realist, in the sense of taking place in the real

[20] An extensive literature exists in metaphysical grounding which may be a preferred perspective for researchers to adopt insofar as grounding facilitates the adoption of a foundation on which partial ordering can be well-established, with qualities of intranstivity, irreflexivity, and asymmetry, thus supported. See Schaeffer (2009, 2012) regarding this body of work.

world, with geo-historically specific co-ordinates. This fact provides a founding ontology for any further claims that follow about the phenomenon. This holds even where researchers state ideological positions that are distinct from realist positions, such as social constructionist perspectives (see Chapter 4). The conceptual framework presents the researcher with the apparatus to articulate the precise relations between any adopted philosophical position, and the universe of discourse that is the geo-historical situation of their research project.

This first commitment, to a geo-historically specific research situation, sets constraints regarding claims that can be made from the research project analysis or conclusions, given the conditions that will apply at any time and place. Such conditions may be stated in quantitative terms, such as via specific timestamps or latitude and longitude records, or they may be qualitative, such as a statement of circumstances relevant to actual study scenarios including “ceteris paribus” conditions (Suppes, 1962, p. 258; Boumans, 2016). These would include, for example, environmental or social features that may influence the research outcomes. Given the inclusion of both qualitative and quantitative records in the representation of the research project, qualitative techniques are explored that are relevant to providing maximal meaningfulness for our phenomena of interest in our research project in Chapter 7.

The aim in utilising this conceptual framework in psychology research is to establish a representation of a theory of expected relations for the research project, within which use of the GLVM can be framed. This theory of expected relations constitutes a possible realisation of the GLVM model. It is proposed in what follows that the most relevant apparatus for establishing a conceptual framework that takes the research situation as a primitive or base element is a set-theoretical formulation. This is because, as we have stated in Chapter 2, set theory has already been characterised as broad and universal enough to be relevant to mathematical, statistical, and empirical structures (Suppes, 2002;

Ferrario & Schiaffonati, 2012). The body of work of Patrick Suppes provides resources to think through the application of set theoretic structures to not only the theoretical but also empirical aspects of psychology research projects, cohering these terms in a single system. Before drawing from Suppes's works, however, let us return to Spearman and his 1904 paper, looking for what he thought important for characterising the empirical situation of the research itself.

3. Empirical situations and logic in Spearman's work

In Chapter 2, it was noted that in psychological science, researchers rely on the structures of classical logic in the formulation of knowledge. As also stated in Chapter 2, at the time Spearman (1904) developed the mathematical model that became the foundation of the GLVM, the relation between logic and mathematics was undergoing revolutionary transformation (Dauben, 1992; Rescher, 1968). What was introduced in this historical moment was a generalisable function-argument or syntactic form, following Frege (1879/1967) and Boole (1854). Generalisation was thus facilitated in mathematical practice - where once, structures were fixed to specific applications, now there was a way where multiple values could be conceivably associated with the same numeric or logical structure (see Suppes, 1974; Muller, 2011; Ferreriós, 2016). Multiple applications for the same structure could also now be conceptualised, as multiple values could be input into the function and the function itself did not determine application of that same function.

Spearman's original development of factor theory using linear matrix algebra was an application of mathematical structures suitable for the exactness of engineering to psychological phenomena (Lovie & Lovie, 1996). Spearman's concern in developing factor theory was an attempt to make precise the definition of intelligence, using mathematical formulae. He aspired to locate conceptual uniformity for intelligence, later named as g (see Chapter 2). Spearman notes, for example, that there remained a fundamental problem for defining intelligence - one of connecting the "Tests of the

Laboratory" with the "Intelligence of Life" (Spearman, 1904, p. 224-225). What Spearman stated was required in this regard was "a perfectly impartial representation of the whole of the relations elicited by the experiments" (p. 222), from "the entire available data" (p. 225). Spearman was aware of the problem of what he called "errors of observation" (p. 223). Several reflections follow in his 1904 paper regarding the impact of the empirical situation on the mathematical calculations in his correlational approach to the latent variable or common factor:

> For having executed our experiment and calculated the correlation, we must then remember that the latter does not represent the mathematical relation between the two sets of measurement which we have derived from the former by fallible processes. The result actually obtained in any laboratory test must necessarily have in every case been perturbed by various contingencies that have nothing to do with the subject's real general capacity; a simple proof is the fact that the repetition of an experiment will always produce a value somewhat different from before (p. 223)

Further:

> These errors of observation do not tend to wholly compensate one another, but only partially so; every time, they leave a certain balance *against the correlation*, [italics original], which is in no way affected by the number of cases assembled, but solely by the size of the mean error of observation. (pp. 223-224)

In Spearman's view, efforts at replication of a study would always induce variation in study outcomes. He makes here a strong claim about what he perceives as a negligible effect on mean error given an increased number of participants - a claim challenged by psychometric convention, as it has been developed today to deal with mean error (see MacCallum, Widaman, Zang, & Hong, 1999). What is of interest here however is the claim that specific qualitative explication was required, for quantitative precision to prevail. With such precise delineation, Spearman envisaged that errors of observation could be minimised, and that irrelevancies to the concept of interest could be eliminated. In his account, researchers need to collect qualitative information in three steps. These steps are:

a) To "properly define the problem at issue" (p. 226). For Spearman, this demanded specificity to the greatest degree for the definition and the circumstances of the research work, as "in practice we are forced to introduce a large number of conventional restrictions, and for profitable work these must be explicit and unequivocal" (p. 226);
b) To decide on factors that "should be rejected as foreign to our purpose" (p. 226);
c) To search through earlier studies, working in the present context to eliminate factors found to influence the problem defined in a). In the 1904 paper, Spearman conducts analysis of potential confounds for practice-levels, age, and sex, for his definition of intelligence. Quantitative analysis of influence of these "impurities" (p. 236) should be obtained, so that the influence can be evaluated against the results of attenuation calculations.

In the 1904 paper, a rationale is also set out for the range of values adopted for manifest variables such as sensory discrimination, and consistency or reliability of results from earlier studies is evaluated. Validity of his final definition for intelligence is asserted in respect of definitions already established in the literature (p. 238). Spearman further notes the degree to which the measuring instrument or specific test may influence results, specifically in respect of sensory discrimination tests. Overall in summary, when developing a research project that makes use of the original version of the mathematical latent variable model, Spearman stated what was needed was report of: i) the explicit definition of the psychological phenomenon; ii) restrictions that are relevant to this definition in terms of the context in which it is applied; iii) factors or variables that we are choosing not to include in analysis; iv) an account of the particular conditions for the study that may influence results; v) consistency of findings in terms of reliability; vi) matching the conceptualisation of the phenomenon to any earlier works, assuring validity; and vii) assessment of instrumental efficacy.

In present-day GLVM analyses, outcomes may or may not be reported alongside any or all these qualitative specifications. This thesis proposes that any employment of the GLVM in psychological research is more robust when the outcomes of analysis are reported in the context of information that Spearman himself indicated was vital. A systematic approach to recording such information further facilitates the comparison and integration of studies through time. One interesting aspect of the kind of information that Spearman specifies here is the degree to which his 1904 recommendations are echoed in recent recommendations for pre-registration of all psychology studies following the outcomes of the reproducibility project (Open Science Collaboration, 2015). Pre-registration is also proposed to have a role in reducing quality uncertainty in research outcomes (Vazire, 2017), primarily because it supports transparency regarding researcher practices, important for separating, for example, prediction from postdiction (Nosek, Ebersole, DeHaven, & Mellor, 2018). One advantage of the set-theoretical approach to the conceptual framework established in this thesis is that it provides an overarching structure in which pre-registration processes may themselves be unified. This is important as pre-registration processes differ (see below for more detail on this). The next section briefly introduces proposed links between the conceptual framework and pre-registration, before going on to examine the set theoretical formulation of scientific knowledge of Suppes (2002).

3.1 Pre-registration and the GLVM

It is impossible to conduct psychological research today without acknowledging the impact of the outcomes of the reproducibility crisis in changing the norms for research practices (Open Science Collaboration, 2015; Wai & Halpern, 2018). Suggestions for improvements to scientific practices include adopting higher standards of evidence (Simmons, Nelson & Simonsohn, 2011), more direct replications of studies (Makel, Plucker, & Hegarty, 2012; Vazire, 2018), greater transparency (Nosek et al., 2015; Vazire, 2017, 2018), and pre-registration of research projects (see Open Science Collaboration, 2015; van't Veer & Giner-Sorolla, 2016). Notably, the recommendations for what to include in pre-registration processes and databases differ across different fields of psychology research (c.f. Munafo et al., 2017; Grand et al., 2018; Davis et al., 2018). In addition, the distinct proposals made for pre-registration databases to date do not seem to serve as a unifying framework that facilitates the knitting together of the whole of the research project. Rather, the proposals seem to reflect the principles perceived as most influential considering the question of replicability by present day experts, in the specific field. In the majority, the pre-registration processes are aimed at transparency in respect of planned hypotheses for a study (see Nelson et al., 2018; Davis et al., 2018), rather than transparency about overall project assumptions, or specifically, assumptions in statistical modelling such as those that accompany the GLVM. Recommendations for pre-registration for social psychology studies for example (see van't Veer & Giner-Sorolla, 2016) differ from the more general psychology study recommendations (Open Science Collaboration, 2015; see Appendix 1 for tabulation of these differences).

The guidance of experts on what is of import in a specific field is of course to be highly regarded. Question may remain though, as to what extent these expert-informed databases are comprehensive, as the elements that are recommended as inclusions for the different pre-registration processes vary (see Appendix 1 for tabulation of some differences). This stands to be problematic in the future, as researchers may selectively employ pre-registration processes for projects that notionally investigate the same psychological phenomenon, with respect to the use of the GLVM. This leaves the a priori commitments potentially unavailable for standardised investigation and comparison. Pre-registration is a step in the right direction, but the processes described in the literature so far do not extend to questions of integration with a priori declarations made under other pre-registration criteria. Further, pre-registration processes are not aimed at solving the problems specified in Chapter 4 of this thesis.

It is recognised in the psychological research community today that the larger proportion of decisions that researchers make in the process of producing research simply go unreported, in the write-up of outcomes, even with the new pre-registration for hypotheses (Flake & Fried, in press; Barry, Chaney, Piazza-Gardner, & Chavarria, 2014; Flake, Pek, & Hehman, 2017; Slaney, 2017). This conceptual framework aims to upend the trend for the GLVM, in this regard, setting out a formalised structure for the clarification of researcher decisions. A primary goal of formalisation proposed in the conceptual framework as informed by principles of Suppes (2002) is scientific progress like that seen in the philosophy of mathematics through the 20th century, as formalisation was pursued in that field. For Suppes, true scientific advance only follows processes of formalisation are undertaken in a specific scientific field (Suppes, 2002, p. 22). With respect to the kinds of problems outlined for the GLVM in Chapter 4, psychometric science stands to benefit from standardisation and formalisation, in this regard. The next section introduces Suppes's approach to maintaining transparency in scientific modelling, and his approach to formalisation using set-theoretical principles.

3.2 Set-theory for conceptual frameworks

Two problems swim directly into view when considering an application of set theory as described in Chapter 5 to a psychology research project that includes analysis using the GLVM. Firstly, as per Spearman's criteria above, explicit recording of empirical aspects of our research situation is required. It is however notoriously difficult to fit abstract formal notation suitable for set-theoretical analysis to disparate empirical phenomena and properties (Boumans, 2016). Secondly, it has already been noted that both qualitative and quantitative information must be integrated in the definition of the phenomenon that is of interest in research projects. Such information does not easily form patterns suitable for recursive representation, as typically demanded for set-theoretical representation.

Both considerations indicate something of the difficulty addressed by Suppes (1959, 2002) through his years of innovating set-theoretical approaches to models of scientific theory. Suppes remained throughout his work insistent on the integration of representations of empirical situations with abstract theory (see Suppes, 1974; Muller, 2011; Ferrario & Schiaffonati, 2012; Thompson, 1988). Suppes's body of work on set theory for scientific research then provides explicit direction on the interweaving of abstract theoretical terms and empirical situations, in a hierarchy of models (Muller, 2011; Suppes, 1962, 2002; Boumans, 2016), and this is explored, in what follows.

4. Suppes and the informal-structural view of science

Suppes states he had an interest in set-theoretic formulations for models of theory or *representations* from the very beginning of his career in philosophy of science (see Suppes, 1988, 2002)[21]. Suppes's work in philosophy of science is prolific for its breadth,

[21] It should be noted that Suppes has a substantial body of work addressing probabilistic structures in scientific theories. This body of work is not addressed in the present thesis; however, given what Suppes (2011) specifies in terms of probabilistic approaches to phenomena under his Informal Structural View of scientific practice, which must rely on axioms for occurrences, or events, there is every reason to expect that the current conceptual framework would be suited to addressing probabilistic accounts of psychological

covering paradigms from the physical sciences and formal sciences to the human sciences (see as an example Suppes, 1961; Ferrario & Shiaffonati, 2012). In his 1969 anthology of collected papers, Suppes notes motivation to combine the themes of logic, modelling, and mathematics in ways that could support investigations in empirical sciences (p. 3). As stated above, Suppes is possibly best known for his work in measurement theory. Measurement is of course a practice that involves representation, but more than that, for Suppes these representations have a vital role in science, in that they track *invariances.* For Suppes (2002, p. 11), invariance was the property of interest behind axiomatic approaches to measurement, and it is the property which gives representations meaning (p. 97). Invariance expresses constancy, and it is a condition for confirming identity and uniqueness as it is relevant to a representation. With invariance, measurement is not just a random allocation of numerals to empirical properties, but rather, some aspect of orderedness or the structure of number is reflected in the relations established by the measurement system, between some quantitative system, and an empirical situation.

In Suppes's earlier work, his interest in invariances was evidenced in a commitment to axiomatisation (see Suppes, 1951; Suppes & Winet, 1955; Thompson, 1988) with a view to clarifying invariance under transformation. These are known as morphisms. A morphism is a precise transformation under which a set of empirical relations are said to hold, as discussed above in the section on measurement and the GLVM above. Suppes (2002, p. 111) notes that empirical systems typically are very complex - and this is certainly true for instances of psychological phenomena. As stated in the section covering Markus and Borsboom's (2013a) account of axiomatic approaches to measurement, tracking psychological phenomena in ways that are amenable to representation under morphism remains always difficult, because of the complexity of the

phenomena. See chapter 8 for further detail on notions of phenomena in psychology research as events.

phenomena, themselves. In Suppes's later work, the commitment to axiomatisation remains, but what can be observed is a slow shift in the direction of increased recognition of the complex empirical circumstances in which research takes place, with reduced focus on the constraints of morphic structures (Muller, 2011). In his final works, uncertainty in scientific processes is pursued as a topic particularly relevant to probabilistic modelling (Suppes, 2016). Uncertainty as a theme for psychometric practices will be discussed in Chapter 7. What remains in Suppes's work throughout, however, is an orientation to the clarification brought about, by axiomatisation, in model form.

4.1 Suppes's account of models

Suppes's early axiomatic formulations cite indebtedness to Tarski. Tarski (1933) in his earlier work had followed Frege (1879/1967) in distinguishing between two types of entity in logical argument systems, constants, and variables. Tarksi (1956) accounts for a formulation of model that adds one more entity type. Tarski's concept of a testable model is {U, f, R}, where U = a set-theoretic structure, describing the domain or universe of overall discourse, f = a defined map or model relevant for the context of the problem in question, clarified in axiom terms, and R = variables defined over f. For Tarski (1956), such models could be tested for deductive closure, regarding their status as a model. All terms of the model were specified as a system of axioms, which could be tested by holding constant all axioms, except for one. If the outcome of the test meant that the model produced a 'true' value, the model was a model, and if it produced any value other than 'true', it was not considered a model.

In *Logics Appropriate to Empirical Theories*, Suppes (1975) maintains the model structure of Tarski, but dispenses with strict deductive logical tests, in favour of a pluralist perspective that considers the elements that are of import for any empirical scientific situation. The meaning of a model is maintained as a structure consisting of domains for scientific phenomena, with relations defined over the domains, and with certain conditions held to be satisfied (see Moulines, 2006). Such an approach facilitates inclusion of elements of different logical type, as:

> the exact analysis of the relation between empirical theories and relevant data calls for a hierarchy of models of different logical type. Generally speaking, in pure mathematics the comparison of models involves comparison of two models of the same logical type, as in the assertion of representation theorems. A radically different situation often obtains in the comparison of a theory and experiment. Theoretical notions are used in the theory which have no direct observable analogue in the experimental data. In addition, it is common for models of a theory to contain contiguous functions or infinite sequences although the confirming data are highly discrete and finitistic in character. (Suppes, 1969b, p.25)

This simple structure gives the basic scheme for all elements in the conceptual framework described below for the GLVM. For each component of the research project, the same structure is maintained, and an ordered set is adopted. More detail is given regarding this structure below. The other technique of import employed by Suppes in his formulation of representations of invariances is axiomatisation, and we turn now briefly to consider his approach in this formulation technique.

4.2 Suppes's approach to axioms for set-theoretic models

Axiomatisation sets a series of constraints or co-ordinates, under which claims can be assessed within their own framework, for truth or falsity (Suppes, 2001). Axiomatisation allows for logical assessment of the series of claims derived under their exercise. At base, logical assessment makes use of classifications of true or false, no matter the degree of generality or the order of the logic, be it first or higher order. Axiomatisation is of value in circumstances where generalisation from one situation to another is sought, because with this logical checking process, axiomatisation has within its structure preservation of the representations by which invariances can be tracked and secured.

The axiomatic technique is cited by Suppes (2001) following Knorr (1975) as having roots in the 4th and 5th centuries BC, with preservation in Book V of Euclid's *Elements* where Eudoxus's theory of proportion was the first detailed recording of axiomatic form. This and the work of Aristotle in the *Posterior Analytics* (1994, 74a-17) are important because they are proven not for single geometric objects, but for magnitudes in general. This format for the approach was enduring, lasting at least up until the 18th and 19th centuries when several problems emerged. The discovery of non-Euclidean geometries by Bolyai, Lobachevski, and Gauss and developments in other fields such as those presented for mathematics in Chapter 5 led to new techniques for axiomatisation. Suppes (2001, 1954) states the main ingredients for the modern approach to axiomatising theories are: (i) statement of primitive concepts for the theory; (ii) statement of any prior mathematical or formulaic basis assumed; (iii) statement of the axioms; and (iv) characterisation of the models, for the theory. These steps, taken together constitute the representation of a theory.

4.3 Connecting axioms to the conceptual framework

The conceptual framework developed within this thesis sets out a formulation of a research situation and its expected outcomes as a theory of expected relations. The axiomatic form applies to the structure of the research project elements, which are models of identical structure, but distinct natures. What this means is that each research element is standardised as a representation. The primitive concepts are the sub-elements, and these are defined according to the model form first articulated by Tarski (1933). In the next section, Suppes's modification of Tarski's original formulation of a deductive mathematical model is discussed, which leads naturally on then to the definition of a minimal set of elements suitable for a conceptual framework where the GLVM is employed.

4.4 Conceptual framework, formulation

As stated above, Suppes (2002) followed Tarski (1956) in his formal definition for a model. Tarski designed his approach for articulating deductive closure for mathematical theories, or models. In the conceptual framework for the use of the GLVM, deductive closure is neither desired nor possible, given the empirical nature of the research situation that must be accounted for in the representation of the project that includes the GLVM. Instead, a basic relational structure is adopted, informed by Suppes's approach to axiomatisation of set-theoretical models following Tarski, preserving meaningful ordering of the sub-elements. The basic structural form for conceptual framework elements is consistent across all elements, being an ordered relational structure, $<U, A, V>$. *U* is the universe of discourse for this element, *A* is the specification of the representation in terms of principles or definitions, and *V* stands for variables, as relevant for the element. While the sub-elements form a set, insofar as they belong together, the ordering format here gives a relational structure to the research project element, as a representation. What this means is that each subsequently defined sub-element falls under or is a part of the previous sub-element, in the order. Example sub-element entries for elements needed for an implementation of the GLVM are described below and are followed with two empirical examples of retro-fitted conceptual frameworks, in Chapter 8.

In Chapter 7, rationale for a network systems approach to conceptual frameworks is discussed. Such an approach means that there is no strict hierarchy assumed in the conceptual framework, in terms of relations between elements, even though ordering is presumed, within-elements. Advantage however may accrue when we take as the universe for an element some other element, already defined in model form, when considering the role of constraint setting for applying the conceptual framework to the analysis outputs from the GLVM. The advantage comes about insofar as it is possible to consider such a hierarchical arrangement as characterising a constrained field. When for example, constraints founded via a philosophical perspective such as Borsboom's or Haig's as described in Chapter 5, are adopted, which is in turn used to develop a model which falls under the GLVM paradigm of perhaps Bartholomew et al. or McDonald, these will have specific implications for what needs to be reported for a meaningful GLVM analysis to follow. The constrained fields support specification of the kinds of conditions that apply to the element(s) or ultimately the GLVM, as limiting restrictions within which any research outputs can be considered.

5. The conceptual framework - elements for the GLVM

There are some key elements for research projects involving the GLVM. These are research concept, construct, theory, model, variable, data, and phenomenon. The universe of the research project itself will require specification as the research situation, and a researcher element facilitates the account of philosophical or ideological commitments. We will address some formulation for these one by one, below, and two worked examples are included in Chapter 8. One important aspect of the kind of information that fits in the relational structure as stated above is that it will include qualitative information. With that in mind, criteria for trustworthy qualitative research information is explored in Chapter 7. The basic structure of research project elements for the GLVM is set out, in what follows in this chapter.

5.1 Elements

Above it has been said that an important feature of defining the elements for a research project is the representation of sub-elements, following Suppes/Tarski. Specifically, the sub-element structure works to specify: i) U = the universe for the element, or the relevant context for it; ii) A = specific defining features for an element; and iii) V = any variable values that are relevant for the element. An element for a model may include, for example: i) U = theoretical formulation for the phenomenon of interest; ii) A =

definition of the model in words or in syntactic form; iii) *V*= other variables of influence for the element. When taken as a whole, the elements of a project formulate a representation of the research project. This representation functions as articulation both of the a priori commitments of the researcher prior to beginning a research project, as well as the constraints within which research project outcomes are interpretable. As much information should be provided as facilitates an assurance of quality certainty, for an element, which may be guided by the specific pre-registration process chosen by the researcher (see Appendix 1 for examples, and for more detail regarding satisfaction of quality certainty, see Chapter 7). A check of the pedagogy of production of APA style psychology reports in the *Publication Manual of the American Psychological Association* (American Psychological Association, 2010) suggests researchers should include as much information as would render the study reproducible by a fellow researcher. The kind of elements that are included for the whole project, and specific formulation of information within the elements and the relation of the elements to each other remains freely determined by the researcher. Linking of specific sets of elements together is possible, prior to inclusion in the whole of a research project representation. This may serve well where a series of item batteries are used in evidence of a singly conforming latent variable as was discussed in the section on infinite behaviour domains, above (see McDonald, 2003). A brief summary of key information for a GLVM example is presented in Table 6, below. In the detail that follows regarding the research elements, we make use of an example of a of straightforward employment of CFA in a research program, to provide an example of methodological inclusions in the conceptual framework, and to point out the ways the problems of Chapter 4 are addressed, in closing this present chapter. In Chapter 8, an example from the psychiatry literature is developed for an example adapted from substantive peer reviewed reporting literature.

Table 6

Conceptual framework for the GLVM general description and layout

Element	Universe/domain	Formulation	Variables
Situation	Field in which this study takes place including background (e.g. type of psychology)	Defining details of substantive research project situation, purpose, type of study	Features of the study situation or environment. Also include population and sample parameters
Researcher	Context of involvement in research project production (e.g. university affiliation)	Researcher name(s)	Ideological, philosophical, institutional or other commitments
Research Concept	May be defined for a specific research situation or may be the universe/domain sub-element for a whole field of research	Description of the concept	Relevant alternative formulations of the concept that are explicitly not adopted
Research Construct	Research concept	Description of the specific construct for this population	Any population specific parameters of relevance, any specific exclusions from the construct
Theory	Context of theory e.g. field of study, background, grand theory, translational, foundational, methodological	Theory description	Variables identified for this theory

Model	Context as theory or hypothesis - is this a substantive, statistical, or data model	Definition of model, specifically relations amongst variables	Variables relevant to model (e.g. as defined in phenomena or variables element), also variables known but excluded
Variable	Context of application for variable – model for which variable is relevant	Definition of variable	Range of values and calculation method; manipulated versus non-manipulated, moderator or mediator
Data	Research situation, model, variable, or phenomena for which the data is collated.	Data description	Aspects of data e.g. type of information, whether transformation has been applied, aspects of data collation, handling of missing data and data deliberately left out
Phenomenon	Context of phenomenon (e.g. research situation, theory, model)	Definition of phenomenon	Space-time location, properties for organism and environment, nature of relation between organism and environment (e.g. proximal, distal, modal)
Quality control	Research situation	Listing of quality control protocols adopted	Any considerations relevant to quality control application in the research situation

5.2 Addressing problems for the GLVM with the conceptual framework

In Chapter 4, a series of concerns for present-day use of the GLVM were set out. In

Table 7 below, each of the concerns covered in the previous chapter are listed, with the address of the concern made possible for the GLVM in the conceptual framework listed in the final column. For the key concerns of conditional independence, factor indeterminacy, and circularity, you can see that the suggestions each involve a resolution whereby a domain element is clarified for the latent variables, manifest variables, and links between the research situation, the concept of interest, and the construct, for the latent variable. Latent variables and manifest variables are not explicitly distinguished in the general table above but would be uniquely identified via separate and nominated variables elements. A retrofitted example conceptual framework for a published research paper that makes use of the GLVM follows in Chapter 8, which facilitates a direct example of what would be included, in addressing aspects of a research project. In what follows the table below, some generic aspects of the elements are explored, with some detailed address of the problems from Chapter 4 is included.

What follows is by nature abstract and shows just some of the links between elements that are possible to provide a structure for the support of research claims made following GLVM analysis. As described in Chapters 4 and 5, part of the address of the problems to do with applications of the GLVM comes from establishing consistency between research elements in a representation of the research project. Part of the problem in working to specify how this looks is that each project will make use of substantially different approaches, with respect to what is specified in the research elements. Some examples of problem solutions are contained in what follows, with the aim of remaining informative, without being too abstract.

Table 7

Problems for GLVM in research applications addressed in the conceptual framework

Type	Problem	Specific issue

Theoretical	Conditional Independence	Specify evidence for independent existence of phenomena behind the latent variable (typically the research concept) in the phenomena element; make this the universe for the research concept element. Include relevant statistical assumptions in GLVM model element (variables sub-element).
	Factor Indeterminacy	Specification of the philosophical perspective on the GLVM model (variables sub-element); specific common property held to underscore item outcomes on the manifest variables in the variables sub-element for the research construct.
	Circularity	Specify the research situation as the domain for the research construct; explicitly declare the research concept as the domain for the research construct (which then becomes the domain for the latent variable).
Application	Between vs within subjects	Declare in the variables sub-element of the latent variable element the conceptualised relation regarding cross-sectional data and within-subjects processes (or as the relation requires).
	Causation	Specify extra-statistical assumptions about causation in the variables sub-element of the model element.
	Equivalent models	Declare alternative model in the model element (variables sub-element).
	Variable structure	Provide any relevant information about the assumed variable structure in the latent variable element (variables sub-element).
Quality control	Unidimensionality index	Address in (variables sub-element) for the latent variable element; declare this as a secondary link to quality control element.
	Model fit	Address in (variables sub-element) for the model element; declare this as a secondary link to quality control element.
	Measurement invariance	Address in (variables sub-element) for research construct element; declare this as a secondary link to quality control element.
	Reliability and validity	Address in quality control element (variables sub-element)

5.2.1 Research situation

It has been stated several times that the research situation is understood via local realism as a founding constraint that sets the initial conditions for the universe or domain of discourse for a research project. For psychology research generally in the wake of the credibility revolution, the role of the context of study or research situation for psychological phenomena has surfaced as an under-considered feature in questions of reproducibility (Van Bavel, Mende-Siedlecki, Brady, & Reinero, 2016). Methodological consideration of context for research in psychology, however, has a longer history than this, with origins in Spearman (1904), as noted above. Writing half a century later, Brunswik (1952, 1955) in his work on ecological and representative design for psychology studies emphasises something unusual amongst his peers. This was a focus on the impact of the environment on an organism. The environment, for Brunswik, stood to have as great an impact on research study outcomes as any inherent difference, in the individual. So important did Brunswik conceptualise this to be, that in his 1952 monograph setting out a conceptual framework for psychology, he argues for correlational designs for research that included correlational analysis of environments, prior to investigation of any psychological variable. Psychology as a field has made little use of this insight subsequently (Hammond & Stewart, 2001), perhaps to our collective detriment as suggested by Van Bavel et al. (2016), who comment on contextual sensitivity with respect to questions of generalisability.

In terms of the ordered relational structure for the research situation element, when specifying the details relevant to the domain or universe, U, for a research situation, parameters that are relevant to the inferences made from the study findings should be specified. Geo-historical contextualisation may include the physical locations and times of the conduct of the research. The formulation sub-element, A, would include description of the study and its context as relevant to the field of psychology, while in the variables sub-element, V, may contain further information which may be expected to vary in any replications or future studies, for example, the number of participants/site. It also may be where a researcher chooses to record any aspect of the situation that may have an impact on reproducibility, of findings.

5.2.2 Researcher

In Chapter 5, it was noted that research credibility is enhanced by inclusion of information regarding the researcher's own commitments and philosophical positioning, in respect of the research. This already generally occurs in qualitative psychology research, whereby qualitative researchers typically account directly for their expectations, as well as their prior investigations (Tracy, 2010; Creswell, 1998). To the degree that it has already been made clear in this thesis that use of the GLVM constitutes as much a qualitative as quantitative enterprise, GLVM researchers benefit the research community when they account directly for their expectations and prior investigations. Ponterotto (2005b) specifies several stances informed by philosophy that a researcher may seek to endorse as a position relevant to their perspective on the project, including emic or etic positionings, ideographic or nomothetic orientations, qualitative or quantitative approaches, positivism or post-positivism, constructivism, interpretivism, or other stances such as the various realisms discussed in Chapter 5. As clarified in Chapter 5, each of these will have distinct implications for what is specifiable in subsequent elements for a research program, in terms of consistency. For example, an etic perspective is considered to imply that the phenomenon of interest is one that is characterised by universal law or principle - applying equally to all people or species in the situation (Ponterotto, 2005b). An emic perspective on the other hand is contextualised to a local community - so population parameters for statistical testing of the GLVM for example would be curtailed and finite with respect to a local population. For the researcher element, the universe of discourse sub-element, *U*, may include the researcher's affiliations and institutional commitments, while the formulation, *A*, may include any ideological or philosophical perspectives that they see as relevant. Such specification may also include any deliberate negation in the variables sub-element, *V*. For example, a researcher may wish to declare that they are deliberately not social constructivist in orientation, to exclude possible interpretation of their findings in this way. Such specification facilitates construction of expected relations within which the research outputs can be understood and interpreted, as described above.

5.2.3 Research concept

The primary role for specification of the problem in Spearman's (1904) terms or the research concept, becomes clear when the research situation is taken as the domain, *U*, for the research concept, which can then be connected as the domain for the research construct. Following the discussion of Markus (2008) in Chapter 5, clarity is thus facilitated in respect of the possible, and actual, populations over which the concept and the construct respectively are considered to apply. It also facilitates comparison of distinct constructs for the same concept, where for example the GLVM may be used in support of building or adapting substantive theory (see Wilson, 2013a for an example), or comparing equivalent models with distinct constructs. The formulation sub-element, *A*, would include description of the concept, and in the variables sub-element, *V*, it may be expected that any specific restrictions or omissions for the concept are clarified. Another important inclusion in the variables sub-element, *V*, for the research concept would be independent evidence for the existence of the phenomena relevant to the concept, where such exists. This is in aid of solving problems to do with circularity and the operation of the constraint of conditional independence as described in Chapter 4, by putting in place the series of constraints that operate in reference to the research claims and outputs.

5.2.4 Research construct

In specifying the research construct following the distinctions between concepts and constructs set out in Chapter 5, it can be imagined that there is benefit in demonstrating consistency of the representation of the research project in the conceptual framework, by linking the previously defined research concept element as the domain sub-element, *U*, for the research construct. The formulation, *A*, would then contain a precise definition for the construct, which includes specification of the common property relevant to the research concept, which can be assessed in some way, using scientific practice (Slaney & Racine, 2013). The variables sub-element, *V*, can be expected to include any specific omissions, as well as any specific points about application. If the construct is conceptualised as formulating an infinite behaviour domain following McDonald (2003), this can be clarified in the variables sub-element. An important aspect of defining the research construct is the role that this plays in demonstrating the kind of consistency in the research project representation that will facilitate steps towards overcoming the problems to do with the use of the GLVM. For example, nominating an infinite behaviour domain constitutes one technique highlighted by McDonald (2003) that can be interpreted as a resolution aimed at factor indeterminacy.

5.2.5 Theory

Adams and Buetow (2014) note that theory permeates the whole of a research project in distinct ways, but in ways that are in the majority marginalised in psychology research. Adams and Buetow (2014) perceive that psychology has dispensed with a focus on theory, in what they characterise as a paradigm-wide obsession with method, comments echoed in other voices in the field (Cliff, 1992; Wilson, 2013a; McClimans et al., 2017; Muthukrishna & Henrich, 2019). In an example for the GLVM, in the theory element, a researcher may choose to nominate the theoretical latent variable modelling form which they have chosen – for example McDonald (1999), or Bartholomew et al. (2011), or some other GLVM framework that comes from beyond the psychometric stable – for example Skrondal and Rabe-Hesketh (2004) or Muthen (2002).

One of the important aspects of such a nomination is what it offers in terms of both self-audit and community review, when it comes to assessing the research claims that follow the completion of the project considering the research methods. This is particularly in respect of consistency of the perspective as it is applied to the research elements. For example, as we had stated in Chapter 3, the posterior moments position that follows from the theoretical GLVM of Bartholomew et al. (2011) asks that the latent variable be treated as a random variable, which would need to be specified in the variables sub-element, *V*, for an element that clarified the latent variable domain. Bartholomew proposed that the structure of the latent variable could be determined by the use of the GLVM with empirical data. Where a researcher specifies a theoretical GLVM of Bartholomew et al. (2011) in the description sub-element for the theory element, *A*, it would be expected that the latent variable element also specifies in the variables sub-element the range and structure for the latent variable, given such specification. Other elements that would contain some specific information following the nomination of the Bartholomew et al. (2011) theoretical GLVM in the theory element would include a nomination made for the model element in the variables sub-element, *V*, regarding the resolution of the problem of factor indeterminacy using the posterior moments position, where individuals are conceptualised as located by the model at positions on the latent variable. There would also need then to be an inclusion in the variables sub-element, *V*, for the variables element for the latent variable some rationale for how between-subjects analysis is applied to within-subjects processes, for the psychological phenomenon or construct, in question.

5.2.4 Model

Psychometric modelling has been conceptualised as “modeling the encounter of a person with an item” (Whitely, 1983, p. 184). While this may describe the province of some educational testing, and even personality assessment in psychology research, it falls short of all that we ask the GLVM to do, in psychometric contexts. The GLVM is employed today extensively in questionnaire construction and production, but this does not constitute the full extent of its work as a model in our field. For example, it is more and more frequently being used in discernment of correlations between symptoms in large databases from psychiatric patient studies (Marquand et al., 2016). An example that makes use of the GLVM in computational psychiatry literature is set out in Chapter 8.

In connecting methodological theory to the CFA model in the model element as an example, an appropriate universe or domain, *U*, for the CFA is the GLVM. In the formulation sub-element for the model, *A*, the model and its relations can be specified including error structure and identification assumptions. Assumptions that accompany the CFA particularly may be included in the variables sub-element for the model, V, and this may include, for example, distinguishing between the latent and manifest variables, the nature of the relationships, and the handling of error.

It was noted in Chapter 2 that even though both Bartholomew et al. (2011) and McDonald (1999) propose versions of the GLVM, they advocate for different kinds of uses for the model – Bartholomew et al. in processes of data reduction and latent variable discovery, typically in EFA analyses, McDonald in test refinement, typically in IRT analyses. In this way then, for some studies, there will be an important connection between theory and a substantive psychological model which may take form as the GLVM, and for other studies, this will not be as important (see also Haig, 2014 regarding some processes of abduction, which may not include directly a substantive model). Where substantive theory informs model structure for the GLVM, an important aspect of the model is the degree to which it is informed by theory that guides expectations and predictions for a certain set of relations amongst variables in the model (Edelsbrunner & Dablander, 2018). In this way, taking a pre-defined substantive theory element as the universe of discourse for a model sub-element facilitates the connecting of the model back to the theory components.

In considering other resolutions for the concerns presented in Chapter 4, aspects of assumptions that are to be checked with the specific formulation of the model also can be beneficially identified in the variables sub-element, *V*, including model fit, model invariance, model equivalence, and statistical independence checks. To the extent, that causal inferences are made from the phenomenon behind the latent variable to the manifest variable outcomes, extra-statistical information needs to be provided to support the causal inference in the variables sub-element, *V*. This is because there is not enough information in the model itself to support a causal inference, as was discussed in Chapter 4 and 5.

5.2.5 Variables

One important aspect of variable definition for the GLVM is the degree to which variables must be clearly specified for their range and relations (Spearman, 1904; Mulaik, 2010). This is also important for the latent variable, where an infinite behaviour domain is specified in aid of working towards addressing factor indeterminacy, following McDonald (2003). Mulaik (1986, 2010) notes that definition of the term variable differs, in the psychometric literature, which may mean researchers adopt different positions that should be declared, regarding their interpretation of what a variable is and does. For example, Mulaik notes the term variable has been described as that which "pertains to a set of mutually exclusive and exhaustive predicate classes having a common form" (Mulaik, 1986, p. 315). But also, "[a] common definition of a variable is a quantity that may take any one of a set of possible values" (Mulaik, 2010, p. 51), or "[a] property on which individuals in a population differ" (p. 51). The definition that Mulaik (2010) finally adopts for latent variable modelling is:

> A variable is a functional relation that associates members of a first set (population) with members of a second set of ordered sets of real numbers in

> such a way that no member of the first set (population) is associated with more than one ordered set of real numbers at a time. (p. 51)

For Mulaik (2010), there is also a classificatory system that is relevant for variables. Variables may be discrete, where the variable could take, at most, a countably infinite number of values, where each value can be put in one-to-one correspondence with members of a set of integers, or continuous, where the variable value could take uncountably infinite possible values. Mulaik (2010) notes the question of whether variables are discrete or continuous however is less important than the question of whether the population that is theorised for the construct in question is finite, or infinite. For infinite populations, it is not possible to arrive at actual counts for each variable value, due to the uncountable infinite order, so, we take expected values for an uncountably infinite population, given what we find from samples (Mulaik, 2010, p. 54). What this means in reference to the conceptual framework is that there is a role for setting the domain for any variables which are set up as a research element, either latent or manifest, to reference a declared research concept or construct. Remembering as highlighted above, concepts following Markus (2008) range over possible and actual cases, while constructs range over actual cases for a population.

We also saw an interpretation in Chapter 4 from Bartholomew (1987) that latent variable model variables are random variables. Strictly speaking, this indicates a much more complex process mathematically than the operations of addition and scalar multiplication would reveal. The only way to make sense of a random variable in this scenario is as a random real-valued quantity, where the values depend on an experimental outcome that operates under pure chance. Where a researcher seeks to make use of the property of randomness, what is needed is some specification of the relation between randomness and the research situation structure stands to inform both peers and future researchers about the constraints they themselves may need to apply for either study comparability or variable applicability.

5.2.6 Data

Data as a term is not simply defined, a fact that is as true of psychology as it is of sciences, generally. Leonelli (2014, p. 400) for example defines data as “mobile pieces of information, which are collected, stored and disseminated in order to be used as evidence for claims about specific processes or entities.” Key aspects of this definition include that data formulate evidence for scientific inferences, and that data can be communicated effectively to different stakeholders and utilised in different circumstances for different ends. Leonelli notes that the assumptions that are made in the production of data impacts upon the ways that data play roles in evidence for subsequent claims. Specifically, Leonelli (2014, pp. 409-410) cites information about the study setting, chosen categories for data, and the data scale or structure for representation of phenomena as key characteristics. These can be recorded in the variables sub-element, *V*, for data. The domain for data, *U*, can be varied, given what we have said above. It may be that data are collected as part of a larger separate study, or indeed these days, passive data collection with no real intentional purpose in the collection stage is fast being accumulated, for example, from smartphone usage patterns (for example Cao et al., 2017). The formula or principles sub-element, *A*, will give a general description in terms like the definition of a variable. Specific properties relevant to data are specifiable in the variables sub-element, *V*. For example, data scale structures, unit of measurement where appropriate, any modelling aspects relevant for the data point such as whether its value is determined over some averaging function.

5.2.7 Phenomenon

As stated above, researchers are interested in invariances in general, in science. What is interesting to think through for psychology is the question of what sort of invariances are of interest, and how these might present in data patterns, of the sort that

Spearman first noticed. Correlations are interesting in the context of invariances because correlation is formulated entirely from a set of relations. Relations can only be possible between two points of difference. Even the logical relation of equality or a relational structure of constraint relies on two distinct elements or attributes that are brought into relation with each other. When we describe isomorphism, what we are describing is perfect correlation, from one system, to another. In this scenario, there is no variation that occurs in one system without the same variation occurring in the second system. However, there always remain two systems, and correlation, particularly in the context of latent variable modelling, is only of interest when there is variation in the variation – when there is some variance, in invariance. These are the patterns that were of interest to Spearman - he conceptualised these variances in the invariant common factor as individual differences in students.

Haig (2014) notes that phenomena have been characterised in the philosophy of science literature in terms of law-like generalisations (p. 33), "that includes objects, states, processes, events, and other features that are hard to classify". One concern with characterising psychological phenomena as law-like generalisations is that it does not easily facilitate an understanding of the ways that social and institutional factors may influence the existence of the phenomenon of interest – situational concerns, under a local realist perspective. Given this, an events perspective on psychological phenomena is endorsed in this thesis. Examples include the occurrence of symptoms of mental illness, or the exercise of mental ability. These events occur within specifiable organism-environment relationships, and at least four pieces of information can be said to characterise any given psychological phenomena: i) space-time location; ii) property relevant to the organism; iii) property relevant to the environment; iv) property of the relation between organism and environment (see Petocz & Mackay, 2013; Maze, 2001; Boag, 2011). In characterising the phenomena for a psychological research project, these four pieces of information included in the variables sub-element, *V*, supports basic characterisation of the phenomena that is of interest as described in the description sub-element, A, which may have as a domain sub-element, *U*, both the data element and the concept element. Circularity thus becomes evident, in the structure of the conceptual framework, in that the phenomenon both has a role in the invariance that is of interest, in the research concept, and the variances in the invariance, which are of interest, in the data. Such non-linearity will demand special evaluation, in terms of the support it provides for inferences made following the completion of a GLVM analysis in a psychology research project. Chapter 7 goes on to describe a network approach to systematisation and introduces some techniques that aid in assessment of its quality control function. Further discussion of quality control then is postponed, to the next chapter.

6. Summary

Following the examination of themes of measurement for the GLVM, this chapter began with acknowledgement of Spearman's requirements for detail to support analyses following applications of his factor theory. Spearman suggested we need to account for an explicit definition of the psychological phenomena; any restrictions relevant to the definition; any variables we choose not to include in our model; note of ceteris paribus conditions associated with the particular study that may influence results; a report of consistency or reliability of our findings; reconciliation of our current conceptualisation of the phenomena to earlier conceptualisations and finally, some address of any instrumental uncertainty associated with the methods used. The conceptual framework makes possible a standardised presentation of this exact information, in a model structure following Suppes (2002), given a local realist approach to the geo-historical research situation. The universal structure of elements means that a definition of the concept can be stated in a principle sub-element for an element for a concept; any explicit restrictions and exclusions for the concept can be noted in the variables sub-element, for the concept element. Also possible is to record in the variables sub-element for the concept element are any relevant details for earlier studies. The conditions of the research situation are notable in the variables sub-element for the research situation. Reliability, validity, and instrumental uncertainty inclusions may each be included under the quality control element, which remains as yet not comprehensively described, but which places us now at a perfect point to turn to Chapter 7. In this chapter, we will learn more about constructivism, a role for a network systems approach to the conceptual framework facilitating inference from the best systematisation, and finally, how all this fits together as a suite of activities that serve in overall quality assurance, for projects making use of the GLVM.

In summary, what is claimed here is that that the syntactical GLVM model as presented in different forms, for example, by McDonald (1999), Bartholomew et al. (2011), and Skrondal and Rabe-Hesketh (2004)[22] only becomes meaningful in the context of psychology research when it is presented in a conceptual framework, as is exemplified above. This is because any output that is obtained in the process of constructing a statistical analysis that includes reliance on the key assumptions of latent variable modelling as discussed in Chapter 4 demands an account of these very assumptions, because the assumptions themselves do not apply in all circumstances, in all conditions related to psychological research, equally. What has been set out above is a structure for the conceptual framework, what has not yet been presented is rationale for the systematisation of such information, and this is what follows, in the next Chapter.

[22] Note that it is possible to conceptualise the syntactic definition of the GLVM as specified within the Model/Variable/Data elements of the GLVM, however the whole of the conceptual framework is argued here as necessitated, by the very definitions that are needed for the formulation of the model in the context of psychology research.

Chapter 7: Conceptual Framework – Constructivism and Systematisation

1. Objectives

The objectives of this chapter are to elaborate a systems view of the conceptual framework, and to examine a stance informed by mathematical constructivism, in considering the representation of the use of the GLVM in the conceptual framework. Each of these is carried out with a focus on how systematisation and constructivism support evaluation of quality certainty, in respect of research claims made following GLVM analyses. Argument for a realist interpretation of the geo-historical situation of research was presented in Chapter 5, section 5. Examination of classical proof distinct from constructivist mathematical proof was presented in Chapter 3, section 3. In the present chapter, in recognition of the fact that psychological research using the GLVM relies on modelling, which involves constructing models, mathematical constructivism and the history and nature of constructivist proof is explored. The specific interests in mathematical constructivism for the conceptual framework are: i) how constraints are formulated for theorems and proofs in analytic distinct from axiomatic ways that support pluralism in logic; ii) how contexts and constraints combine in the process of producing constructivist proof; iii) in what way constraints are connected to meaningfulness; and iv) how strong intersubjectivity works in mathematical constructivism.

The previous chapter introduced the conceptual framework, having drawn on insights from the work of Suppes (2002) in philosophy of science. The conceptual framework, as presented, is informed by a set-theoretical model structure, suitable for representing the invariances of interest to psychology researchers, in a way that facilitates transparency regarding the assumptions and decisions that a researcher makes in the process of conducting a project. Tracking the 20th century evolution of thought in the field of mathematics that followed in the wake of the introduction of set theory, this chapter notes particularly the role of cognition for mathematical activity, and the way that cognition is embedded in contexts, with specific constraints that apply for cognition, and for the context in constructivist mathematics. Drawing on Ferrerios's (2016) *Mathematical Knowledge and the Interplay of Practices*, a conceptualisation of cognitive mathematical practice is explored, which highlights the importance of context for knowledge, the role of a community of practitioners in evaluating contributions to knowledge, and the importance of logical consistency for statements made in procedures of verification under constructivism.

Following this, the account of cognitive systematisation of Nicholas Rescher (1979, 2016) is presented in its role for the conceptual framework, along with his process of inference from the best systematisation. Derived via objective pragmatism (Rescher, 1997), the formulation of cognitive systematisation is connected back to local realism as presented in Chapter 5. A case is then built for inference from the best systematisation using local realism grounded in the geo-historical context of the situation. Inference from the best systematisation extends to address the alternative scientific processes of description or prediction for which the GLVM may be employed, beyond explanation. It is argued that inference from the best systematisation is the most appropriate inferential system for a conceptual framework for the use of the GLVM, given its generality. Next, an assessment of systematisation for a representation of a research project is discussed in the last sections of the present chapter, in aid of providing an assessment of consistency as discussed in the section on constructivism. The purpose of this is to provide an assessment of the degree to which the conceptual framework provides a quality control activity, for the representation of any research project. With this, all keys are in place for presentation of two worked examples of the elements of the conceptual framework retro-fitted to previous studies, with some comparison made of the outcomes included in the final conclusion in Chapter 8.

2. Constructivism

In Chapter 5, it was proposed that all psychology research at minimum is realist, insofar as research occurs in real situations that have geo-historical specificity. The proposal for the conceptual framework for use of the GLVM itself however is best be understood as realist constructivist. The position may be considered as philosophically unusual, as realism and constructivism have been presented as contrasting perspectives in opposition to each other (Borsboom, 2005), both in philosophy of science (for example, Henderson, 2018) and philosophy of mathematics (for example, Friend, 2014). The contrast comes about because realism is typically characterised as aimed at truth regarding a mind-independent world, while constructivism typically takes its constraints on truth as articulated in constructions, often in the minds of researchers in a field (c.f. Borsboom, 2005).

In this thesis, local realism is adopted for the conceptual framework in reference to psychology research, in recognition of the fact that a research project occurs in space and time with interest in some phenomenon or event with relevance to psychological functioning. Realist *constructivism* is necessary for the use of the GLVM in psychology research, as the GLVM itself is a model, constructed in representation of some aspect of the psychological phenomena of interest. While an account of local realism was presented in Chapter 5, no account of constructivism has yet been presented. Constructivist approaches are explored below with a view to systematisation of the conceptual framework with specific reference to constraint setting via domain specification, in aid of quality control in respect of evaluation of research outputs. The aim is to support transparency and cohesion of disclosure of researcher decisions and research practices. Constructivism in mathematics is of interest to us to the degree that mathematics is taken to be the most certain and universal of scientific languages (Grattan-Guinness, 2008). Constructivism has come to play a central role in formulations of mathematical practice in the 21st century (see Ferreriós, 2016; Wagner, 2017).

2.1 Constructivism in mathematics

Mathematics for our purposes here is understood as "a science of the real world, just as much as biology or sociology are" (Franklin, 2014, p. 1). This perspective on mathematical practice is endorsed in inquiry in philosophy of mathematics which looks to what mathematicians treat as real, in their practices (see Maddy, 2011; Shapiro, 2000; Feferman, 1998). The refrain extends a trend in mathematics away from classical foundationalism which began at the end of the 19th century (Troelstra, 1991). Foundationalism in mathematics proposes that the only valid mathematics is one built on secure foundations with an incontrovertible basis (Shapiro, 2000). The strength of logical deductive closure for a proof under foundationalism implied that the conclusion was applicable in a universal law-like sense (see Shapiro, 2004), relevant in the same way to all times and all places.

Constructivism, broadly defined, maintains attention to the very practices that are involved in the explication of a mathematical object (Troelstra, 1991; Ferreriós, 2016). Constructivism emerges in mathematics in the intutitionism of Brouwer (1907/1981). It can be understood as a reaction to the introduction of Cantor's ideas about set theory (Troelstra, 1991). In brief, the early forms of constructivism maintained that mathematical proofs should provide a pathway to the existence of the object that was at stake in the proof. It was not enough just to assert the existence of the object as proof as Cantor had done for uncountable infinities, for Brouwer (Troelstra, 1991; Bauer, 2017). Any proof in the early forms of constructivism should lead back to the natural numbers, and a finite number of steps should lead to decidability involving the natural numbers, in the early formulations of constructivist mathematics (Troelstra, 1991).

The emergence of constructivism at the turn of the 20th century was met with suspicion by classical foundationalists working at that time such as Hilbert (see Ferreiros, 2008; Mandelkern, 1985). Foundationalists took the three laws of thought described in this present thesis in Chapter 2 as foundational to all knowledge. Kleene (1952, p. 46) notes that in 1908 Brouwer publishes a paper: "*The untrustworthiness of the principles of logic*", which took up the theme of challenging the validity of classical logic. Typically, proof in classical logic works by relying heavily on the law of the excluded middle from the laws of thought. Statements must be either true, or false, under this principle, and nothing else. In order that to prove something is true as a mathematical statement, one can suppose that it is false, show that holding the statement as false produces a contradiction, and thus conclude that the statement must hold. This is known as proof by contradiction (Hardy, 1940/1992). Constructivist proof, on the other hand, is silent about the law of the excluded middle, it neither accepts or denies this law, and does not therefore utilise proof by contradiction (Bauer, 2017; Maddy, 2012).

While relaxing requirements to reference the law of the excluded middle could be considered as a move towards generalisation in terms of remitting one of the logical laws and allowing for a broader array of phenomena to be considered, the conditions for existence specified under constructivism are also logically weaker, with fewer rather than more inferences accepted as valid. This means it is a more restrictive logical structure than classical logic, demanding greater clarification of the context for mathematical proof, in order to connect the proof to truth (Brouwer, 1923/1967). For Brouwer, notions of universal and eternal absolutes were no longer valid, and in the words of Kleene (1952, p. 49):

> Brouwer opened our eyes and made us see how far classical mathematics, nourished by a belief in the 'absolute' that transcends all human possibilities of realization, goes beyond such statements as can claim real meaning and truth founded on evidence.

What this meant in the history of mathematics was a fundamental transformation or revolution, in mathematical practice (see Dauben, 1992; Shanin, 1997). It is noted in the literature that the birth of the constructivist paradigm for mathematics irretrievably altered what mathematics was taken to be, in a way where present practices cannot be construed as being on an ontological continuum containing past practices (see Noble & de Castro, 2017). In constructivism, emphasis is maintained not on truth in the abstract, but rather the procedures that support verification of the statement. Whether the verification is of a statement as a true statement, or verification of a false statement, it is not the case in the constructivist paradigm that verification of the true automatically prescribes verification of the false, as is the case under classical logic with the law of the excluded middle (Avigad, 2000). Rather, what is demanded in constructivism is contextualised verification, as constrained by the mathematical object or situation in question. What this means is that evidence must be accumulated for the mathematical object that is directly relevant to the object, or, constrained by, the object's existence.

Recognition of verification of the context of practice in a series of constructions for specific mathematical objects is clear in Brouwer's (1923/1969) formulation of mathematics as founded in perceptions of mathematician's minds (Grattan-Guinness, 1982). Endemic to this view of mathematical practice is recognition of the role for the present researcher, but also a role for the community of practitioners who have had historical roles in shaping what is taken to be of import in present-day practices, or who may be involved in assessment and evaluation of current outputs from mathematics. In this way, the community of practitioners take a vital role in the formulation of new knowledge for a field (Ferreiros, 2016). In circumstances where deductive certainty under foundationalist ideals for proof is not available, the community of practitioners plays an essential role in evaluating or verifying a contribution in the field of mathematical knowledge.

When considering the interplay of practices and theory in contexts in the field of mathematics, Ferreriós (2016) proposes a hypothetical conceptualisation of mathematical practice as most relevant to what is done when creating advances in the field. This is an interesting proposal, since mathematics is the home of deductive certainty in abstract thought - not hypothetical, but always true, or real. It is a combination of freedom and necessity in the form a hypothetical structure that drives an attentiveness to constructs in mathematical theory and practice, following Weyl (1951, pp. 538-539):

> The constructs of the mathematical mind are at the same time free and necessary. The individual mathematician feels free to define his notions and set up his axioms as he pleases. But the question is, will he get his fellow mathematicians interested in the constructs of his imagination. We cannot help the feeling that certain mathematical structures which have evolved through the combined efforts of the mathematical community bear the stamp of a necessity not affected by the accidents of their historical birth. Everybody who looks at the spectacle of modern algebra will be struck by this complementarity of freedom and necessity.

Ferreriós (2016) notes that the necessity here is one that is formulated in an intersubjective network - mathematics having the strongest form of intersubjectivity perhaps known, in science, as a universal language for sciences (Grattan-Guinness, 2008). A set of axioms is formulated for a specific phenomenon in mathematics, contextualising the eventual proof to be situated within the series of axioms. To arrive at the conclusion for the proof, however, one must accept and follow the axioms. There is no other way to the proof, but via the context elaborated in the representations formulated, in the axioms.

The hypothetical formulation of mathematical knowledge then makes a pathway via an accepted set of axioms in a community of practitioners, to what is considered admissible in a proof, which is also accepted by a community of practitioners. Even where what is taken as certain in mathematical knowledge changes through time, what does not change is the necessity of following the relations between axioms, to derive a proof outcome. Strong intersubjectivity secures objectivity for a proof, but what cannot be avoided is the context of the network of relations established among the axioms, which are the ground on which the strong intersubjectivity or objectivity is assured (Ferreriós, 2016, p. 175; Wagner, 2017).

All of this is important with respect to the use of the GLVM in psychological research, because as seen in previous chapters, deductive closure under strict true/false logical determination does not exist for the GLVM model. Several practices are accepted as providing quality assurance or control, regarding inferences made from use of the GLVM (see chapter 4). Verification of truth claims occurs via reliability and validity testing, amongst other protocols. There is however currently a question presented in the psychometric literature itself regarding the veracity of these assurance processes (Maul, 2017; Davis et al. 2018; Vazire, 2018). Research communities in psychological sciences are stating that more information is needed about the conduct of research, in support of claims made following completion of research projects (Fried & Flake, 2018; Nelson, Simmons, & Simonsohn, 2018; Vazire, 2018). With question about the inferences made following GLVM practices and the repeated call for transparency in view, the role for the community of practitioners in assessment of claims made from the outcomes of empirical practices in psychology research perhaps has never been more indispensable for psychology research.

The vital role played by a conceptual framework that makes apparent the linkages between different research project elements in a coherent representation of the whole of the research project is clear, in this regard. The conceptual framework becomes a communication tool for verification of claims about the research concept that is of interest – the psychological phenomenon at stake - in the project. In this capacity, there is a secondary role for mathematical constructivism with respect to the conceptual framework. While deductive closure is not possible for the GLVM, tracing the pathway entailed by mathematical constructivist practices stands to reveal the means by which the hierarchy of models formulated under the conceptual framework for a research project may act as a series of interrelated constraints or "links that restrict the admissible" (Ferreriós, 2016, p. 247). Here what is important is the consistency maintained within and between in these element-models, in representations pertinent to the research project. The logical test of consistency can be drawn into an evaluation of a minimal path in the relations between the represented project elements. The limits of human cognition as relative to logical consistency has been pursued in the philosophy of mathematics (see Maddy, 1997; Ferreriós, 2016), and will be further explored in the section on Rescher below (see Wagner, 2017). Questions of consistency pertain more closely to logical rather than to the kinds of meta-mathematical considerations that are at play in describing mathematical constructivism, and to these we will now turn our attention.

2.2 Consistency

As highlighted in the previous chapter, one guiding element for the conceptual framework is distinction between consistent and inconsistent terms and relations, when it comes to verifying claims that follow from analyses conducted via use of the GLVM in a psychological research project. From a mathematical perspective, Harsanyi (1983) states that if a theory is consistent, its mathematical existence is assured. In terms of the logic of constructivism then, consistency plays an important role for the structures pertaining a mathematical object (Ferreiros, 2016). Inconsistency itself has a substantial role in the development of set theory in mathematics. As presented in Chapter 5, Dauben (1992) notes that Cantor's innovation represents an innovation in thinking logically about numbers, in terms of utilising the concept of consistency. Cantor's work in 1874 distinguished denumerable from non-denumerable sets using a distinction grounded in the consistency/inconsistency polarity.

Brouwer's advance on Cantor's set theoretical proposal is noticing the distinction that remains however, between consistency and existence:

> Now the following question arises: suppose we have proved by some method, without thinking of mathematical interpretations, that the logical system, built up out of certain linguistic axioms, is consistent, i.e. that two contradictory theorems can occur at no stage of development of the system; suppose further that afterwards we find a mathematical interpretation of the axioms (which of cause will require the construction of a mathematical system whose elements satisfy certain given mathematical relations); does it follow from the consistency of the logical system that such a mathematical system exists? Such a conclusion has never been proved by axiomaticians, not even for the case where the given conditions involve that it is a mathematically constructible system that is required. Thus, for instance, it has nowhere been proved that a finite number, subjected to a provably consistent system of contradictions, must always exist. (Brouwer (1907/1981, p. 141)

While consistency may be presented for a logical mathematical system or even a conceptual framework for substantive psychology research, it does not guarantee the existence of the objects that are of interest, in the conduct of the reasoning that utilises, consistency. This is where reliance on local realism grounded in the geo-historical context of the research situation comes into play; consistency provides a set of constraints within which some result may be formulated, but the geo-historical situation of the research project provides a way to connect those constraints, to reality in facts. Verification then can take place, in virtue of the constructions pursued.

2.3 Verification and constructivism

Classical logic was presented in Chapter 2 of this thesis as consisting in three principles, these being: identity, non-contradiction, and excluded middle. A mentioned above, in producing classical proof, mathematicians relied on a relationship between truth and falsity to demonstrate contradictions when an axiom or principle was assumed as false. If a contradiction occurred when the axiom or principle was assumed false, then the axiom or principle must be true, under constructivist mathematics (Bauer, 2017). One advance pursued under constructivist mathematics is relaxing the requirement for the law of excluded middle to hold, in mathematical proofs (Bauer, 2017). Constructivist mathematics does not accept or deny the law of excluded middle; it just remains silent about it, in a way consistent with much modern mathematics where only identity and the law of non-contradiction are maintained as foundational to mathematical work (Maddy, 2012).

Under a constructivist approach, then, it is not the case that it can be assumed that the condition of truth automatically prescribes the circumstances of falsification (Bauer, 2017). That truth proscribes falsity holds, under classical logic, but not under constructivist logic. Therefore, constructivist logic applies more constraints and tests, than classical logic, in order to reach, truth. Such constraints are formulated under verification conditions, for constructivist proofs in mathematics (Kapsner, 2014). More statements are demanded which contextualise the mathematical object in question.

One important aspect of this for the conceptual framework is the way that truth and falsity come apart when reflecting on the GLVM, because it is a constructed model, with layers of principles, constraints, and assumptions, which feature in its structure, as seen in

Chapter 6 (see Maraun, 2017). However, the GLVM involves three key conditions that we have already described which inextricably take the model a step away from the world. These are the constraint of conditional independence, the problem of factor indeterminacy, and the incidence of circularity. These three key features of the GLVM demand some record keeping for the representation of the psychological phenomenon reflected in the latent variable, beyond the model representations. Such records formulate the steps of verification for any claims made following analysis utilising the GLVM. Markus (2013, p. 810) noted that one problem that existed for the correspondence account of psychological science is that it can encourage an idea of building a theory of truth that locates some truth-maker in the world, without reference to the network within which the theory, the cognition-independent phenomenon, and the justification, reside. An idea that a claim for truth gives you automatically the grounds for falsification, or the reliance on binary logic pursued by Haig and Borsboom (2012), leaves researchers potentially less likely to consider alternative models and theories according to Markus (2013), and is likely to result in under-representation of the true level of complexity of situations. By grounding representation with the geo-historical research situation and then making use of domain links between elements and explicit declaration of omissions and negations, both laws of identity and non-contradiction still do important work in the production of scientific knowledge. However, the process of verification takes place with reference points made clear in a way that is not possible utilising negation under the law of the excluded middle.

Verification is not merely of interest to the researcher in question; more important is its function in communication with the scientific community. The community and the context both play a vital role in discernment of the connection, then, between meaning, for the model and the representations of phenomena, and truth. In a way, here the relation between community and context can be understood as re-connecting that which was broken apart by Frege in the innovation of the argument-function relation. This innovation had allowed for a de-contextualisation of a syntactic structure – the argument-function relation, as stated in earlier chapters had portability, across different fields and paradigms. However, late 20th century advances in mathematics and logic suggest that in applications, meaning and truth are inextricable from each other. As Brady and Rush (2009, p. 492) state:

> There is a close connection, perhaps to the extent of inextricability, between meaning and truth, such that it may even be the case that both (minimally) are needed to express not only the nature of deduction, but also truth itself. And even if we were to suppose that knowing the meaning of p necessarily involves knowing when p is true, this does not entail that knowing the truth conditions for p gives us its meaning.

Brady and Rush (2009) go on to explore the degree to which meaning may be located by connection with the real world, as well as located within a semantic or modal system, which has no direct connection with the real world. The question of situatedness or context is an important one for the use of the GLVM in research project analysis, as we have stated that all research projects occur in situations. Contextualisation of a research project lends constructability to the conceptual framework apparatus. This is important for the GLVM, because the problem of factor indeterminacy and its resultant infinity of solutions for the latent variable, as well as the application of statistical analyses to data take researchers into the realm of consideration of uninstantiated properties, considering real-world circumstances.

2.3.1 Uninstantiated properties and verification

What was noted in Chapter 4 with respect to statistical analysis of psychological phenomena in inferential processes is that articulating something about psychological phenomena may involve consideration of uninstantiated properties (see also Luce, Krantz, Suppes, & Tversky, 1990). This may include, for example, where in considering a range of possible scores on a latent variable, it may be that there are no manifest variable value combinations exist that would provide the correlational evidence for the latent variable scale structure that is proposed. As a shift occurs from real-world facts, such as data points gathered on tests completed by individuals, to possible-world uninstantiated properties, such as the scores that notionally an individual could have had, researchers become more reliant on structures of inferential reasoning. There is a necessary shift in this circumstance to correspondence truth, about real-world states of affairs, to coherence truths, or alignment of statements with the structure of inductive argument. Statistical inferences in their generality extend beyond the explicit information given in data, and always involve some subjective decision-making from the researcher in this respect (Abelson, 1995) about how to deal with for example, uninstantiated properties. Some quality assurance process is thus needed, regarding claims that are made following the conduct of statistical analysis, such as those covered in Chapter 4 connected to the reliability and validity of quantitative conclusions.

What this means is when reporting research outcomes, fellow researchers need to be able to see into the sets of assumptions, principles, and instantiated and uninstantiated properties that are relevant to the conclusions expressed for the research project. Uninstantiated properties may arise for values within manifest variable ranges; they also play an important role in the limits that are used to construct statistical knowledge. For example, infinity is a term that meaningfully participates in assumptions about inferences from latent variable modelling practices that rely on normal distributions of variables. Infinity however remains always as characterised best by its uninstantiated properties, whether we consider countable or uncountable infinities, following Cantor (1883). It is not a meaningless term, even though we struggle to make the term precise or exact, for the purposes of inferential logic. We have also above described the way that ranges of values for variables may remain uninstantiated as properties in terms of parameter outcomes found from real-world data, and its analysis. Meaninglessness can be understood as occurring "only in the case of a term occurring outside the scope of any predicate" (Brady & Rush, 2009, p. 499). In respect of such an occurrence, the authors note that no further generalisation is suitable for any term occurring outside of the scope of declared predicates. In respect of the conceptual framework for the GLVM, meaninglessness may be considered to apply for example where a solution proposed for a latent variable falls outside of a pre-specified domain for an element that we have articulated, in the conceptual framework.

2.4 Meaningfulness - assessing elements and quality assurance

Meaningfulness had been designated as a third value in many-valued logic systems described by Suppes (1959), Narens (2002), and Rescher (1968). In simple form, this meant that determination could be made regarding a status of true/false/meaningless, for any statement. Analysis of the meaning of statements, and their logical import had followed in the wake of the development of logical calculus seen in Boole and Frege, and was furthered in semantic considerations and language-based analysis, following the logicism of Russell and Whitehead, through to the metalogic of logicians and mathematicians such as Tarski and Turing (see Kneale & Kneale, 1962). Brady and Rush (2009) demonstrate, as per the quotation above, that meaning and logic however cannot easily be disentangled, and that logical determination must be founded via the meaning of the statements. Suppes (1959) explored meaning and meaningfulness in connection to statements of measurement. This 1959 exploration instituted the beginning of Suppes's efforts to build a model-based account of science that tracked invariances in nature, beginning with the physical sciences (see also McKinsey & Suppes, 1955).

As described in Chapter 2 of the present thesis, historically the condition of invariance was associated with the operation of natural law in science. This meant that natural laws were unchanging and eternal, and it seems this is one way to characterise what Spearman was searching for, in his conceptualisation of *g* or cognitive ability as an objective measurement (see Spearman, 1904). Suppes (1959, p. 134) notes that the findings in physical sciences through the 20th century led to the unavoidable conclusion that objectivity in terms of science was garnered in a relative network (see Weyl, 1922). For mathematics, invariance was found in the various number systems, and Suppes (1959) notes that much of mathematics takes place without declaration of its relative network. Constructivist mathematics invites practitioners not to take constants for granted, but to contextualise them without relying on the logic of the law of the excluded middle to do all the heavy lifting, providing adequate grounds for verification of the statements, in consistent structures.

For psychology research that makes use of the GLVM, analysis inevitably takes place in empirical research situations. Recently, the role of measurement has been re-cast as one of meaningful representation of reproducible findings, effects that are presumably, invariant (Cano, Pendrill, Melin, & Fisher Jr, 2019; Stevens, 1951; Mundy, 1987). Given the gaps that are present for the GLVM in terms of its basic structure with reference to the constraint of conditional independence and the problems of factor indeterminacy and circularity, reproducibility is enhanced in research projects to the degree that detailed representation is made of the features and factors that were relevant to the conduct of the project. Researchers can perform a self-audit for quality assurance regarding the ordered relations in elements and between elements, in a conceptual framework as is briefly described at the end of the chapter. The overall assessment a researcher should evaluate, and one that the community of researchers who receive, read, and make use of a researcher's findings should also be invested in is an assessment of the overall consistency and systematicity of the research project elements, in light of the realist foundations in the project situation. One theorist who addresses consistency and systematisation is Nicholas Rescher, and to his work, we now turn.

3. Nicholas Rescher

Rescher is a German-born American philosopher working in mathematical logic and analytic philosophy, instituting what is known as methodological pragmatism (Rescher, 1977; Jacquette, 2009), which maintains focus on the development of technical criteria in aid of evaluating knowledge claims. Rescher's body of work is extends across six decades, but one topic that repeats over those years is the investigation of the relevance of logic for and the application of logic in the empirical sciences (see Rescher 1968, 1995, 2005b). He is responsible for innovating several distinct approaches to the use of logic in applied contexts, based on pragmatic grounds (see Rescher, 1969). There remains throughout his body of work an intention to address science as a practice shared amongst practitioners, seeking to cohere elements from the history of Western intellectual history in ways that facilitates adoption of what he describes as objective pragmatism (Rescher, 1997). Objective pragmatism is distinct from methodological pragmatism in that it recognises while there is no ontological distinction between the mind-independent world and the social world within which science as a practice is situated, we are well served to the degree that we recognise that science *is* a socially situated practice (Rescher, 1997, 2005a). Science as a human practice remains always fallible for Rescher (see 1979, 2005b), and our interaction with reality, particularly in situations of scientific research projects, remains always mediated by social and cognitive constraints. Any claim for objectivity of knowledge then must demonstrate both the contextualised social and cognitive constraints, and how they are overcome in order to create objective knowledge.

Importantly, Rescher adopts a coherentist theory of truth for objective knowledge (see Rescher, 1979), which as we will see in what follows, is distinguished from a foundationalist perspective, in ways resonant with the distinction made in mathematics between foundationalism and constructivism that we have discussed above. Consistent with constructivism, Rescher's coherentist account of truth maintains attention to the set of circumstances relevant to the situation within which truth is given to arise. Rather than setting a metaphysical given for what truth 'is', Rescher (1979) takes these circumstances into account, we shall see, in his account of scientific truth, in a way that is informed by his theorems on consistency, which we will now examine.

3.1 Rescher and consistency

Rescher and Manor (1970) note that whenever we are making conclusions based on probabilistic inferences, we are already reasoning from what are likely[23] to be inconsistent premises. Rescher (1979, pp. 165-166) defines inconsistency as present for a body of knowledge when "it lacks consonance, internal regularity, and self-concordance". One aspect of inconsistency that is important to note in respect of the history of logic is that under classical logic, where inconsistency is present in a series of premises or statements, then contradictions are present amongst the statements such that the principle of explosion applies. This is where any proposition whatsoever is given to logically follow (Jeffrey, 1981; Carnielli, Coniglio, & Marcos, 2001). Brandom and Rescher (1979) distinguish between contexts where statements A and not-A are true for some A, and A and not-A being rendered true by explosion for all A. Contextualisation facilitates the recognition of the distinction between A and not-A being true together collectively, and A and not-A being true together *distributively*. Deductive proof works with systematisation of a set of

[23] To be truly probabilistic is to also acknowledge the possibility that our processes can actually and completely randomly produce consistent relations and premisses for our conclusions.

premises or statements collectively, whereas a distributive approach facilitates consideration of the elements of a system in isolation, one-by-one (Brandom & Rescher, 1979). These authors go on to develop their account of inconsistency as evaluated over semantic formulations of closed deductive systems under conjunction. The conceptual framework designed in this thesis is as already stated, not a closed deductive system. It is proposed that what will best serve the community of research practitioners and indeed the broader community who make use of the findings of psychological research is an evaluation of the conceptual framework, specifically, its consistency. Such assessments are recorded for the quality control element introduced in Chapter 6. Ultimately, in the future, with the framework made amenable to query via software with entry of project details into digitised format, calculation of proportion of constraints applied across all elements included in the research project may serve as an index of consistency via constraint for a project. In the meantime, we rely on comparisons between researcher self-audit, and assessment from the scientific community who have an interest in the GLVM analysis from the research project, to provide an indication of the overall consistency attributable to the conceptual framework for a project making use of the GLVM.

In connection to the GLVM, inconsistency is a feature of analyses because of the three concerns that are present for every GLVM analysis as highlighted in Chapter 4. These include the assumption of conditional independence intrinsic to every use of the model, the logical concerns regarding circularity of argument, and the mathematical concerns of factor indeterminacy. Each of these elements alone induces the possible feature of inconsistency in the structure of our conclusions, specifically because internal regularity of our inferences cannot be assured, under the assumptions. An assessment of quality for each research element with a view to verification of the quality by a community of fellow practitioners supports then inferences made following analysis utilising the GLVM, in psychology contexts.

Consistency is a crucial concept not just because of the three key problems for the GLVM discussed above. Inconsistency will also be relevant to us for another reason. As we seek to understand the series of constraints that are at play in the representation of our research project that is exemplified in the conceptual framework, it may well be that there are mutually incompatible specifications for the elements of the project, that even though they are incompatible, still form part of the whole representation of the project. One example already discussed in this thesis is where a researcher may adopt the conceptual framework informed by a realist constructivist perspective for representation of a research situation, but also wish to take a social constructionist view of the meaning of the items, for research participants. Social constructionism would imply that the meaning of the items for the research participants is completely uniquely determined for every participant-researcher relationship represented in our subject pool. A realist constructivist position on the other hand takes the context of the research situation into account in such a way where the meaning of the items can be generalised or understood as shared, for all participants. Interpretation of GLVM analysis in each respect would be different, no matter the actual content of the items themselves, but also each ideological position, relevant at different points in the project, may also remain inconsistent. There is nothing that actually forces a researcher to close-out on a specific claim under one interpretation or the other. Further, it may be that the researcher prefers actually not to limit the possibility of interpretation to one ideological position or the other, particularly where for example future studies may be taken up that make use of the present data, outcomes, findings, or methodology. Methodological pluralism is valuable in this regard, as it maintains focus on technical constraints relevant to the local situation of the research project, articulated in a systems view (Rescher, 1977). Interpretation of research outputs remains guided by the structures or relations that are evidenced in the conceptual framework, while local systematicity acts as a guide regarding technical consistency.[24] We have mentioned systematisation a number of times, and in the next section, we will define and introduce the concept, before then exploring its role in inferential processes.

Before we go on to the next subsection, a brief recap. It has been said that mathematical constructivism provides several tools for the conceptual framework. Firstly, it situates the context for an inference or claim to knowledge with the inference or claim to knowledge - the context itself is understood to have a vital role in asserting the inference or knowledge as a contribution in the field to which it belongs. Secondly, we can trace the distinctions between finite/infinite domains following Cantor's thought and acknowledge that constructivism asks us to specifically account for how we see the relation between the finite and infinite, when we think about our research elements (in mathematical constructivism, the research element is a mathematical object). Thirdly, constructivism references not merely thoughts in an isolated researcher's mind as per Borsboom (2005), but makes use of a series of constraints between collections of statements to articulate something about a mathematical object which must be accepted as working principles within a community, to arrive at the conclusion engendered by the mathematical theorem or proof. In this way, the relationship between the community of practice and the researcher must be acknowledged. Fourthly, we have said there is a role for consistency in helping us to make judgements about the way that a body of statements function together. Because our conceptual framework addresses empirical situations and includes several assumptions that are necessary to utilise the GLVM, which are impossible to overcome with any sense of deductive closure in model representation form, inconsistency can be expected in our account of the research project in the framework. What is aimed for is to

[24] Note that there are a number of many-valued logics which could beneficially be integrated into a fully automated conceptual framework structure which make use of such evaluative evidence as plausbility ratings in a way similar to but distinct from Bayesian inference - see Rescher and Manor (1970), Carnielli and Malinowski (2018) for examples of such reasoning processes.

maximise systematicity and reduce to the greatest extent possible the quality uncertainty for reported research outcomes. Systematicity for the framework has yet to be addressed, and it is to this we now turn.

4. Towards systematisation

The value of systematisation for knowledge is elucidated perhaps nowhere more clearly than in Kant's Critique of Pure Reason:

> In accordance with reason's legislative prescriptions, our diverse modes of knowledge must not be permitted to be a mere rhapsody, but must form a system. Only so they can further the essential ends of reason. By a system I understand the unity of the manifold modes of knowledge under one idea. This idea is the concept, which determines *a priori* not only the scope of its manifold content, but also the positions which the parts occupy relatively to one another. The scientific concept of reason contains, therefore, the end and the form of that whole which is congruent with this requirement. The unity of the end to which all the parts relate and in the idea of which they all stand in relation to one another, makes it possible for us to determine from our knowledge of the other parts whether any part be missing, and to prevent any arbitrary addition, or in respect of its completeness [to discover] any indeterminateness that does not conform to the limits which are thus determined *a priori*. The whole is thus an organised unity and not an aggregate. It may grow from within, but not by external addition. It is thus like an animal body. . . . (translation via Kemp Smith, A833 = B861)

Rescher (1979, p. 14) defines a system this way:

> A system is a collection of interrelated entities, the relationships among which are such that information about them affords a basis for inferring conclusions regarding the structure, modus operandi, or temporal history of the system as a whole.

In the book *Cognitive Systematisation: A Systems-Theory Approach to a Coherentist Theory of Knowledge*, Rescher (1979, p. 4) advocates for a view of systematicity as an ideal for knowledge which integrates it into an organised and rationally co-ordinated whole. This is informed by the ideals of Kant as exemplified in the quote above, but not limited by the transcendental idealism that pervades Kant's approach. Transcendental idealism limits the construction of knowledge to universal concepts, which inhabit a single human mind (see more comment on of Kant in this respect in the section on Hegel, below). In tracing the history of development of the idea of systematisation, Rescher (1979, p. 9) notes that Kant's role in launching an explicit theory of cognitive or knowledge organising systems. This approach unites both material systems and intellectual systems under a single metaphysical formulation, with no distinction made between them (Rescher, 1977, 1979). Rescher absorbs the ideas of cognitive systematisation but introduces a modern approach to pragmatism informed by realism in respect of truth. This he describes as pragmatic idealism, where the validity of a claim for truth can be assessed with reference to the coherence of the claim. An important aspect of the assessment of coherence however is that it acknowledges that the social-linguistic structures that provide the means for an assessment of coherence themselves are derived from the contact we have as human beings, with reality. In this way truth is always traceable back, to correspondence with reality (Marsonet, 1994). For the conceptual framework described in this thesis, systematisation from the representation of the geo-historical situation of the research project provides a similar through-line, to the facts of the context of the research study.

Rescher traces the origins of the idea of the systematicity of knowledge to Aristotle in *Posterior Analytics*, specifically to Aristotle's inquiry into the nature of truth, noting that truth is always accounted for as contextualised, in a system, from the earliest of Western intellectual history (see also Hoyningen-Huene, 2013). What is important is not only that a system arrives at a determination of truth; key is the role that systematisation plays, in arrival at a determination of truth. Rescher (1979) adopts a distributive or process view of truth determination, looking at the elements that make up a body of knowledge and their interlinkages, as primary. Such an attitude to the decomposition of a system is important when considering the conceptual framework for the use of the GLVM in psychology research. This is because it is already the case in psychology research practices that fellow researchers may extract a single component of a previous study for use or replication in their own study. Each element in a conceptual framework for GLVM analyses therefore should be able to be considered as its own unique ordered system in the larger system of statements that make up the representation of the research project.

Rescher (1979) suggests that systematicity is more than just a handy strategy in the formalisation or organisation of knowledge; he proposes that systematisation in fact plays a role as a criterion of knowledge. This Rescher states is introduced into Western intellectual history by Hegel, in his critique of Kant. The Greek syllogistic relation between reality and knowledge in syllogistic form is this (Rescher, 1979, p. 36):

Reality is a coherent system

Knowledge agrees with reality

Knowledge is a coherent system.

As described by Rescher, Kant inverts this syllogism, so that rather than basing a conclusion about knowledge on premises regarding reality, he infers a conclusion about reality from premises about knowledge:

Knowledge is a coherent system

Knowledge agrees with (empirical) reality

Empirical reality is a coherent system.

Rescher traces Hegel's own retracing of Greek thought, where Hegel asked - how do we really know that knowledge itself forms a consistent system? The Hegelian inversion finds consistency for a new thesis or contribution in the body of scientific knowledge when it comports or is consistent with the already existing body of knowledge. This body of knowledge, which is taken in the broader scientific community to be true knowledge – is knowledge that is understood to characterise reality, itself. In Hegel's developments, then, systematicity does not just order what is known. What is also needed is demonstration of consistent links between old knowledge, and new knowledge. The formulation or systematicity of a collection of elements has an important role in the process of truth determination. Systematisation then becomes an arbiter or a qualifying test, for the acceptability of claims of fact (Rescher, 1979, p. 38). It is only with systematisation, that we can truly say that we have new knowledge. An important inclusion in the conceptual framework then is evidence of systematic link, between previous research and the present project, relations that can be clarified in the variables sub-elements, for theory, model, variables, and even data elements.

A cognitive system for Rescher (2016) is one that addresses an issue for knowledge in a comprehensive way. Cognitive systematisation is characterised by taking the whole domain of relevant information into account. Coherence, consistency, and harmony between relations and elements of the system are features that are important, here. Rescher (2016) states that systematisation as a principle provides a set of norms for inductive acceptability, in terms of the cohesion between elements, for a single research project, and in the way the research project representation fits, in the larger body of scientific knowledge that makes up the field, in our case, of psychology. When researchers conduct self-audit in terms of estimated consistency as links between domains for their research project representation and add this to the quality control element, they are providing information about the perceived cognitive interpretability, of what is included in the representation, of the project.

4.1 Cognitive systematisation – contra foundationalism

So, what would a cognitive system look like, in Rescher's account? Rescher counterposes foundationalist approaches to systematisation such as that exemplified in the step-by-step axioms of Euclidean geometry, to a network model of scientific knowledge (Rescher, 1979, p. 41). The linear chain of successive tiers of foundationalist systems give the means for justification of knowledge in the linear structure. As Rescher (1979, p. 43) points out, the Euclidean model of cognitive systematisation cannot be underestimated for its impact on the whole of the development of intellectual history in the West. But it is not the only possible model. For analyses that involve probabilistic inferences such as those that follow from the use of the GLVM, an inferential interlacing network model of cognitive systematisation is appropriate, in that the linkages that are of local import for a particular research element can be interwoven into a full account of the empirical situation that is of interest.

Important implications of a network model for the conceptual framework for the GLVM are the following. Firstly, what is made possible is determination of local features of the representations in the system; where consistencies can be evaluated relevant to the element and the near neighbours of the element affected by the consistency. A problem with foundationalism is that one problem in the system of axioms that makes up the system confounds the whole of the system. In a network approach, the elements can be treated individually without necessary disruption to other aspects of the relational structure. Therefore, the explosion that we described above which would for example follow a contradiction or inconsistency in a foundationalist system does not occur in a network system.

Secondly, similar to the way that a set is defined by its elements; in a network system there is no formal structure into which the system is forced - the system is defined by its elements, and the relations between the elements define the system in question. The opportunity to connect elements in the conceptual framework developed in this thesis has already been elaborated in Chapter 6. One other key structure for the conceptual framework is the ordinal structure within an element, and the capacity to connect these elements to any other elements by making use of this ordinal structure. The actual relations between elements are free to vary, within constraints that are set by the situation of the research project itself.

The theory of expected relations that then takes form in this network system seeks to co-ordinate the facts that are of concern for the research project situation and the application of mathematics and statistics, in the use of the GLVM in psychological research. It seeks to co-ordinate them in a way that makes clear the feedback loops which bear representation, and which are excluded under a linear foundationalist axiomatised approach (Rescher, 1979, p. 49). Foundationalism makes use of the efficiency of self-evidence contained in a set of systematised axioms, but this is purchased at the expense of attentiveness to the real-world phenomena itself, and the kinds of structures or processes that may be relevant to actual scientific reasoning. A network system aims to include the apparatus to address the intricacies of complexity in reasoning and in the real world. Rescher (1979, p. 63) points to the failures in axiomatised accounts of completeness for systems that followed in the wake of the work of Kurt Goedel, which forever challenged the possibility of production of an account of mathematical practice that was consistent and complete. What we take as of import for our conceptual framework is the notion that systematicity itself is ground for acceptability of a representation for a research project. In the words of Rescher (1979, p. 80): "[t]he principles of systematicity now represent presumptive principles regulatively governing the conduct of inquiry."

What this means is that selection principles should routinely privilege systematicity amongst elements that are presented in the conceptual framework. As an example of how this is relevant – in the element devoted to the researcher as seen in section 5.1.2 of Chapter 6, in the variables sub-element a researcher may choose to declare in their philosophical stance with respect to values such as simplicity, or complexity, or plausibility, or uniformity. Where the researcher self-audits subsequent research elements, or the research community evaluates the researcher's findings considering their conceptual framework, evaluation should be conducted consistent with the declared value(s). In this regard as well, element specification can be used to articulate roles for the kinds of processes taken to ensure research veracity in psychometric paradigms, including for example, reliability and validity assessments. In this way, support for the inferences made from a GLVM analysis is structurally supported by articulation in the conceptual framework of the relations by which reliability and validity tests are considered by the researcher to have a role, in the research project that is at stake. Rescher (1979, p. 102) describes a necessary relation between theoretical coherence and pragmatic legitimation, which is grounded in the situational relevance or application, of the system in a mutually determinative domain. In balancing the realism as correspondence in the geo-historical situation of the research project with constructivism for the GLVM as a model guided by principles of cognitive systematisation, the conceptual framework articulated in this thesis aims for a similar mutual determination.

4.2 Ontological systematicity and cognitive systematicity

This sense of a mutually determinative domain grounds what Rescher (1979, p. 124) describes as "ontological systematicity", such that "[i]f the world were not orderly . . . then there would be no uniformity in information-gathering... and consequently, there would be no avenue to the acquisition of knowledge of the world" (p. 121). Rescher goes on to say that a level of ontological systematicity in the world must persist as a causal precondition for the success of a systematising inquiry. Cognitive systematicity in no way serves as deductive evidence for ontological systematicity. However, Rescher argues that cognitive systematicity provides the means for inductive reasoning about ontological systematicity. In addition, for Rescher, "[c]onceptual systematicity affords the prime – perhaps the sole – entryway through which evidence of ontological systematicity can be secured" (p. 126).

What this means for us in our conceptual framework for use with the GLVM is that careful clarification must be made regarding the relevant real-world domain over which both the research project has run, and the real-world domain over which results are expected to be generalised. When domains are linked, the framework takes shape as a hierarchy of models, where each lower-level element inherits the previous level element, as an originating domain. However, the relations are not limited to hierarchies, as seen in Chapter 6 section 5.2.7, where psychological phenomena may be linked not only to data, but also to the research concept. These principles apply in cognitive systematicity and serve as methodological legitimation for claims made from analysis under the GLVM. Cognitive systematicity is designed to serve our inferences in organising the structures of knowledge. In this way, project aims and researcher values will each have a role in formulating cognitive systematicity and will involve several relative judgements that will also be served by clarification in the conceptual framework. Integration of these relative grounds for judgement into the conceptual framework provides a means for assessment of overall plausibility for claims that result from the use of the GLVM in research contexts, and we will see how this takes place in Chapter 8. Meanwhile to learn something about the pattern of inference from systematisation, we will look now at Rescher's contrast between inference to the best explanation, and inference from the best systematisation.

5. Inference from best systematisation

5.1 Inference to best explanation - a distinction

In terms of cohesive approaches to making scientific inferences from psychometric research techniques such as the GLVM, inference to the best explanation has been proposed as an appropriate basis for a rationale, for inferences (see for example Haig, 2009). Inference to the best explanation (IBE) was originally proposed by C.S. Peirce and championed by Harman (1965) in the form recognised today in psychology literature. Rescher (2016) sets out the logical structure of IBE in this way:

F is an established fact.

$E_1, E_2, \ldots, E_n$ are possible explanations for *F*

Among these, E_1 is the best-available explanation

Therefore, E_1 is to be accepted, provisionally, as the case that affords the correct and actual explanation of *F*.

5.2 Inference to best explanation - concerns

Rescher (2016) notes that problems regarding inference to the best explanation persist to the degree that the desiderata that characterise 'best' may clash with one another for any given context or situation. For example, the best explanations, following Rescher's examples, may maximise:

a) The use of here-and-now availability of information;

b) Simplicity or elegance;

c) The probability of evidential premises;

d) The probability of the explanatory conclusion relative to the premises;

e) The extent to which the premises render the conclusion likely;

f) The definiteness of the conclusion;

g) The extent of evidence considered;

h) The breadth and scope of explanatory principles;

i) Similarity to other explanations.

It is possible to have two equally good explanations that remain inconsistent with each other, and to have no method to disentangle their worth. In GLVM analysis, this may be the case for example where two CFA models demonstrate comparable fit statistics. There is also no obvious universal way to account for the preferencing of maximising any of the explanatory criteria listed in a)-i), so there remains no general criterion from which we can say the 'best explanation' is chosen.

A systematic account of inferences on the other hand allows that researchers can make use of a rationale justified on the basis of methodical links between representations of research elements, no matter whether the project is invested in description, explanation or prediction of psychological phenomena. Future researchers also can investigate elements of the project with reference to any criteria with respect to the 'best explanation', free of an assumption about what constitutes the best explanation, per the initial project. For this, Rescher (2016) advocates for inference from the best systematisation over the best explanation. Criteria for best systematisation makes available the set of working assumptions that guided the inferences, and also works to place the current set of inferences in a broader historical context (e.g., what has come before in terms of bodies of knowledge about this particular psychological phenomena; what might best support future researchers). In the words of Rescher (2005b, p.49) "[s]ystematisation is a resource of cognitive validation that is significantly different from explanation" (p. 49). Systematisation asks that researchers think in a big picture or metatheoretical way about the whole of the research project, and how the GLVM analysis fits into it. A good systematisation can be judged on a single criterion – that of the consistency of the construction. An explanation in and of itself need not be true, nor even probably true, but where it is systematised, we are given the means to understand how it fits into a bigger picture or the facts-in-general that are pertinent to the situation (Rescher, 2016).

Rescher (2016) defines the basic pattern of reasoning for inferences from best systematisation this way:

F is an established fact.

S is the best-available systematisation of determinable facts relevant to *F*

S entails *X*

Therefore, *X* is the case.

Systematisation does all the heavy lifting for any further inferences, and can support any sort of preferred explanatory criteria, wherever the GLVM is used in explanatory capacity. In Rescher's (2016, p. 150) words systematisation "seeks to reduce to the greatest feasible extent the room for anomalies, discrepancies and loose ends". As stated above, the system for Rescher (1979, 2005b, 2016) becomes not just an organisational tool which structures what is already accepted, but also serves as an arbiter of what can be further accepted. In addition, it is made clear that the inferences that are made following use of GLVM in an analytical process follow from what has already accepted as true and as conditions contextualised in the research project space, rather than making inference to some external desired outcome that may invite opportunistic use of evidence. Systematisation may facilitate inclusion of processes of description and prediction as well as explanation in our research project schema. No disjunctions are implied between alternative explanations, either, and they may each continue to serve for some perspective or in some respect for a set of elements, in a project.

There is no guarantee that inference from the best systematisation arrives at truth. Rather, what is attempted in this articulation is the setting of optimal conditions under which it is possible to arrive at truth, utilising coherence and correspondence in ways that are appropriate to the research project context, and in ways that take shape from the research project context. In the words of Rescher (2016, p. 153):

> [A]ny thesis - explanatory ones included - that forms part of the optimal (available) systematization of the facts-in-general must for that very reason square with our best understanding of the overall situation. And on this ground alone we can plausibly view it - from the practical point of view - as endowed with a reasonable warrant for acceptance - for truth estimation.

Construction under maximal systematisation then yields a distinct set of advantages in respect of self-audit, for a researcher, and in evaluation, by the research community. For example: a) intelligibility can be scrutinised; b) scientific adequacy of any statements can be pursued via logical grounds; c) quality control gives a means of testing acceptability or correctness for any claims of fact; d) there is a foothold for a criterion of knowledge - a rational construction of what counts as systematised knowledge contribution, versus what does not.

5.3 Quality certainty assessment in a systematic conceptual framework

It was suggested above that a calculation can be conducted which evaluates for any single research project the proportion of elements that are linked in consistency via domain constraints. Such a calculation was recommended as an inclusion for the quality control element as the researcher performs a self-audit of the overall project representation that is made in the conceptual framework for projects that include GLVM analysis. Best systematisation follows when there is high consistency in the project representation as evidenced by the effects of the domain constraints.

It has been proposed already that in efforts directed at the reduction of quality uncertainty in psychology research outputs, there is benefit to supporting transparency following Vazire (2017), but also that quality uncertainty is not only founded in a lack of transparency, it is also in part a function of quality management practices (Wankhade & Dabade, 2006). Some techniques from the qualitative literature that target research credibility and trustworthiness of qualitative information in aid of quality management are discussed in Chapter 8. To the extent that these are incorporated into the conceptual framework, it would be expected that the quality assessment of the overall research project would increase. For the GLVM, in Chapter 4 several quality control techniques with respect to research outputs were discussed. These include reliability and validity assessment, as well as practices undertaken to manage invariance, unidimensionality, generalisability, and model-data fit (Thompson, 2004). In considering the representation of quality control in a conceptual framework for the GLVM, the domain, U, would typically be the research situation, while the formulation, A, beneficially would include a listing of the types of quality control engaged in, for the GLVM outcomes, with any deliberate exclusions listed in V, the variables sub-element. A researcher could beneficially consider how the techniques combine in a quality control effort, where for example, reliability and reproducibility concerns should be assessed prior to validity checks (Thompson, 2004). A final addition to the formulation, A, for quality control in reference to the GLVM may include a researcher's self-assessment, of the consistency of their framework. In this way, transparency regarding researcher decisions is not only support, transparency regarding what the researcher thinks about the cognitive consistency of their decisions, is also made available to the scientific community. Such information may be of value in evaluating research claims no matter whether the assessment takes place prior to, throughout, or following the conclusion of a research project, making use of the GLVM in psychological research.

6. Summary

Reference to constructivism in mathematics for use of the GLVM provides us with several indispensable resources for the conceptual framework, as described across the preceding sections. Specifically, the focus on existence of the specific mathematical object of interest in the construction of mathematical proof, and consistency of statements in respect of the object provides guidance about how psychology researchers can work towards verification of the phenomena of interest, in research making use of the GLVM. Because the GLVM precludes the possibility of achieving deductive closure given the assumption of conditional independence, the problem of factor indeterminacy and the logical limit of circularity, correspondence truth to states of affairs remains challenging to establish when using the GLVM, and coherence of statements becomes a more important consideration in verification, of claims. To this end the conceptual framework stands to provide evidence in support of the phenomena of interest, with demonstrated consistency calculable as the proportion of research elements presented in the conceptual framework that are interlinked via domains. The assessment, included in the quality control element of the conceptual framework, provides both the researcher with a self-audit, and may be scrutinised by the community of practitioners to whom the project in question is of interest and relevance. Realist constructivism for a conceptual framework for the use of the GLVM is situated in two respects – the geo-historical situation in which the research occurs, and transparently situated in a body of expert practitioners, who take a role in assessing systematicity of the account of the project and the contribution that the project makes to the field of psychology, as a whole.

Chapter 8: Applied Examples

1. Objectives

In this brief chapter, two worked examples of the conceptual framework are presented, for two different uses of the GLVM, where the conceptual framework is retrofitted to previously published studies. It should be noted that this is not the typical use for which the conceptual framework is built. The point of the conceptual framework is to complete it as one completes the pre-registration processes that are described in more detail in Chapter 9, as one undertakes a study that makes use of one of the models that fall under the rubric of the GLVM. Completion of such a conceptual framework for a particular study facilitates comparisons of applications of the GLVM model for specific psychological phenomena – something that is at present challenging to conduct.

In thinking through psychology research that makes use of the GLVM, the advantages of adopting perspectives informed by mathematical constructivist accounts of set theory may be observable in the following worked examples. As described in the previous chapter, under mathematical constructivism, there is recognition that one must clearly account for the construction of the mathematical object in question, in order to prove that the same object exists. While proof is not pursued nor is available to us with use of the GLVM, as described particularly in Chapter 4 of the present thesis, what we can acknowledge is the benefit of pursuing the steps toward proof that makes up the evidentiary process in constructivism in mathematics. The conceptual framework described in this thesis provides a means of evidencing the decisions made through the process of utilising the GLVM. As described in Chapter 4, and as seen in the examples below, the subjective decisions and choices made by researchers have substantial implications for the conclusions that can be made and the conclusions that are made, as a process of completing analysis following the gathering of empirical data. The first example looks at analyses of a neurocognitive battery for schizophrenia in the computational psychiatry literature, the second looks at the development of a short-form questionnaire for depression and anxiety in the patient-centred psychiatric care literature.

2. Applied example - computational psychiatry

Computational psychiatry is a newly emerging field with stated aims of bridging neuroscience and the more traditional methods of assessing psychological functioning and phenomena (Huys, Maia, & Frank, 2016). An example that makes use of the GLVM is presented in recent schizophrenia literature, '*A confirmatory factor analysis of the MATRICS consensus cognitive battery in severe mental illness'*, by Lo et al. (2016). This report in the example below serves as an example of retrofitting the conceptual framework. As will be seen, there are several gaps that thus present in the conceptual framework, where the details of researcher decisions are not included in the brief six-page report that appears as a journal article.

In respect of the report, The National Institute of Mental Health's Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) took as a goal the development of a standard neurocognitive battery to assess improvements or change in cognition for those experiencing this disorder (Nuechterlein et al., 2008; Grey et al., 2014). Following initial consultation with experts regarding appropriate tests from already established assessments, a number of tests were combined for the battery that aimed to maximize criteria for test-retest reliability, replicability, connection to functional outcomes, responsivity given pharmacological change, practicality, and tolerability (Kern, Green, Nuechterlein, & Deng, 2004). This is recognized as the MATRICS Consensus Cognitive Battery (MCCB: August, Kiwanuka, McMahon & Gold, 2012) and has subsequently been employed in mapping biomarkers for schizophrenia (Tregellas et al., 2014; Sui et al., 2015) and pharmacological comparisons (Keefe et al., 2011). At least one study confirms distinct outcomes for individuals who experience schizophrenia when compared with healthy controls on the MCCB (August et al., 2012). Confirmatory factor analyses have been previously employed in evaluating the ten cognitive tests that make up the MCCB, which includes domains of verbal and visual learning, reasoning, attention, processing speed, social cognition and verbal and non-verbal working memory (Burton et al., 2013; Harvey et al., 2013; Lo et al., 2016). These studies have compared different structures for the variable relationships in a factor analysis of the battery outcomes as continuous scale structure outcomes (see Lo et al., 2016).

There are a number of studies which have sought to confirm distinct factor structures across the MCCB tests. The specific study chosen for the example conceptual framework is Lo et al. (2016). It is partly chosen as a brief paper of six pages, with minimal detail included regarding qualitative aspects of both the study and the researcher's decision-making. The aim in presenting it here is to demonstrate what sorts of additions the conceptual framework would contribute in terms of formulating an informed perspective on the conduct of latent variable modelling in the context of psychiatric research that makes use of computational methods.

In Lo et al. (2016), reporting of an attempted confirmatory factor analysis is presented, which aimed to examine the factor or latent variable structure for the MCCB across what is described in the paper as "a larger, more diagnostically diverse sample of participants" (p. 79). The paper is comparing its samples to Burton et al. (2013) and Harvey et al. (2013). Each of these earlier studies had utilised factor analysis techniques for the MCCB items across schizophrenia-specific samples. Harvey et al. (2013) maintained there was support for a single latent variable in a continuous scale structure reflective model, representing global cognitive ability, following Nuechterlein et al. (2008). Burton et al. (2013) on the other hand had found no sound evidence for a unifactorial structure for schizophrenia-spectrum disorders, but some support for three latent variables in a model for the battery. With this information, what is set out below is a possible conceptual framework representation, for Lo et al. (2016). It is noted that MANOVA had been performed after the CFA to examine effects of age, gender, and diagnosis on outcomes, in the original paper; no address is made of the MANOVA analyses, in what follows.

In what is included below, elements are presented where these are supportive of claims made following the use of models from within the GLVM corpus, these include, the research situation, the researcher(s), research concept, construct, model, variables, data, phenomena, and quality control. Tables are presented in each element section. The first row in each table presents what information has been included or can be sourced via direct citation, from the brief report of Lo et al. (2016). The second row in the table presents what information was missed in the report, which would support claims made in the conclusion of the paper. The conclusion presented is that there is an effective replication of Burton et al. (2013) and Harvey et al. (2013), with evidence suggested as found for a three-factor structure with the factors nominated as representing processing speed, attention, and learning.

2.1 Research Situation

Table 8
Research situation element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	This research paper presents a secondary analysis of information collected in a “Thinking Skills for Work” program for individuals with severe mental illness conducted across a few sites in North America. Data collected for this program included the MCCB battery. As listed on the website http://www.matricsinc.org/mccb/ (accessed 24 February 2019), the MCCB aims to function as an outcome measure that can assess	The “Thinking Skills for Work” program targeted cognitive remediation for individuals with severe mental illnesses who were receiving vocational rehabilitation support in a community-based program. Three hundred participants are recorded for the study. The aim of the study is to replicate a three-factor latent variable model as presented in Harvey et al. (2013) and Burton et al. (2013). Lo et al. (2016) do not include the	There are three locations for randomised controlled trials, all are community mental health centres with stated ethics approval from relevant institutional boards - Mental Health Center of Greater Manchester (*n* = 72 and 47), Thresholds Inc, Chicago (*n* = 31, 92), Brooklyn mental health agency (*n* = 58). Different criteria applied for participants across the different studies beyond common criteria of 18 yrs of

	cognitive change in pharmacological intervention or cognitive remediation for schizophrenia and related disorders.	social cognition component of the MCCB battery in their analyses, following the rationale of Burton et al. (2013) which states that social cognition differs from task cognition.	age, a diagnosis of severe mental illness, an assessed desire to work, and no history of other neurological disorders. Study 1 added not having previously benefited from supportive employment programs, Study 2 added mild verbal learning/executive function impairment.
Further information needed	The explicit field in which the research is situated is not specified.		

2.2 Researcher(s)

Table 9

Researcher element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	Researcher affiliations are listed as Boston University, Department of Psychological & Brain Sciences, Dartmouth College Geisel School of Medicine Department of Psychiatry, and Boston University Centre for Psychiatric Rehabilitation, Department of Occupational Therapy, Psychology, and Psychiatry.	The researchers conduct this study under the National Institute of Mental Health grant R01 MH077210 and National Institute on Disability, Independent Living, and Rehabilitation Research grant H133G090206. Names as given on this paper include Stephen B. Lo, Kristin L. Szuhanym M. Alexandra Kredlow, Rosemarie Wolfe, Kim T. Mueser, and Susan R. McGurk.	Inter-rater reliability was confirmed prior to MCCB administration.
Further information needed	The research situation also is a domain in which the researchers are functioning, with reference to the "Thinking Skills for Work" program		Ideological commitments or any researcher stances in respect of the research situation are not specified, these may be philosophical positions such as realist, or values pursued, such as usefulness or parsimony.

2.3 Research concept

Table 10

Research concept element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)		"Impaired cognitive functioning" in individuals with schizophrenia and other severe mental illness (Lo et al., 2016, p. 79). In Table 1 the mental illnesses as listed are schizophrenia, schizoaffective disorder, bipolar, major depressing and other severe mental illnesses.	
Further information needed	It is not explicitly declared in this paper, but it could be expected that this research concept extends across multiple studies which would constitute the relevant domain for the concept.		In Table 1 on p. 80 it shows nearly 9% of the total participants were classified as "Other", no indication of what these are is included, and no indication of how the battery developed for schizophrenia could be expected to be relevant for or generalised to other severe mental illnesses is presented.

2.4 Research construct

Table 11

Research construct element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	Impaired cognitive function is assessed across a number of cognitive tests, including "processing speed, verbal learning, visual learning, verbal and nonverbal working memory, reasoning, attention, and social cognition." (p. 79)		
Further information needed		Several other tests seemed to have been used in a diagnostic capacity in respect of impaired cognitive function; no explicit reconciliation of these to the impaired cognitive function as assessed in the factor analysis is presented. The tests include Trail Making Test Part B, Brief Psychiatric Rating Scale, Positive and Negative Syndrome Scale (PANSS), and 'Difficulty in Abstract Thinking" item from PANSS (p. 80).	Any specific negations or exclusions for the construct, in reference to the specific tests of included in the formulation, can be included here.

2.5 Theory

Table 12

Research theory element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	Nuechterlein and Dawson (1984) present a theoretical review of earlier literature extending back to Bleuler who was perhaps the first to describe cognitive processing differences for individuals with schizophrenia symptomatology. The MATRICS itself is formulated by expert panel which helped select the most reliable and valid assessments available for cognitive differences in schizophrenia-spectrum patients. Although not explicitly declared, the background theory on the MATRICS seems to inform the present study.		Seven separable domains were identified for cognitive deficits in schizophrenia, including speed of processing, attention, working memory, verbal learning and memory, visual learning and memory, reasoning and problem solving, and verbal comprehension. An eighth domain was added as social cognition in the original research due to awareness of specific deficits for this ability in population experiencing symptoms of schizophrenia (Nuechterlein et al., 2004). This variable however was excluded in the present study on the rationale that the other measures in the battery specifically focused on neurocognitive rather than social ability (see Burton et al., 2013).
Further information needed		The MATRICS battery ultimately represents not just aspects of cognitive processing differences. This is because different aims for the NIMH, the Food and Drug Administration (FDA) in North America, and the team of people who came together to develop MATRICS were addressed in the construction of the battery to include pharmacological assessment and also intervention assessment across large multisite trials (Nuechterlein et al., 2008). The formulation needs to be clarified, in this respect, so that the pathway to cognitive processing differences is clear.	No theoretical reconciliation on expected differences between schizophrenia and other severe mental illnesses is presented

2.6 Model

Table 13

Research model element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables

Described in report (or cited reference)	The theoretical and expert-informed selection of battery tests for cognitive ability for individuals experiencing schizophrenia formulates the universe of discourse for the psychological model, which ultimately is connected to the CFA (as specified in the above 'Theory' element).	Confirmatory factor analysis with maximum likelihood estimation serves in tests of fit and stability for the three latent variable solution for the MCCB. The three latent variables were processing speed, attention/working memory, and learning (Lo et al., 2016, p. 81).	Model fit indices tested included the standardised root mean square residual (SRMR) at < 0.08, the root mean square error of approximation (RMSEA) at < 0.06, the comparative fit index (CFI) at > 0.95, and the non-normed fit index (NNFI) at > 0.95. The chi-square was not utilised in this process due to the large sample size. In this study, two distinct CFA models are tested for two different population groups, those with schizophrenia, versus those with some other severe mental illness. The first CFA was constructed without correlated errors between items but with latent variables correlated, the second with both correlated. Correlations were permitted between five pairs of variables with "theoretical overlap" (p. 81), one of these manifest variables (tests) appears in four out of five of the correlations. Model fit was further tested for this final model on just schizophrenia spectrum individuals with better fit than the one factor model.
Further information needed		No explicit strategy is described which takes the psychological model variable of "executive function" and transforms it into latent variable for "learning". No explicit mention made of which characteristics of executive function would not be addressed/would not be expected to be included as an item in CFA, or why this was the case.	Rationale for the correlation of errors, particularly given the unusual pattern in the correlation of errors with one manifest variable (test) from the processing speed latent variable correlating across manifest variables from each of the other latent variable correlations.

2.7 Variables

Table 14

Research variable element for Lo et al. (2016) example

Sub-elements	**Universe/domain**	**Formulation**	**Variables**

Described in report (or cited reference)		Three latent variables formulate the CFA in this paper, with manifest variables as listed correlating highly with each latent variable. i) Processing speed, the latent variable for Trail Making Part A; the Brief Assessment of Cognition in Schizophrenia (BACS) symbol coding subtest; the BACS category fluency subtest, and the Neuropsychological Assessment Battery (NAB) mazes subtest manifest variables; ii) Attention/memory, the latent variable for the Continuous Performance Test- Identical Pairs version (CPT-IP); the Weschler Memory Scale 3rd Edition (WMS-III) spatial span subtest, and the Letter-Number Span test manifest variables; and iii) Learning, the latent variable for Hopkins Verbal Learning Test Revised (HVLT-R) and Brief Visuospatial Memory Test Revised (BVMT-R) manifest variables.	
Further information needed	The universe for these variables is the model as specified in the 'Model' element.		Lo et al. (2016) do not report how tallying or scores for the model variables were calculated, but we would expect to find this here. No tests of validity were confirmed in Lo et al. (2016), but these could also be reported here.

2.8 Data

Table 15

Research data element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)			Normality tests are mentioned in connection to MANOVA – no mention is made of confirmation of normality for manifest

			variables in the CFA.
Further information needed	The universe for data for this project is the three studies, as the research situation accounted for above.	Data is not described by Lo et al. (2016), but to the degree that there are notable properties to report about the variables included in the model, they should be reported here.	What would be valuable to see from Lo et al. (2016) would be relevant properties of the data in terms of for example missing values and handling variation from normal distribution.

2.9 Phenomena

Table 16

Research phenomena element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)		The study occurs on baseline data that was collected for a cognitive enhancement program that is not fully described in Lo et al. (2016). Even though it is not described here in detail, the phenomena of interest are cognitive function in individuals specifically experiencing the symptoms of schizophrenia.	
Further information needed	The universe is the research situation - impaired cognitive functioning in diagnoses of schizophrenia and severe mental illness.		While they are not noted by Lo et al. (2016), in this section differences between the three study sites, as well as differences that may be relevant to interviewer/participant relationships at each of the sites could beneficially be included here.

2.10 Quality control

Table 17

Research quality control element for Lo et al. (2016) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)			

Further information needed	Very limited information is reported for reliability, validity, or other quality control measures taken. Some analysis is conducted with MANOVA investigating impact of diagnosis, symptom severity, and age on individual performance over the latent variables. No reconciliation is offered from the between-subjects correlation, to within subjects processes, connected to each or any of the cognitive tests in question. Inter-rater reliability is stated as confirmed; no indices are reported.	

2.11 Review of presented elements for Lo et al. (2016)

Several questions follow from the retro-fitted conceptual framework for the Lo et al. (2016) study. Firstly, filling in the conceptual framework revealed that the structure of the CFA in Lo et al. (2016) differed from that of the earlier studies it cited in terms of the correlation of errors. Neither Burton et al. (2013) nor Harvey et al. (2013) report for their CFA analysis that errors for any variables were correlated[25]. There is also no explanation given of the conceptual reduction of executive function in the construct of impaired cognitive function, to learning as a latent variable in the CFA. One reason that an account of such a reduction is important is that executive function generally is described as including a component for inhibitory control (Diamond, 2013). Where inhibitory control is affected, this could be expected to impact on performance on processing speed tasks. Lo et al. (2016) state in respect of the correlated errors that there is theoretical reason to expect that error on one of the variables from processing speed would be expected to be correlated with errors across other variables that load onto the other latent variables, but no indication is given as to the theoretical relationship between the specific manifest variables that are nominated as suitable for error correlation. Specification of explicit links between research

[25] It is possible that errors were correlated in the models of Burton et al. (2013) and Harvey et al. (2013) and just remained unreported as such. Filling in conceptual frameworks for all studies could in the future prevent the mis-specification of models based on omission of such information.

theory, concept, construct, variables, and data in a conceptual framework would clarify the relationship that is needed for any future attempts at replication.

Other differences exist in comparison to the reports of earlier studies that were revealed in retrofitting the conceptual framework. With respect to filling in the data element, Harvey et al. (2012) had deviations from normality in their data and so had used a modified maximum likelihood method, while Burton et al. (2013) had normal distribution, but missing data. No mention is made of data variation for Lo et al. (2016). This is important, as it is known that invalid performance features frequently for individuals experiencing severe mental illness, leading to likely data variation or absence (Sawyer, Testa, & Dux, 2017). Retrofitting the conceptual framework revealed that there is possibly error attributable to data variation, but a reader from the scientific community does not have access to information that would help them to understand the nature of data error for this GLVM analysis for this project. Filling in a conceptual framework for a research project with such features made clear supports maximal transparency, in this regard.

Reliability and validity indices are not reported explicitly in Lo et al. (2016), although some indices are reported in Burton et al. (2013) and Harvey et al. (2013) respectively for reliability and validity. The scientific community is not given information about the conduct of quality control practices for the Lo et al. (2016) study, although recommendations in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) would suggest that these would have been conducted in the process of factor analysis. Availability of this information in a standardised conceptual framework that could be made available perhaps separately to the peer-reviewed journal article report would support the research community in assessing the outputs from this GLVM analysis. Making such information available not only supports community evaluation of the contribution to knowledge made by the conduct of the project, it also serves to educate researchers about the steps needed to conduct GLVM analysis in research projects (see

Wilson, 2013a). This closes our commentary, on the Lo et al., (2016) study report.

3. Applied example – patient-centred psychiatric care

In the patient-centred care literature, the Mood and Anxiety Symptoms Questionnaire Dutch translation short form MASQ-D30 (Wardenaar et al., 2010) has been developed as a brief 30-item version of the older 90-item Mood and Anxiety Symptoms Questionnaire (MASQ), originally developed in Clark and Watson (1991), and validated in Watson et al. (1995a). One important point to note for the MASQ and its distinct versions are the variations in numbers of factors relevant to the questionnaire that are presented across different latent variable model studies of its items. Clark and Watson (1991) is cited in both Wardenaar et al. (2010) and Watson et al. (1995a) as the relevant initial body of work for distinguishing diagnoses for anxiety and depression via differential outcomes across three distinct and basic factors: negative affect (NA), positive affect (PA), and somatic arousal (SA). These factors together are known as the tripartite model (Clark & Watson, 1991). The original version of the tripartite model had three subscales that together made up the somatic arousal construct, these included General Distress: Anxiety, General Distress: Depression, and General Distress: Mixed (see Watson et al. 1995a). The tripartite model was originally proposed in research that noted some limitations of earlier scales that assessed mood and distress disorders, informed by DSM-III diagnoses (see Clark & Watson, 1991). These limit that included: i) use of clinician-assessed checklists for symptoms versus Likert-scale self-report; and ii) the limitation of old scales to assessment of positive and negative affect only, without including recognition of somatic arousal.

The original tripartite model relevant to the MASQ was developed and validated using principal components analysis (PCA: Gorsuch, 1983; Everett, 1983). It should be noted that PCA does not fall under the GLVM, as there is no latent variable in a principal components approach, what is modelled are weighted sums of observed or manifest variables given outcomes for one other observed variable (see Borsboom, 2005). Subsequent studies of the tripartite model however have made use of both exploratory and confirmatory factor analyses for the MASQ, and as stated above, the number of factors that are recorded across the subsequent studies differs, ranging from one factor (see Watson et al. 1995) to a new five factor model with items added from other scales (see den Hollander-Gijsman, de Beurs, van der Wee, van Rood and Zitman, 2010). Subsequent studies have also sought to develop shorter versions of the MASQ from the initial list of 90 items. For example, Wardenaar et al. (2010) present a 30-item version of the MASQ, the MASQ-D30, developed from Dutch psychiatric hospital Routine Outcome Monitoring (ROM) data, making use of confirmatory factor analysis to evaluate model fit in the second sample, with what they report as good fit indicators, following.

In the conceptual framework presented below, a validation study of the MASQ which includes testing of the MASQ-D30 in young people in an Australian mental health clinic is presented, from the study of Lin et al. (2014). The stated aim of this study is to compare outcomes from analysis of the MASQ with the MASQ-D30 in a sample of young people referred to or seeking help from a mental health service in Australia.

Similar to the example above, in what is included below, tables for elements are presented where these are supportive of claims made following the use of models from within the GLVM corpus, these include, the research situation, the researcher(s), research concept, construct, model, variables, data, phenomena, and quality control. The first row in each table presents what information has been included or can be sourced via direct citation, from Wardenaar et al. (2010). The second row in the table presents what information was missed in the report, which would support claims made in the conclusion. of the paper.

3.1 Research Situation

Table 18

Research situation element for Lin et al. (2014) example

Sub-elements	**Universe/domain**	**Formulation**	**Variables**
Described in report (or cited reference)	This research paper presents a validation study of the MASQ and the MASQ-D30, conducted with a 2003 sample of adolescents and young adults referred to but not necessarily treated by a youth mental health service in north and west metropolitan Melbourne, Australia.	The youth mental health service ("Orygen") which aims to support young people in early psychosis prevention, non-psychosis intervention, and in crisis evaluation and referral. Data for the MASQ was collected from 147 young people aged 15-24 years. The aim of the study is to replicate the three-factor tripartite model on which the formulation of the MASQ-D30 is founded, providing evidence for the validity of a quick assessment for depression and anxiety for young people.	The inclusion criteria for participants in this study were young people referred to the Orygen Youth Health program who are described as "help-seeking" (p, 779), exclusion criteria were a known organics cause for presentation, known intellectual disability (IQ < 70), and lack of ability to speak English.
Further information needed			

3.2 Researcher(s)

Table 19

Researcher element for Lin et al. (2014) example

Sub-elements	**Universe/domain**	**Formulation**	**Variables**
Described in report (or cited reference)	Researcher affiliations are listed as University of Birmingham, UK, Orygen Research Centre, Melbourne, Australia, Institute of Brain, Behaviour and Mental Health, UK, University of Groningen, Netherlands, and Maastricht University, Netherlands.	The researchers conduct this study funded by the Colonial Foundation (Australia). Names as given on this paper include Ashleigh Lin, Alison Yung, Joahnna Wigman, Eion Killackey, Gennady Baksheev, Klaas Wardenaar.	
Further information needed	The data analysed in this study was originally collected and presented in Godfrey et al. (2005), and was reanalysed with respect to the MASQ in Buckby et al. (2007a, b), the research situation functions then as the research domain for the researchers.		Ideological commitments or any researcher stances in respect of the research situation are not specified, these may be philosophical positions such as realist, or values pursued, such as usefulness or parsimony.

3.3 Research concept

Table 20

Research concept element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	This study makes use of the MASQ-D30 which was developed on an adult sample in the Netherlands (with a Dutch translation of the items).	The tripartite model for depression and anxiety as originally formulated in Clark and Watson (1991) is made up of factors for Anhedonic Depression, Anxious Arousal, and General Distress.	
Further information needed	The research area of interest described here is a dimensional approach to the diagnosis of depression and anxiety.		Although not discussed in the current study, the original formulation of the tripartite model in Clark and Watson (1991) also included a hierarchical structure for General Distress, which included factors for General Depression and General Anxiety.

3.4 Research construct

Table 21

Research construct element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	Evidence for the items chosen that make up the three factors of the tripartite model for the MASQ-D30 is presented in Wardenaar et al. (2010). This study made use of the original 90-item MASQ administered to two samples of adult psychiatric hospital admissions between 2002 – 2007. A principle components analysis was used to select the top 10 items for each factor.		

Further information needed		Original development of the questionnaire in Wardenaar et al. (2010) relied on administration of the Composite Diagnostic Interview (CIDI, WHO version 2.1) for diagnosing depression and anxiety disorders. Other scales included the Beck Anxiety Inventory (Beck et al., 1988); the Inventory of Depressive Symptomatology (Rush et al., 1996); and the Distress Four Dimensional Symptoms Questionnaire (Terluin et al., 2006). The study of Lin et al. (2014) however only used The Centre for Epidemiologic Studies Depression Scale (Radloff, 1977), and compared depression outcomes, only.	Any specific negations or exclusions for the construct, in reference to the specific tests included in the formulation, can be included here.

2.5 Theory

Table 22
Research theory element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	The development and validation of the MASQ-D30 is cited as valuable in Lin et al. (2014) because it reduces the administrative burden connected to assessment of variations in affect and hyperarousal in young people.		
Further information needed		The original MASQ development by Clark and Watson (1991) was an attempt to establish distinctions between diagnoses of depression and anxiety for DSM-IV, and also grounds for a diagnosis of mixed anxiety-depression.	No further connection of diagnoses of depression and anxiety with either categories for the same disorders in the DSM-V or ICD-10 is discussed, nor is any potential theoretical connection to other mental health disorders, for this item set.

3.6 Model

Table 23
Research model element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	The expert-informed selection of items for each of the factors formulates the universe of discourse for the psychological model..	Confirmatory factor analysis with weighted least squares (WLSMV) and polychoric correlations serve in tests of fit, validity, and internal consistency for the three latent variable solution for the MASQ-D30. The three latent variables were general distress, anhedonic depression, and anxious arousal (Lin et al., 2014, p. 778-779).	Model fit indices were tested for 1-, bi-, and 3- factor models for both the MASQ and MASQD-30, with the bi-factor model reported as having the best fit for both the MASQ (CFI = 0.93; RMSEA = 0.09) and the MASQD-30 (CFI = 0.96; RMSEA = 0.10). In the results section of the article, the MASQ-D30 is reported as having "adequate fit" compared to the 1-factor and 3-factor models (p. 780). It is not reported why the chi-square test is not utilised, or why other fit indicies are not reported. In the discussion section it is reported that the 3-factor structure was represented in this study. Internal consistencies are calculated and compared with Spearman-Brown adjustments for length for the MASQ-D30. These are reported as "similar" to the Wardenaar et al. (2010) results but no criteria for judgement is presented.
Further information needed		The Wardenaar et al. (2010) study utilised maximum likelihood estimation for the CFA; Lin et al. (2014) uses WLSMV in CFA. While the categorical nature of the factors is given as a rationale, no rationale of the changed statistical metrics for the model are given (see for more information Beauducel & Herzberg, 2006).	

3.7 Variables

Table 24
Research variable element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables

Described in report (or cited reference)		Three latent variables formulate the CFA in this paper (see Wardenaar et al., 2010): i) General distress, including items such as "Felt confused", "Felt dissatisfied with everything", "Had trouble making decisions"; ii) Anhedonic depression, including items such as negatively scored "Felt really talkative", "Felt successful", "Felt like I had a lot to look forward to"; iii) Anxious arousal, including items such as "Was short of breath", "Was trembling or shaking", "Felt nauseous".	
Further information needed	The universe for these variables is the model as specified in the 'Model' element.		Lin et al. (2014) do not report how tallying or scores for the model variables were calculated, but we would expect to find this here. No tests of validity were confirmed in Lin et al. (2014), but these could also be reported here.

3.8 Data

Table 25

Research data element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)			Corrections for skew are reported for the inter-item correlations for internal consistency (p. 779).
Further information needed	The universe for data for this project is the three studies, as the research situation accounted for above.	Data is not described by Lin et al. (2014), but to the degree that there are notable properties to report about the variables included in the model, they should be reported here.	What would be valuable to see from Lin et al. (2014) would be relevant properties of the data in terms of for example missing values and handling variation from normal distribution. This is important as Wardenaar et al. (2010) had included modifications for non-normal data, suggesting this data

			would also have non-normal properties.

3.9 Phenomena

Table 26

Research phenomena element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)		The study occurs on baseline data that was collected for a comorbidity analysis that is cited but not fully described in Lin et al. (2014).	
Further information needed	The universe is a subset of the research situation – depression and anxiety as presenting in a clinical sample of young people seeking support from a mental health service in southern Australia.		Criteria for the diagnosis of depression and anxiety have changed between DSM-III, when the tripartite model was originally established, DSM-IV, when much of the work validating the original MASQ was conducted, and DSM-V, which is the present diagnostic gold standard in mental health diagnosis. No reconciliation of these differences in diagnostic categories is presented.

3.10 Quality control

Table 27

Research quality control element for Lin et al. (2014) example

Sub-elements	Universe/domain	Formulation	Variables
Described in report (or cited reference)	This study attempts to replicate the factor structure of Wardenaar et al. (2010) in a sample of young adult data.	The tests of validity and reliability include reproduction of the factor models and previously established correlations, the production of internal consistency, Greatest Lower Bound statistics, and an IRT	Correlations are established with the CES-D only for convergent validity.

		model of the results.
Further information needed		No information is given about the decision to shift from Pearson's correlations used in Wardenaar et al. (2010) to polychoric correlations in this study, nor why the 3-factor model is reported as being the most appropriate fit when the bi-factor model seems to offer best fit statistics. No information is given about why no correlation pattern is sought with the anxiety and distress factors, to seek convergent validity across the full model.

3.11 Review of presented elements for Lin et al. (2014)

In a similar pattern to that seen in the analysis presented in the retro-fitted conceptual framework for Lo et al. (2016) above, a number of questions emerge from the above tabulation for Lin et al. (2014). Firstly, and perhaps most importantly, the study seems to provide evidence for a different model to that described in the original study that evidences the background for the MASQ-D30 – a bifactor hierarchical model, distinct from a 3-factor model. The authors describe the 3-factor model as having adequate fit, while the bifactor model is described as having better fit, in the results, with the 3-factor model presented in the discussion section, with no account of the bifactor model. What is possible to notice here is the selective privileging of both what is reported in terms of fit statistics for a CFA, but also the subjective or study-specific nature of the import of the weight given to the statistical outcomes that are reported. A key statistic of note with respect to the fit of the CFA model to the data is RMSEA. Lin et al. (2014) cite Brown (2006) as the relevant text referenced with respect to the decision-making with respect to model fit. Brown (2006) cites an acceptable RMSEA of ≤ .05 as representative of good model fit, and RMSEA of .06 or less as indicating "reasonably good" (p. 74) model fit, following Hu and Bentler (1999). The RMSEA statistics reported for all of the models tested in Lin et al. (2014) are equal to or greater than 0.07, indicating less than good reasonable model fit. The same chapter in Brown (2006) also states that CFI should be greater than 0.95, which is only achieved for the bifactor model of the MASQ-D30 in the Lin et al. (2014) study, and is not achieved across any of the other model types tested.

There are other differences in the methods employed by Lin et al. (2014) compared to earlier validation studies which at least should be explained as variances from the earlier study methods, as revealed in retrofitting the conceptual framework. For example, while the use of WLSMV estimation in the CFA is described as employed on the basis that the indicators for the factors are categorical, earlier studies had used maximum likelihood methods for CFA estimation. The difference is not noted in the Lin et al. (2014) study, and neither is the use of polychoric correlation matrices. Spearman's correlations are reported for the MASQ-D30 factors compared to four factors of the CES-D scales that include depressed affect, happiness, somatic and interpersonal factors, and it is stated that the MASQ factors correlated in similar fashion, without any statistics being reported. The study of Wardenaar et al. (2010) had employed specific separate scales that were correlated with the MASQ-D30 factors, each distinctively for depression, and anxiety, and general distress. Presumably the fact that the Lin et al. (2014) study is a retrospective analysis of data collected in 2003 for a different purpose meant that it was not possible to correlate scales that would specifically confirm the convergent validity of each of the factors. This may however be important as other studies which attempted and did not find evidence for the tripartite model as presented in Boschen & Oei (2006, 2007) for example report greatest weakness with the anxiety factor. This closes our commentary, on the Lin et al. (2016) study report.

4. Summary

In this section two examples have been presented of retrofitted conceptual frameworks from earlier completed studies in the computational psychiatry and patient-centred psychiatric care literature, respectively. These retrofitted examples raise a number of questions for the applications of the GLVM, which may have been answered where the conceptual framework had been applied in the development of the research project, and made available to the scientific community through something equivalent to the present pre-registration processes that are increasingly utilised, in psychology research. In the final section of this thesis, a presentation that contrasts the results of each of these examples will be made available, as will a summarisation of what has been said as the chapters have progressed, detailing the historical features of the development of the GLVM, that have been absorbed into the development of this conceptual framework tool.

Chapter 9: Conceptual Framework – Conclusion

1. Objectives

The objectives of this chapter are to provide an overview of the chapters, and to compare and contrast the worked-through examples of the conceptual framework presented in the previous chapter, before concluding with some future directions. We return to the definition of what a conceptual framework is and integrate insights from literature on qualitative research regarding trustworthiness, with a view to enhancing the quality of inferences made in psychological research that makes use of the GLVM. Throughout the previous chapters, it has been shown that a conceptual framework for the use of the GLVM in psychological research has the benefit of transparent disclosure of the critical links between representations of research phenomena, concepts, constructs, and variables. Such disclosure supports clarification of the means by which error enters the latent variable modelling process, and it supports reproducibility of research outcomes in terms of systematised disclosure of substantive and theoretical concerns, for each project. It has been shown that the conceptual framework can be used to record independent evidence about the existence of the phenomenon behind the latent variable, and this is vital for any reasoning that seeks to rely on the existence of the same phenomenon, as the model cannot be used to garner evidence for existence. This is because of the key features of conditional independence, the problem of factor indeterminacy, and the logical concern of circularity.

In this thesis, we have seen that latent variable modelling since its inception relied on mathematical structures and processes to obtain its results (see Chapters 2 and 3). Brief coverage has been presented of the history of development of mathematical set theory (see Chapter 5, section 3) as well as the subsequent emergence of mathematical constructivism.

To develop the conceptual framework, this thesis looked to 20th century developments in philosophy of science, including Suppes's (2002) informal structural view to represent invariances in science, and Rescher's (1979) approach to cognitive systematisation. This included examining the techniques used to understand inconsistencies in a network and connecting these techniques to the formulation of constraints for reasoning (see Chapter 7, section 2.4). Using these insights and grounding them on a local realist foundation, this thesis has have advocated for a constructivist network relational structure to support a theory of expected relations, for a research project (see Chapter 6, section 5). Reference to classical logic has guided much of the historical inquiry throughout the thesis, with a shift noted in the role for logic in a constructivist paradigm with respect to evaluating meaningfulness within a series of constraints, for a research project (see Chapter 7, section 2.4). This shift accompanies transitions in mathematical practice, from reliance on classical structures, to constructivist approaches that take verification of claims by expert community throughout the conduct of inquiry without reliance on the law of the excluded middle as key to the articulation of good knowledge.

These constraints are important in psychological research, as the research concepts that are employed always take initial form in geo-historically situated circumstances but are not confined to a specific population as a psychological construct is, following the insights of Markus (2008) (see Chapter 5, section 4). In describing a psychological phenomenon, the concept of 'model' in the history of blending of psychometric and psychological research practices has helped to define what it is that we are attempting to say something about (see Chapter 1 section 4.2). It has been shown that the strictures of a specifically mathematical modelling process do not lend themselves easily to saying something about what psychological phenomena are (see Chapter 3 section 2.2). Spearman never achieved a full proof for the split of test scores into components for latent and manifest variables based on patterns of correlation. Even though there remains no proof for the latent variable model, it is noted in Chapter 4 that psychology researchers have come to utilise the GLVM plus a series of supplementary statistical tests and confirmations of, for example, fit. This is in order to say something about what a construct or psychological attribute is, including a series of steps that aim at quality assurance regarding these inferences (see Chapter 4, section 3.3.4). We acknowledged that Spearman had been most interested in trying to develop evidence for his concept of *g*, or general cognitive ability. Spearman worked towards what he conceived of as a rigorous solution to the problem of definition of intelligence or cognitive ability by using mathematical processes from matrix algebra. We saw that mathematicians of the time tried to explain to Spearman in a variety of different ways that his technique could not secure proof under strict mathematical constraints (see Chapter 3 section 4.2). While Spearman tried to solve the problem of factor indeterminacy, we have acknowledged that Spearman's solutions do not work, following Maraun (1996), McDonald (2003), and Mulaik (2010). Factor indeterminacy remains a problem for the GLVM to this very day (see Chapter 4 section 3.1.2).

We tracked the inclusion of statistical approaches in latent variable modelling techniques and saw Fisher's attempts at securing proof for the maximum likelihood approach, witnessed the gaps, and acknowledged that this implies similar gaps whenever we use maximum likelihood in latent variable modelling solutions today (see Chapter 3 section 5.4). We also saw that a number of problems accompany the use of latent variable modelling analysis in researching psychological phenomena, including the aforementioned factor indeterminacy in light of conditional independence, measurement invariance, unidimensionality, model identification and equivalent models, and questions of causation, particularly where between-subjects data is used in an attempt to say something about within-subject processes (see Chapter 4 section 3). With these problems, we explored whether any of the generalised accounts of the latent variable model solved any of the problems that are present with the use of the GLVM. We found the answer to this question to be negative (see Chapter 3 section 2.4).

We further appreciated that psychologists may attempt to use the latent variable model in an effort to explain psychological phenomena, particularly where model advances are offered, such as when a new variable gets added to a confirmatory factor analysis model, perhaps as a result of exploratory factor analytical approaches (see Chapter 6 section 3). We looked at what is needed for causal explanation and found that extra-statistical information from beyond the GLVM model output must be presented wherever causal explanations are sought (see Chapter 4 section 3.2.3). Extra-statistical information about the process of latent variable modelling includes any principles, assumptions, independent evidence, influences on outcomes, choices about the form of model used for project analysis and so on. This information is vital in interpreting the findings that follow from GLVM analysis, yet there has been no effort to date to systematise the collection of this information into a universal structure that facilitates communication of research findings to a broader community. The conceptual framework in this thesis fills that gap.

In Chapter 5, we looked at recent contributions in latent variable modelling literature, and the connection these present to philosophy of science. We looked at whether ontology for a latent variable in a latent variable model can be logically derived from the way that practitioners use the latent variable model in their work. We said that while it is certainly possible to do this, there would be good reasons to think about what sort of logical statements could conclusively be made on the other side of such a rationale. Also we noted that such an ontology does not provide any decision making support as we progress through our research project, to be able to understand problems, such as for example when we may have strayed from the true nature of the psychological phenomena that we are interested in, by virtue of some aspect of the research situation.

In light of these concerns, we have noted that there is a role for exploring metaphysical grounding for research projects, but also that it is unlikely to be the case that the ground can be well set just by making inferences about what latent variable modellers think when they conduct their research (see Chapter 5, section 2 and 5). We surmised it might be the case that a research project is well supported when we have accounted for something like a theory of expected relations - which will account explicitly for researcher assumptions, premises or hypotheses, for a research situation. Specification of these can take place in a conceptual framework that is robust to both qualitative and quantitative inputs and the kinds of varied information needed for representation of substantive, mathematical, and statistical practices. In Chapter 6, we look at Suppes's body of work on empirical modelling practices, and in Chapter 7, we draw on Rescher's network model, which supports inferences from the best systematisation for any specific research project. All of these features are presented in two worked examples in Chapter 8. In this chapter, we compare the outcomes from Chapter 8 and examine how these support meaningful claims about the events or occurrences that we use to shape our understanding of what a psychological construct or attribute is or does follows. Before that, we return here to the definition of a conceptual framework and look to how the systematised theory of expected relations addresses the features of such a structure.

2. Conceptual framework revisited

2.1 Review

Generally defined, a conceptual framework is described in the literature as "the system of concepts, assumptions, expectations, beliefs, and theories that supports and informs your research" (Maxwell, 2013, p. 9). With respect to the conceptual framework formulated in this thesis, it has been shown that a conceptual framework is not only a system that supports the conduct of research. It is also an overarching system that: i) facilitates improved inferences from analyses that are performed in research which utilise any of the models that fall under the rubric of the GLVM by making possible the clarification of the logical pathway to those inferences and systematising the presentation of information beyond just the model form; and, ii) offers infrastructure for scrutiny about research processes and outcomes to peers, the broader community, and future researchers. In fields beyond psychology research, conceptual frameworks offer professionals a means to cohere their practices under ethical responsibility, such as in the international accounting standards (Murphy & O'Connell, 2013), and in responsible conduct of research codes in technology and engineering (Steneck, 2006; Vallor, Green, & Raicu, 2018). The conceptual framework presented in this thesis presents the possibility that researchers valuing an ethical obligation towards transparency will account for the integrity of their research decisions in a framework general enough to include assumptions and decisions connected to mathematical, statistical, and empirical content, no matter whether the framework is used in describing, explaining, or predicting psychological phenomena. It supports sustainability in research practices to the degree that multiple studies may be combined using the simple element structure, facilitating a new level of integration for meta-analyses conducted over studies, and supporting true comparison for replications. Because completion of the conceptual framework facilitates logical planning and testing of assumptions prior even to the conduct of the GLVM analysis, it has benefits akin to those of conceptual analysis in early recognition of logical confounds which should indicate a research program should change direction or in some cases, be terminated (see Petocz & Newbery, 2010).

This thesis has presented rationale for a conceptual framework that is realist constructivist. Realist, in that we take the research situation as an initial constraint characterising a network of relations relevant for representation of the project, and constructivist, in that use of the GLVM involves employment of a model that is a representation of psychological phenomena that does not offer the possibility of deductive closure under isomorphism. This aspect of the GLVM means that several researcher decisions are required to account for the connection between the phenomenon of interest and the model, as no one-to-one relationship is possible between these. The conceptual framework as a theory of expected relations forms a constructed representation of the research project that can be evaluated for consistency. Such an evaluation provides a quality control measure in reference to reductions in quality uncertainty that may be involved with claims made following the conduct of the research project. Taken as a whole, then, the conceptual framework for a research project should define the local research situation, and each element in the framework that represents some singular aspect of the research project should lend itself to scrutiny by the scientific community.

Chapter 4 introduced several reasons that a conceptual framework is essential for any project that makes use of the GLVM, in terms of presenting decisions, assumptions, and problems that a researcher must address in utilising the GLVM. The core set of these concerns is the three-fold existence of factor indeterminacy, the problem of circularity, and the nature of the assumption of conditional independence. These three concerns mean that we cannot rely on the application of the GLVM itself to provide evidence for the phenomena that notionally sits behind the latent variable. Independent evidence gathered regarding the phenomenon could be accounted for transparently in the conceptual framework. We have said that the conceptual framework must remain as general as possible, in structure. Firstly, this is in order to maintain relevance for the different techniques utilised in psychology research, including, mathematical, statistical, substantive, quantitative, and qualitative processes. Secondly, this is in order to maintain relevance to the different kinds of roles that the GLVM can play, these being descriptive, explanatory, or predictive roles in the production of scientific knowledge. Finally, this is in order to accommodate the possibility of several distinct philosophical positions that may be adopted in the process of utilising the GLVM.

A number of different perspectives on science and mathematics have been presented in the preceding chapters. Largely the advantage of the conceptual framework as presented is a function of clarifying qualitative conditions for the domains that are relevant to the project representation. We have talked about quality control as remedial for the lack of deductive closure in the making of good knowledge and the problems for our statistical inferences from latent variable modelling techniques. The quality control measures will be accounted for in qualitative terms, so we need some understanding of what the qualitative literature tells us about how to make good knowledge.

It is noted here that because of the broad variation in kinds of circumstances that are relevant to psychological research, a network systems approach to accounting for the relational structure is preferred. A network systems approach does not exclude a researcher from making use of a strictly axiomatic approach within their project or any aspect of their project, but the network approach does allow for any additional aspects of the project that fall outside of a strictly axiomatised account to be included in the overall representation of the research project.

2.2 Conceptual frameworks and scientific inferences

Two aspects of GLVM literature that are impossible to ignore particularly since the turn of the 21^{st} millennia are the causal explanatory account of the GLVM, and advocacy for the model as a measurement technique. In the metrology literature, Mari, Maul, and Wilson (2018) for example argue for causal explanatory accounts of general measurement circumstances. Noting that the example of Spearman's *g* does not have properties that occur in a network of lawful relationships as physical properties typically do, which would make g amenable to quantitative measurement, Mari et al. (2018) suggest that a causal explanatory role for the property in some observed phenomenon may suffice as a structure that lends itself to measurement principles (see also Cano et al., 2019). Here the role of the GLVM analysis is to 'create' or cause a measurement of a psychological phenomenon. In this thesis, we have explored the limits of causal reasoning with respect to the GLVM, as well as the limits of interpreting use of the GLVM as a measurement process. Where researchers seek to rely on causal explanatory accounts of the GLVM, or to characterise the modelling process as measurement, qualitative extra-statistical information is needed which support such interpretations of the practices, as seen in Chapters 4 and 5. The conceptual framework offers a standardised way to present such decisions, integrating qualitative and quantitative aspects of research decision making. To think into how quality assurance might be secured for qualitative information, we will turn now to the research literature on qualitative methodology, looking to how this literature addresses assurances of veracity, regarding its research outputs, seeking guidance in respect of how this informs us in constructing our representation, of methodology.

2.3 Qualitative and quantitative evaluation

2.3.1 The problem of construct validity for psychological research

The reproducibility crisis in psychology research has been associated with what have been labelled as "questionable measurement practices" (QMPs: Flake & Fried, in press). QMPs are defined by Flake and Fried as "as decisions researchers make that leave questions about the measures in a study unanswered" (p. 5). The authors note that such practices threaten the validity of conclusions made following research projects in respect of any claimed measurement of phenomena of interest. Examples provided include creation of on-the-fly scales for a phenomenon of interest, which may also involve opportunistic item tallying in making up what would be included as manifest variables in the GLVM. Little theoretical or substantive evidence may be referenced, for such creations. Other examples highlighted in Flake and Fried already noted in the psychometric corpus in respect of QMPs include imposing numbers on an unknown latent structure (Slaney, 2017), and the increasing body of literature that suggests that the linguistic structure of questionnaire items may account for a good deal of systematic variance in tally scores that largely otherwise to date has remained obscured (see Slaney, 2017; Maul, 2017; Michell, 1994). Any and all these problems bring concerns for any evaluation of reliability and validity of findings (Flake & Fried, in press; Slaney, 2017; Maul, 2017) – qualities that we bank on in evidencing the veracity of our findings in psychometric practices.

One of the ways to counter a claim regarding QMPs is to provide a systematised account of the qualitative information in respect of choice of questionnaire items or decisions regarding scale structure, in a conceptual framework such as the one presented in this thesis. This qualitative information regarding the quantitative model would need to be credible to the scientific community who are involved in evaluation of research outputs. A parallel for construct validity in the qualitative research literature is credibility (Shenton, 2004; Guba, 1981). We have noted above that much of the information recorded in the conceptual framework will be qualitative rather than quantitative in nature. The conceptual framework articulates constraints in the relational structure of elements. In attending to the quality control assignment for consistency discussed in Chapter 7, it can be expected that most researchers will optimise for the meaningfulness of entries in their sub-elements, and as much consistency as possible, between elements. The links between domains for elements act as a set of constraints on each other can be conducted in a process of self-audit and can be evaluated by peers in the scientific research community. In the future, this would likely be performed by technology, rather than human review. It is possible to conceive that technology can address the kind of testing involved in many-valued logic evaluations (see Rescher, 1969, 1995) across both qualitative and quantitative information, such as in NoSQL databases (Abramova & Bernadino, 2013). It remains challenging however to think about how technology could possibly ensure the trustworthiness of our research, when so many qualitative assumptions go into the relation between what goes into research data for a data project, and what exists, in the real world. To this extent, what we are interested in is addressing the credibility of our research outputs. Credibility is exactly the principle identified as at stake, in the recent literature on the reproducibility crisis. In the next section, we examine these qualities of credibility and trustworthiness, as described in the qualitative literature.

2.3.2 Trustworthiness, credibility, meaningfulness

Morrow (2005) notes that there are some features of research which extend beyond specification of the kind of data or kind of study that the research purports to be, and trustworthiness is one of these features. Positivists and post-positivists questioned whether trustworthiness could be an attribute of qualitative research outputs, on the assumption that the reliability and validity tests that they could perform over quantitative data were superior to the kinds of outcomes derived from qualitative report (Shenton, 2004; Levitt et al., 2017). We are at a historical moment however where the tables are somewhat turned, and the objectivity and trustworthiness of quantitative outcomes is precisely, at question.

Qualitative researchers have developed a series of practices aimed at enhancing research trustworthiness (Shenton, 2004). Championed originally by Guba (1981), trustworthiness has been described as being made up of four elements; credibility, or an address of internal validity in a study which provides evidence that the study does assess what is intended to be assessed; transferability, or an account of how or whether the research can be generalised beyond its own parameters, dependability, or the inclusion of operational detail such that any attempted replication of such findings could be performed, and confirmability, which are the steps that the researcher takes to ensure integrity of data in respect of fidelity to the phenomena.

We have already described aspects of the conceptual framework that address the latter three of these aspects of qualitative practices in Chapter 6. Transferability has been addressed insofar as we use the distinction between an infinite population and a finite sample in our definitions of research concepts, compared to research constructs. These distinctions can be defined in the element for research phenomena, and more will be described regarding making the distinction in the conceptual framework format, below.

Dependability includes a description of the research situation in detail and formulates the initial constraints as we have set them in this account of a realist constructivist conceptual framework. Dependability can also be supported by recording any pertinent information relevant to the tests, questionnaires, or items used in the study in the relevant elements, whether they take a role in data, variables, or models. Confirmability is supported in the researcher's account of the constraints between the phenomena and the data elements of the representation of the research project. A strong assertion of confirmability for continuous scale structure data in a quantitative paradigm for example may include double conjoint cancellation conditions (Michell, 1990), which ensure an evidential basis for a unit structure for representations of empirical phenomena.

Credibility is enhanced by providing information about the completion of quality control activities as described in chapter 7 and including an overall assessment of the consistency of the framework in the quality control activities. However, we can imagine that a singular assessment of consistency may not provide much insight regarding the true credibility or ultimately trustworthiness, of a research project. Shenton (2004) provides a granular perspective on the kinds of provisions that support robust credibility claims in qualitative analyses. These include: adoption of appropriate and well-established methods; explication of the nature of relationships with any involved or invested institutions; rigorous practices for sampling methods for participants; triangulation of information from different perspectives, techniques to maintain participant honesty; repeated assessment in data collection; referencing instances where the research concept in question is absent or negated; maintaining good quality supervision between researchers and field leaders; enhancing the opportunities for peer scrutiny; making good quality use of reflective practices for researchers; inclusion of background information on the researcher including qualifications and experience; confirmation practices with research participants; and thick description of phenomena and a reconciliation of the present research with earlier studies.

It is apparent from casual inspection that these points for enhancing trustworthiness of qualitative research have a good deal in common with Spearman's (1904) provisions for enhancing the veracity of the outcomes of what he had described as quantitative analysis, in his development of factor theory, discussed in Chapter 6. Spearman suggests we should conduct careful review of definitions used in previous literature to develop our research concept, that we clarify factors that were both in and deliberately out of our considerations, that we record the relevant conditions that may have an impact on our findings, and that we make some effort to assess and address instrumental uncertainty. If we examine what the criteria for trustworthiness for qualitative research asks us to add beyond these specifications, what we find is an expansion into the idea of contextualisation, the notion which already grounds our setting of constraints by beginning our account of the research project from its geo-historical situatedness. Contextualisation takes place in specification of the constraints of the elements represented in the conceptual framework. To refresh the perspective offered by the completion of the conceptual framework for the two retrofitted examples in Chapter 8, we now briefly compare and contrast the outcomes of the analyses for each of the studies.

3. Comparison of two retrofitted conceptual framework outcomes

In Chapter 8, two worked versions of the conceptual framework were retrofitted to previously published studies. Each study attempted in some form a replication of the use of the GLVM in an earlier analysis of a psychological construct, and each conceptual framework as constructed showed that the present studies in fact included a number of variations on the earlier studies that would bring question as to whether the later studies truly constituted a replication, of the earlier construct studies. Both conceptual frameworks from the Lo et al. (2016) MCCB study, and the Lin et al. (2014) MASQ-D30 study showed the value of distinguishing between concept and construct elements for the research phenomena, following the distinctions made by Markus (2008) and further endorsed in the set theoretical protocol established in this thesis. The systematised nature of the linkages between the research elements in the conceptual framework render undeniably the degree to which the later study in each circumstance can be identified as taking place for a distinct research construct as compared to the first study. This is not only with reference to different population groups, following Markus (2008), there are also differences in the formulations of the models, the accepted structures of error, the accepted goodness of fit, and the quality uncertainty reduction steps, in each process. Each researcher must in their minds reconcile exactly how their study functions as a replication of an earlier study or formulates an evidence base for a new construct – in each of the studies covered in Chapter 8, there is indication in the work done for each study that what is evidenced is actually if anything the underlabouring of a new construct, perhaps informed by the existence of some earlier convergent research, more than a direct replication of an earlier established construct, persay.

4. Future directions and conclusion

One question that arises in respect of the introduction of a universal conceptual framework for the GLVM is whether adding more procedures for researchers to complete in the conduct of their research can really deliver scientific benefit in terms of integrity, transparency, systematisation, evaluation of contributions to knowledge, and the resolution of gaps in scientific reasoning. Following suggestions made with respect to developing integrity in qualitative research, for example, Levitt et al. (2017, p. 5) note that procedures-based approaches to improving research rigor can have the counterproductive effect of reducing attention to the goals and context of the research project in question. It is well recognised in psychometric research literature that GLVM approaches such as factor analysis can be employed in substantive research without much attention to the actual statistical complexity of the method, and with much expectation that the method will produce reliable and definitive outputs in reference to research goals (Cliff, 1992; Sijstma, 2016; Wilson, 2013a). Suggestions made in the research literature that attend to improving research methods in aid of reducing for example questionable research practices note already that such recommendations do add to researcher workload and demand for resources in terms of for example increased education (Sijstma, 2016; Wilson, 2013a). Simultaneously, the research literature notes a divide between the skills and resources of the substantive researcher, the skills and resources of the psychometrician, and the degree to which development of substantive theory has perhaps suffered because of the lack of involvement of psychometricians with the conduct of substantive research projects (Borsboom, 2006).

The conceptual framework for the use of the GLVM in psychological research works at the intersection of these communities, providing a tool for communication between substantive researchers and psychometricians, and both groups and the broader research community. Primarily the conceptual framework serves as an evidence base, an evidence base that contains demonstrable constraints that can be scrutinised. The recommendations that follow for enhancing qualitative research integrity per Levitt et al. (2017) suggest an orientation to research fidelity with respect to the phenomenon of interest, and evaluation of utility of the outcomes in respect of research goals as a remedial to methodolatry (Bakan, 1967) under procedures-based approaches. A key feature for the conceptual framework is the recording of distinctions conceptualised by the researcher between the phenomenon of interest, the research concept, the construct, and the variable that is pursued in analysis. Some of these distinctions may collapse together for some projects, where for example, well-defined phenomena already exist in a field of interest (see for example, Wilson, 2013a). However, in the same way that mathematicians once relied on the invariance of number, to find that there are indeed various number systems, we cannot be sure at this point that an increased level of attention to detail will not lead to scientific fruit. Evidence from the history of mathematics would in fact suggest the opposite.

The GLVM remains as a mathematical model. Systematising the extra-mathematical and extra-statistical information involved in the conduct of a research project stands to not only support a claim for an inference from the best systematisation. It supports our knowledge about and development of substantive theory, as psychological researchers are supported in understanding more about the influences that are at play in the conduct of psychology project, influences elided in the reduction of representation of the phenomenon of interest to a single model form for GLVM analysis.

It can be observed in psychometrics literature that worthy procedures promoting step-by-step approaches aimed at enhancing precision and accuracy in measurement practices remain largely overlooked by psychometricians in the field. These include double cancellation conditions in conjoint measurement procedures (Krantz et al., 1971; Michell, 1988, 2017), and logical procedures for data-based test analysis (Slaney & Maraun, 2008). It remains unknown, today, whether fulfilment of tests of such conditions by researchers in the conduct of empirical projects may have beneficially reduced the kinds of problems with reproducibility for psychological phenomena demonstrated in the computational psychiatry example above. This is because such fulfilment would reduce some of the unknown sources of error, for the variables in question. Standardised conceptual frameworks can support growth of knowledge in respect of research practices, as much as they stand to provide evidence that supports a research claim, in virtue of the specified constraints.

What is known is that theory and models are inextricably tied to the processes of production of data for analyses (see Chapter 6) for psychological research making use of the GLVM. To the extent that the conceptual framework proposed in this thesis facilitates disclosure of the links between theory, model, data, and phenomena, growth in understanding of the nature of error associated with latent variable modelling is supported.

The nature of error associated with different forms of the GLVM has been the focus of conceptual frameworks proposed in the past (see MacCallum & Tucker, 1991). The proposal for this conceptual framework at this historical moment works to capitalise on both the recent endorsement of pre-registration for all psychological research, and the possibility that a general framework which supports the accumulation of both quantitative and quantitative aspects of model use under the GLVM paradigm lends itself to a full spectrum of research project situations. These may extend beyond the application of GLVM analysis. The cognitive-historical analysis undertaken in this thesis suggested that logical consistency serves as the fundamental test of a contribution to knowledge in science, and for logical consistency to be attributed to analyses that make use of the GLVM in psychological research, demonstration of links between representations of research project elements is vital because of the limits of the structure of the model itself and the assumptions and principles we must rely on, in deciding to use this model. It can be imagined that the future development of substantive theory in psychology research is best supported in logical consistency no matter whether the GLVM or some other form of analytical technique is endorsed. Whether performed by the research community, or somewhere in our futures, by machines, transparency regarding the links between representations of our project elements stand to reduce both error, and concern over non-reproducibility, of findings. Improving the contextualisation of our psychological knowledge by adding systematised qualitative information to our quantitative analysis ultimately stands us ready to offer an improved science, of the future.

References

Abelson, R.P. (1995). *Statistics as Principled Argument.* Hillsdale, NJ: Erlbaum and Associates.

Abramova, V., & Bernardino, J. (2013). NoSQL databases: MongoDB vs Cassandra. In *Proceedings of the International C* Conference on Computer Science and Software Engineering* (pp. 14-22). New York: Association for Computing Machinery.

Achinstein, P. (1971). *Law and Explanation: An Essay in the Philosophy of Science*. Oxford: Clarendon Press.

Adams, P. J., & Buetow, S. (2014). The place of theory in assembling the central argument for a thesis or dissertation. *Theory & Psychology, 24(1)*, 93-110.

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: a distinct problem in testing the equivalence of intra-and inter-individual model structures. *Frontiers in Psychology, 5*, 1-14.

Agazzi, E. (2011). Consistency, truth and ontology. *Studia Logica, 97(1),* 7-29.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In B.N. Petrox and F. Caski (Eds.). *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

Akerlof, G. A. (1970). The market for "lemons": quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84,* 488-500.

Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood. *Statistical Science, 12*, 162-176.

American Educational Research Association, American Psychological Association, Joint Committee on Standards for Educational, Psychological Testing (US), & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.

American Psychological Association. (2010*). Publication Manual of the American Psychological Association* (6th Ed.). Washington DC: Author.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 5) (pp. 111-150).

Aristotle (1999). *Physics.* R. Waterfield (trans.). Oxford: Oxford Unversity Press.

Avigad, J. (2000). *Classical and Constructive Logic*. Retrieved from https://www.andrew.cmu.edu/user/avigad/Teaching/classical.pdf on 23 March 2019.

Baccelli, J. (2018). Beyond the metrological viewpoint. *Studies in History and Philosophy of Science,* 1-6.

Bandura, A. (1986). Social Foundations of Thought and Action. In D.F. Marks (Ed.). *The Health Psychology Reader*. (pp. 94-106). Englewood Cliffs: Sage.

Barry, A. E., Chaney, B., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education & Behavior, 41(1)*, 12-18.

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological),* 293-321.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology, 34(1),* 93-99.

Bartholomew, D. J. (1984). The foundations of factor analysis. *Biometrika, 71(2),* 221-232.

Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis*. London: Griffin.

Bartholomew, D. J. (1995). Spearman and the origin and development of factor analysis. *British Journal of Mathematical and Statistical Psychology, 48(2),* 211-220.

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's Bonds model. *Psychological Review, 116*, 567–579.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach (Vol. 904)*. London: John Wiley & Sons.

Bauer, A. (2017). Five stages of accepting constructive mathematics. *Bulletin of the American Mathematical Society, 54(3),* 481-498.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9(1),* 3-29.

Bauwens, L., & Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations. *Journal of Econometrics, 119(2),* 381-412.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13(2),* 186-203.

Bell, D. (1981). The place of the Grundlagen in Frege's development. *The Philosophical Quarterly (1950-),* 31(124), 209-224.

Benjamin, A. C. (1962). Coherence theory of truth. In D. D. Runes (Ed.). *Dictionary of Philosophy* (p. 58). Totowa: Littlefield, Adams, and Company.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological bulletin, 107(2), 238-246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological bulletin, 88(3), 588-606.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16(1),* 78-117.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika, 45(3),* 289-308.

BiPM. (2012). The international vocabulary of metrology—basic and general concepts and associated terms (VIM). *Joint Committee for Guides in Metrology, 200,* 1-108.

Black, S. A., & Porter, L. J. (1996). Identification of the critical factors of TQM. *Decision Sciences, 27(1),* 1-21.

Blei, D. M. (2014). Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application, 1,* 203-232.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3,* 993-1022.

Berman, J., & Smyth, R. (2015). Conceptual frameworks in the doctoral research process: a pedagogical model. *Innovations in Education and Teaching International*, 52(2), 125-136.

Boag, S. (2007). ‘‘Real processes’’ and the explanatory status of repression and inhibition.

Philosophical Psychology, 20, 375–392.

Boag, S. (2010). Description and explanation within personality psychology research. In R. Hicks (Ed.). *Personality and Individual Differences: Current Directions.* (pp. 21-29). Bowen Hills: Australian Academic Press.

Boag, S. (2011). Explanation in personality psychology: "Verbal magic" and the five-factor model. *Philosophical Psychology, 24(2),* 223-243.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health,* 6, 149.

Boker, S. M., & Martin, M. (2013). On the equilibrium dynamics of meaning. In M.C Edwards & R.C. MacCallum (Eds.). *Current Issues in the Theory and Application of Latent Variable Models* (pp. 240-252). London: Routledge.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53(1),* 605-634.

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly,* 359-372.

Boole, G. (1854). *An Investigation of the Laws of Thought: on Which are Founded the Mathematical Theories of Logic and Probabilities.* London: Dover Publications.

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic,* 35-37.

Boring, E.G. (1957). *History of Experimental Psychology*. New York, NY: Appleton-Century-Crofts.

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics.* Cambridge: Cambridge University Press.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71(3),* 425–440.

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives,* 6, 25-53.

Borsboom, D. (2015). What is causal about individual differences?: A comment on Weinberger. *Theory & Psychology, 25(3),* 362-368.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review Of Clinical Psychology, 9,* 91-121.

Borsboom, D., Cramer, A. O., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In *The Concept of Validity: Revisions, New Directions and Applications*. London: IAP Information Age Publishing.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110(2),* 203-219.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111(4),* 1061-1071.

Boschen, M. J., & Oei, T. P. (2006). Factor structure of the Mood and Anxiety Symptom Questionnaire does not generalize to an anxious/depressed sample. *Australian & New Zealand Journal of Psychiatry, 40(11-12),* 1016-1024.

Boschen, M. J., & Oei, T. P. (2007). Discriminant validity of the MASQ in a clinical sample. *Psychiatry Research, 150(2),* 163-171.

Bourbaki, N. (1968). *Theory of Sets: Elements of Mathematics*. Paris: Hermann.

Boumans, M. (2016). Suppes's outlines of an empirical measurement theory. *Journal of Economic Methodology, 23(3),* 305-315.

Box, J. F. (1978). *R. A. Fisher. The Life of a Scientist.* New York: Wiley.

Brady, R., & Rush, P. (2009). Four basic logical issues. *The Review of Symbolic Logic, 2(3),* 488-508.

Brandom, R., & Rescher, N. (1979). The logic of inconsistency: a study in nonstandard possible-world semantics and ontology. *The Philosophical Quarterly, 31 (124),* 275-277.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern?. *Psychological Bulletin, 107(2),* 260-273.

Brouwer, L.E.J. (1907/1981). *Over de Grondslagen der Wiskunde, Doctoral Thesis, University of Amsterdam*. Reprinted with additional material, D. van Dalen (Ed.). Amsterdam: Matematisch Centrum.

Brouwer, L.E.J. (1908). The unreliability of logical principles. *Journal of Philosophy, 2,* 152–158.

Brouwer, L.E.J. (1923/1967). On the significance of the principle of excluded middle in mathematics, especially in function theory. In J. van Heijenoort (Ed.). *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*. (p. 334). Cambridge: Harvard University Press.

Brown, J. (1992). *The Definition of a Profession: The Authority of Metaphor in the History of Intelligence Testing, 1890-1930.* Princeton: Princeton University Press.

Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research, Second Edition.* New York: Guilford Publications.

Bruner, J. P., & Holman, B. (in press). Self-correction in science: Meta-analysis, bias and social structure. *Studies in History and Philosophy of Science,* 1-5.

Brunswik, E. (1952). *The Conceptual Framework of Psychology*. Chicago: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62(3),* 193-217.

Bunge, M., & Ardila, R. (1987). *Philosophy of Psychology.* New York: Springer.

Burt, C. (1941). The factors of the mind. *British Journal of Educational Psychology, 11(1),* 45-51.

Byrne, B.M., Shavelson, R.J., & Muthén, B.O. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.

Campbell, J. P. (1982). Editorial: Some remarks from the outgoing editor. *Journal of Applied Psychology, 67(6),* 691-700.

Cano, S.J., Pendrill, L.R., Melin, J., Fisher Jr, W.P. (2019). Towards consensus measurement standards for patient-centered outcomes. *Measurement,* 1-20.

Cantor, G. (1874). Ueber eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen. *Journal für die Reine und Angewandte Mathematik, 77*, 258-262.

Cantor, G. (1878). Ein Beitrag zur Mannigfaltigkeitslehre. *Journal für die Reine und Angewandte Mathematik, 84,* 242-258.

Cantor, G. (1882/1932). *Gesammelte Abhandlungen Mathematischen und Philosophischen Inhalts*. Berlin: Springer.

Cantor, G. (1883). *Über Unendliche, Lineare Punktmannichfaltigkeiten*. Göttingen: University of Gottingen.

Cantor, G. (1895/1915). *Contributions to the Founding of the Theory of Transfinite Numbers.* Chicago: Open Court Publishing Company.

Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., ... & Leow, A. D. (2017, August). DeepMood: Modeling mobile phone typing dynamics for mood detection. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 747-755). New York: Association for Computing Machinery.

Carnielli, W., Coniglio, M. E., & Marcos, J. (2001). Logics of formal inconsistency. In D.M. Gabbay, F. Guenthner, W. Carnielli, M.E. Coniglio, & J. Marcos. *Handbook of Philosophical Logic.* (2nd Ed.). Dordrecht: Springer.

Carnielli, W., & Malinowski, J. (2018). *Contradictions, from Consistency to Inconsistency* (Vol. 47). Cham: Springer.

Carnap, R. (1950). *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Carroll, J. B. (1941). A factor analysis of verbal abilities. *Psychometrika*, *6(5)*, 279-307.

Carroll, J. B. (1954). Individual differences. *Annual Review of Psychology, 5(1),* 127-148.

Cattell, R. B. (1934). Occupational norms of intelligence, and the standardization of an adult intelligence test. *British Journal of Psychology. General Section, 25(1),* 1-28.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40(3),* 153.

Cattell, R. B. (1945). The life and work of Charles Spearman. *Journal of Personality, 14(2),* 85-92.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1(2),* 245-276.

Cayley, A. 1855 Sur la transformation d'une fonction quadratique en elle mˆeme par des substitutions lin´eaires. In 1*889–1898 The Collected Mathematical Papers of Arthur Cayley, Vol. 2* (pp. 192–201). Cambridge: Cambridge University Press.

Ching, A., & Ishihara, M. (2010). The effects of detailing on prescribing decisions under quality uncertainty. *QME, 8(2),* 123-165.

Christensen, K. S., Oernboel, E., Nielsen, M. G., & Bech, P. (2019). Diagnosing depression in primary care: a Rasch analysis of the major depression inventory. *Scandinavian Journal of Primary Health Care*, 1-8.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100(3),* 316-336.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18(1),* 115-126.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, 186-190.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the Association for Computing Machinery, 13(6),* 377-387.

Cohen, G.A. (1978/2000). *Karl Marx's Theory of History: A Defence.* Princeton: Princeton University Press.

Cohen, M.R. and Nagel, E. (1935). *An Introduction to Logic and Scientific Method*. London: Routledge & Kegan Paul Ltd.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19(2),* 300-315.

Cooksey, R., & McDonald, G. (2011). *Surviving and Thriving in Postgraduate Research.* Prahran: Tilde University Press.

Cooper, N. (1978). The law of excluded middle. *Mind, 87(346),* 161-180.

Cooper, P. (2007). Knowing your ‘lemons’: quality uncertainty in UK higher education. *Quality in Higher Education, 13(1),* 19-29.

Corcoran, J. (2003). Introduction. In (G. Boole) *The Laws of Thought,* facsimile of 1854 edition. Buffalo: Prometheus Books.

Costa, N. C., & Chuaqui, R. (1988). On Suppes' set theoretical predicates. *Erkenntnis, 29(1),* 95-112.

Cowles, M. (2001). *Statistics in Psychology: An Historical Perspective.* New Jersey: Lawrence Erlbaum Associates.

Cox, D. R. (1956). *Some Problems Connected with Statistical Inference*. Carolina: North Carolina State University. Dept. of Statistics.

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

Crede, M., & Harms, P. (2019). Questionable research practices when using confirmatory factor analysis. *Journal of Managerial Psychology, 34(1),* 18-30.

Creswell, J. W. (1998). *Qualitative Research and Research design: Choosing Among Five Traditions.* London: Thousand Oaks.

Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement Theory and Public Policy: Proceedings of a Symposium in Honor of Lloyd G. Humphreys*. (pp. 147–171). Urbana: University of Illinois Press.

Cronbach, L. J., Gleser, G. N., & Nanda, H. H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profile.* New York: Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52(4),* 281-302.

Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology, 16,* 137-163.

Cybulski, L., Mayo-Wilson, E., & Grant, S. (2016). Improving transparency and reproducibility through registration: The status of intervention trials published in clinical psychology journals. *Journal of Consulting and Clinical Psychology, 84(9),* 753.

Danziger, K. (1990). *Constructing the Subject: Historical Origins of Psychological Research.* Cambridge: Cambridge University Press.

Dauben, J. W. (1990). *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Princeton: Princeton University Press.

Dauben, J. W. (1994). Topology: Invariance of dimension. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences, 2,* 939-946.

Dauben, J. (1992). Conceptual revolutions and the history of mathematics: Two studies in the growth of knowledge. In D. Gillies (Ed.). *Revolutions in Mathematics*. Oxford:

Clarendon Press (pp. 49–71).
Davies, E. B. (2003). *Science in the Looking Glass: What do Scientists Really Know?*. Oxford: Oxford University Press.
Davies, E. B. (2005). A defence of mathematical pluralism. *Philosophy of Mathematics, 3(13),* 252–276.
Davis, P. J. (1972). Fidelity in mathematical discourse: Is one and one really two?. *The American Mathematical Monthly, 79(3),* 252-263.
Davis, W. E., Giner-Sorolla, R., Lindsay, D. S., Lougheed, J. P., Makel, M. C., Meier, M. E., ... & Zelenski, J. M. (2018). Peer-Review Guidelines Promoting Replicability and Transparency in Psychological Science. *Advances in Methods and Practices in Psychological Science, 1(4),* 556-573.
Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika, 56(3),* 463-474.
Deary, I. J. (2000). *Looking Down On Human Intelligence: From Psychometrics To The Brain* (Vol. 34). Oxford: Oxford University Press.
De Bièvre, P. (2012). The 2012 International Vocabulary of Metrology:"VIM". *Accreditation and Quality Assurance, 17(2),* 231-232.
Demopoulos, W. (1994). Frege and the rigorization of analysis. *Journal of Philosophical Logic, 23(3),* 225-245.
Denzin, N. and Lincoln, Y. (2000). *Handbook of Qualitative Research*, Thousand Oaks: Sage.
Detlefsen, M., Gold, B., & Simons, R. A. (2008). Proof: Its nature and significance. In M. Detlefsen, B. Gold, & R.A. Simons. (Eds.). *Proof and Other Dilemmas: Mathematics and Philosophy* (pp. 1-32). New York: The Mathematical Association of America.
Devitt, M. (1991). *Realism and Truth*. Oxford: Blackwell.
Dewey, J. (1882/1969). Kant and philosophic method. In J. Boydston (Ed.). *The Early Works 1882-1898.* Vol. I. (pp 34-47). Carbondale: Southern Illinois University Press.
Diamond A. (2012). Executive functions. *Annual Review of Psychology, 64,* 135–168.
Diebold, F. X., & Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Economics, 4(1),* 1-21.
Dodd, S. C. (1928). The theory of factors. *Psychological Review, 35(3),* 211-234.
Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92(6),* 1087-1101.
Easton, V. J., & McColl, J. H. (1997). Statistics Glossary v1. 1. University of Glasgow: Enlighten.
Edelsbrunner, P. A., & Dablander, F. (2018). The psychometric modeling of scientific reasoning: a review and recommendations for future avenues. *Educational Psychology Review,* 1-34.
Edgeworth, F. Y. (1908). On the probable errors of frequency-constants. *Journal of the Royal Statistical Society, 71(2),* 381-397.
Edwards, A. W. (2000). The genetical theory of natural selection. *Genetics, 154(4),* 1419-1426.
Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5(2),* 155-174.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48(4),* 1-18.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network pschometrics: Combining network and latent variable models. *Psychometrika, 82(4),* 904-927.

Everett, J. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18(2),* 197-218.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthese, 134(1-2),* 45-69.

Feferman, S. 1998. *In the Light of Logic.* Oxford: Oxford University Press.

Feldman, S., & Johnston, C. (2014). Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology, 35(3),* 337-358.

Ferrario, R., & Schiaffonati, V. (2012). *Formal Methods and Empirical Practices. Conversations with Patrick Suppes* (Vol. 205). Stanford: CSLI Publications.

Ferreirós, J. (1995). "What fermented in me for years": Cantor's discovery of transfinite numbers. *Historia Mathematica, 22(1),* 33-42.

Ferreirós, J. (2007). *Labyrinth of Thought: a History of Set Theory and its Role in Modern Mathematics.* Basel: Birkhäuser.

Ferreirós, J. (2008). The crisis in the foundations of mathematics. In T. Gowers, J. Barrow-Green, & I. Leader (Eds.). T*he Princeton Companion to Mathematics.* (pp. 142-156). Princeton: Princeton University Press.

Ferreirós, J. (2016). *Mathematical Knowledge and the Interplay of Practices*. Princeton: Princeton University Press.

Finch, W. H., & French, B. F. (2015). *Latent Variable Modeling with R.* London: Routledge.

Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics, 41*, 155-160.

Fisher, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices of the Royal Astronomical Society, 80,* 758-770.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A (222),* 309-368.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society, 22,* 700-725.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London A, 144(852),* 285-307.

Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, 98(1),* 39-82.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference.* Oxford: Hafner Publishing Co.

Fisher, P. A., & Robie, C. (2019). A latent profile analysis of the Five Factor Model of personality: A constructive replication and extension. *Personality and Individual Differences, 139,* 343-348.

Flake, J. K., & Fried, E. I. (in press). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science,* 1-22.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science, 8(4),* 370-378.

Fletcher, R., & Powell, M. J. (1963). A rapidly convergent descent method for minimization. *The Computer Journal, 6(2),* 163-168.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment, 7(3), 286-299.

Fontaine, J.R.J. (2005). Equivalence. In K. Kemf-Leonard (Ed.). *Encyclopedia of Social Measurement* (pp. 803-813). San Diego: Academic Press.

Fraenkel, A. A., Bar-Hillel, Y., & Levy, A. (1973). *Foundations of Set Theory (Vol. 67)*. Oxford: Elsevier.

Franklin, J. (2014). *An Aristotelian Realist Philosophy of Mathematics: Mathematics as the Science of Quantity and Structure.* Dordrecht: Springer.

Franklin, J. (2015). Uninstantiated properties and semi-Platonist Aristotelianism. *The Review of Metaphysics, 69(1),* 25-45.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics, 12(2),* 101-128.

Fried, E. I., & Flake, J. K. (2018). Measurement Matters. *APS Observer, 31(30/3).*

Frege, G. (1918/1956). The thought: a logical inquiry. *Mind, LXV,* 289-311.

Frege, G. (1879/1967). Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In J. van Heijenoort (Ed.). *From Frege to Gödel: A Source Book in Mathematical Logic* (pp, 1-82). Boston: Harvard University Press.

Frege, G. (1884/1953). *The Foundations of Arithmetic (trans. J.L. Austin)*. New York: Philosophical Library.

Friedman, M. (1999). *Reconsidering Logical Positivism.* New York: Cambridge University Press.

Gabbay, D., & Woods, J. (2005). The practical turn in logic. In D.M. Gabbay & F. Guenthner (Eds.). *Handbook of Philosophical Logic, 2nd Edition*. Dordrecht: Springer. (pp. 15-122).

Galton, F. (1869). *Hereditary Genius: An Inquiry into its Laws and Consequences (Vol. 27).* London: Macmillan.

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium.* Hamburg: Perthes et Besser.

Gauss, K.F. (1860). *Briefwechsel Zwischen C.F. Gauss und H.C. Schumacher.* Leipzig: Georg Olms Verlag.

Geary, D. C. (2005). *The Origin of Mind*. Washington, DC: American Psychological Association.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25,* 186-192.

Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays.* New York: Basic Books.

Gergen, K. J., Josselson, R., & Freeman, M. (2015). The promises of qualitative inquiry.

American Psychologist, 70(1), 1-9.

Gibbs, J. W. (1901/1906). *The Scientific Papers of J. Willard Gibbs (Vol. 3).* London: Longmans, Green and Company.

Gibbs, J. W. (1902). *Elementary Principles in Statistical Physics.* New York: Charles Scribner's Sons.

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika, 24(3),* 229-252.

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor?. *Psychology Science, 50(1),* 21-43.

Gillies, D. (1992) The Fregean revolution in logic. In D. Gillies (Ed.). *Revolutions in Mathematics*. Oxford: Clarendon Press (pp. 265-305).

Giordani, A., & Mari, L. (2012). Measurement, models, and uncertainty. *IEEE Transactions on Instrumentation and Measurement, 61(8),* 2144-2152.

Given, L. M. (2008). *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412963909

Given, L. M. (2015). What is the role of a theoretical and/or conceptual framework in a qualitative study? In *100 Questions (and Answers) About Qualitative Research* (p. 59). London: SAGE Publications.

Gödel, K. (1992). *On Formally Undecidable Propositions of Principia Mathematica and Related Systems.* New York: Courier Corporation.

Godfrey, K., Yung, A., Killackey, E., Cosgrave, E., Pan Yuen, H., Stanford, C., & McGorry, P. (2005). Patterns of current comorbidity in young help-seekers: Implications for service planning and delivery. *Australasian Psychiatry, 13(4),* 379-383.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21(5)*, 725-740.

Goldstein, H. H., Murray, F. L, & yon Neuman, J. (1959). The Jacobi method for real symmetric matrices. *Journal of the Association for Computing Machinery, 6*, 59-96.

Gorsuch, R.L. (1983). *Factor Analysis*. Hillsdale: Lawrence Erlbaum.

Gossett, E. (2009). *Discrete Mathematics with Proof*. London: John Wiley and Sons.

Gould, S. J. (1981). *The Mismeasurement of Man*. New York: Norton.

Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S., & Tonidandel, S. (2018). From outcome to process focus: Fostering a more robust Psychological Science through registered reports and results-blind reviewing. *Perspectives on Psychological Science, 13(4),* 448-456.

Grattan-Guinness, I. (1982). Psychology in the foundations of logic and mathematics: the cases of Boole, Cantor and Brouwer. *History and Philosophy of Logic, 3(1),* 33-53.

Grattan-Guinness, I. (2008). Solving Wigner's mystery: The reasonable (though perhaps limited) effectiveness of mathematics in the natural sciences. *The Mathematical Intelligencer, 30(3),* 7-17.

Grattan-Guinness, I., & Bos, H. J. (Eds.). (2000). *From the Calculus to Set Theory, 1630-1910: An Introductory History.* Princeton: Princeton University Press.

Gray, J. (2008). *Plato's Ghost: the Modernist Transformation of Mathematics*. Princeton: Princeton University Press.

Green, C. D. (2015). Why psychology isn't unified, and probably never will be. *Review of*

General Psychology, 19(3), 207-214.

Green, C. D. & Groff, P. R. (2003). *Early Psychological Thought: Ancient Accounts of Mind and Soul.* Connecticut: Praeger.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective Evaluation: Improving The Usefulness Of Evaluation Results Through Responsive And Naturalistic Approaches*. San Francisco: Jossey-Bass.

Guilford, J. P. (1965). *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika, 18(4),* 277-296.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19(2),* 149-161.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory 1. *British Journal of Statistical Psychology, 8(2),* 65-81.

Guttman, L. (1965). Structure of interrelations among intelligence tests. In C.W. Harris (Ed.). *Proceedings of the Invitational Conference on Testing Problems* (pp. 25-36). Princeton: Educational Testing Service.

Haack, S. (2007). *Defending Science-Within Reason: Between Scientism and Cynicism.* New York: Prometheus Books.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* Cambridge: Cambridge University Press.

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods, 10(4),* 371-388.

Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *The American Journal of Psychology,* 219-234.

Haig, B. D. (2014). *Investigating the Psychological World: Scientific Method in the Behavioral Sciences.* Boston: MIT Press.

Haig, B. D. (2015). Commentary: Exploratory data analysis. *Frontiers in Psychology,* 6, 1.

Haig, B. D., & Borsboom, D. (2012). Truth, science, and psychology. *Theory & Psychology, 22*, 272–289.

Hamami, Y. (2014). Mathematical rigor, proof gap and the validity of mathematical inference. *Philosophia Scientiæ. Travaux d'Histoire et de Philosophie des Sciences, (18-1),* 7-26.

Hamami, Y. (2018). Mathematical inference and logical inference. *The Review of Symbolic Logic,* 1-40.

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The Essential Brunswik: Beginnings, Explications, Applications*. Oxford: Oxford University Press.

Hardy, G.H. (1940/1992). *A Mathematician's Apology*. Cambridge: Cambridge University Press.

Harman, G. H. (1965). The inference to the best explanation. *The Philosophical Review, 74(1),* 88-95.

Harsanyi, J. C. (1983). Mathematics, the empirical facts, and logical necessity. In H. Putnam (Ed.). *Methodology, Epistemology, and Philosophy of Science* (pp. 167-192). Dordrecht: Springer.

Hart, B., & Spearman, C. (1912). General ability, its existence and nature. *British Journal*

of Psychology, 1904-1920, 5(1), 51-84.
Hartley, R. E. (1954). Two kinds of factor analysis?. *Psychometrika, 19(3),* 195-203.
Haslanger, S. (1995). Ontology and social construction. *Philosophical Topics, 23(2),* 95-125.
Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9(2),* 139-164.
Harvey, P. D., Aslan, M., Du, M., Zhao, H., Siever, L. J., Pulver, A., ... & Concato, J. (2016). Factor structure of cognition and functional capacity in two studies of schizophrenia and bipolar disorder: Implications for genomic studies. *Neuropsychology, 30(1),* 28-39.
Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology, 13(3),* e1002106.
Heene, M., Kyngdon, A., & Sckopke, P. (2016). Detecting violations of unidimensionality by order-restricted inference methods. *Frontiers in Applied Mathematics and Statistics, 2,* 1-13.
Helmer, O., & Rescher, N. (1959). On the epistemology of the inexact sciences. *Management Science, 6(1),* 25-52.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15(2),* 135-175.
Henderson, L. (2018). Global versus local arguments for realism. In J. Saatsi (Ed.). *The Routledge Handbook of Scientific Realism*. (pp. 151-163). London: Routledge.
Herrnstein, R. J., & Boring, E. G. (1965). *A Source Book in the History of Psychology.* Cambridge, MA: Harvard University Press.
Hershberger, S. L. (1994). The specification of equivalent models before the collection of data. In A. von Eye & C. C. Clogg (Eds.). *Latent Variables Analysis: Applications for Developmental Research* (pp. 68-105). Thousand Oaks, CA, US: Sage Publications, Inc.
Hibberd, F. J. (2014). The metaphysical basis of a process psychology. *Journal of Theoretical and Philosophical Psychology, 34(3),* 161-186.
Hobson, E. W. (1923). *The Domain of Natural Science*. New York: Wiley.
Hodrick, R. (2014). *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets (Vol. 1).* London: Routledge.
Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics,* 1523-1543.
Hood, S. B. (2009). Validity in psychological testing and scientific realism. *Theory & Psychology, 19(4),* 451-473.
Hood, S. B. (2013). Psychological measurement and methodological realism. *Erkenntnis, 78(4),* 739-761.
Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck, & R.C. MacCallum(Eds.). *Factor Analysis at 100: Historical Developments and Future Directions.* (pp. 205-249). Mahwah, NJ, Lawrence Erlbaum Associates.
Hotelling, H. (1930). The consistency and ultimate distribution of optimum statistics. *Transactions of the American Mathematical Society, 32(4),* 847-859.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24(6),* 417-441.

Howe, W.G. (1955). *Some Contributions to Factor Analysis.* USAEC Report ORNL-1919.

Howe, M. J. A. (1990). Does intelligence exist. The Psychologist, 3(11), 490-493.

Hoyningen-Huene, P. (2013). *Systematicity: The Nature of Science*. Oxford: Oxford University Press.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal, 6(1),* 1-55.

Huneman, P. (2018). Outlines of a theory of structural explanations. *Philosophical Studies, 175(3)*, 665-702.

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience, 19(3),* 404-413.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2(8),* e124.

Isvoranu, A., Guloksuz, S., Epskamp, S., Van Os, J., & Borsboom, D. (in press). Toward incorporating genetic risk scores into symptom networks of psychosis. *Psychological Medicine,* 1-8.

Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-driven methods to diversify knowledge of human psychology. *Trends in Cognitive Sciences, 22(1),* 1-5.

Jackson, D. N., & Chan, D. W. (1980). Maximum-likelihood estimation in common factor analysis: A cautionary note. *Psychological Bulletin, 88(2),* 502-508.

Jacquette, D. (2009). Nicholas Rescher's systematic philosophy - introduction. In: D. Jacquette, (Ed.). *Reason, Method, and Value. A Reader on the Philosophy of Nicholas*
Rescher. Reading Rescher: Vol. 4. (pp. 1-17). Heusenstamm bei Frankfurt: Ontos.

Jacquette, D. (2014). *Ontology*. London: Routledge.

Jansson, J., & Waxell, A. (2011). Quality and Regional Competitiveness. *Environment and Planning A: Economy and Space, 43(9),* 2237–2252.

Jeffrey, R.C. (1981). *Formal Logic, its Scope and Limits (2nd Ed.)*. New York: McGraw-Hill.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32(4),* 443-482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 34, 183-202.

Jöreskog, K. G. (1970). A general method for estimating a linear structural equation system. *ETS Research Bulletin Series,* i–41.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36(4),* 409-426.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *SAGE Focus Editions* (Vol. 154, p. 294). Thousand Oaks, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling.
Journal of Marketing Research, 19, 404–416.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A Guide to the Program and*

Applications. Mooresville: Scientific Software.

Joreskog, K. G., Sorbom, D., & Magidson, J. (1979*). Advances in Factor Analysis and Structural Equation Models.* Cambridge: Abt Books.

Jöreskog, K.G. and Wold, H. (1982). The ML and PLS Techniques For Modeling with Latent Variables: Historical and Comparative Aspects. In H. Wold and K. Jöreskog (Eds.), *Systems Under Indirect Observation: Causality, Structure, Prediction (Vol. I)* (pp. 263-270), Amsterdam: North-Holland.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23(3),* 187-200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20(1),* 141-151.

Kaiser, H. F. (1976). Book reviews: DN Lawley and AE Maxwell. Factor Analysis as a Statistical Method . New York: American Elsevier Publishing Company, Inc., 1971. pp. viii+ 153. $10.50. *Educational and Psychological Measurement, 36(2),* 586-589.

Kane, H., & Brand, C. (2003). The importance of Spearman's g. *The Occidental Quarterly, 3(1),* 7-30.

Kaplan, A. (1964). *The Conduct Of Inquiry: Methodology For Behavioral Science*. San Francisco: Chandler.

Kapsner, A. (2014). *Logics and Falsifications: A New Perspective on Constructivist Semantics.* Dordrecht: Springer.

Karlin, S., & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis, 10(4),* 467-498.

Kelley, T. L. (1935). *Essential Traits of Mental Life.* Boston: Harvard University Press.

Kelley, T. L. (1940). Comment on Wilson and Worcester's "note on factor analysis". *Psychometrika, 5(2),* 117-120.

Keogh, E., & Reidy, J. (2000). Exploring the factor structure of the Mood and Anxiety Symptom Questionnaire (MASQ). *Journal of Personality Assessment, 74(1),* 106-125.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin, 138(3),* 427-457.

Kim, J. (1995). Explanation. In R. Audi (Ed.). *The Cambridge Dictionary of Philosophy* (pp. 256-257). Cambridge: Cambridge University Press.

Kincaid, H. (2000a). Formal rationality and its pernicious effects on the social sciences. *Philosophy of the Social Sciences, 30(1)*, 67-88.

Kincaid, H. (2000b). Global arguments and local realism about the social sciences. *Philosophy of Science, 67*, S667-S678.

Kincaid, H. (2002). Scientific Realism and the Empirical Nature of Methodology: Bayesians, Error Statisticians and Statistical Inference. In S. Clarke, & T. D. Lyons (Eds.). *Recent Themes in the Philosophy of Science* (pp. 39-62). Dordrecht: Springer.

Kitcher, P. (1984). *The Nature of Mathematical Knowledge.* Oxford: Oxford University Press.

Kleiner, I. (1991). Rigor and proof in mathematics: A historical perspective. *Mathematics Magazine,* 64(5), 291-314.

Kline, R.B. (1998). *Principles and Practice of Structural Equation Modeling*. New York:

The Guilford Press.
Kline, R. B. (2012). Assumptions in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 111-125). New York, NY, US: Guilford Press.
Kneale, W. C., & Kneale, M. (1962). *The Development of Logic*. Oxford: Oxford University Press.
Knorr, W.R. (1975). *The Evolution of the Euclidean Elements. A Study of the Theory of Incommensurable Magnitudes and its Significance for Early Greek Geometry.* Dordrecht: D. Reidel.
Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New dimensions in the quantitative classification of mental illness. *Archives of General Psychiatry, 68(10),* 1003-1011.
Krantz, S.G. (2007). The history and concept of mathematical proof. In Hansen, V. L., & Gray, J. (Eds.). *History of Mathematics*. UK: EOLSS Publications.
Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Vol. I: Additive and Polynomial Representations.* New York: New York Academic Press.
Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., ... & Bagby, R. M. (2018). Progress in achieving quantitative classification of psychopathology. *World Psychiatry, 17(3),* 282-293.
Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2,* 111-133.
Lange, M. (2012). What makes a scientific explanation distinctively mathematical?. *The British Journal for the Philosophy of Science, 64(3),* 485-511.
Lange, M. (2016). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford: Oxford University Press.
Lange, M. (2018). Because Without Cause Scientific Explanations by Constraint. In A. Reutlinger and J. Saatsi. *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations* (pp. 15-43). Oxford: Oxford University Press.
Lauritzen, S. L. (1996). *Graphical Models.* London: Clarendon Press.
Lawley, D. N. (1940). VI.—the estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, 60(1)*, 64-82.
Lawley, D. N. (1942). XIV.—Further Investigations in Factor Estimation. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 61(2),* 176-185.
Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *Journal of the Royal Statistical Society. Series D (The Statistician), 12(3),* 209-229.
Lawley, D. N., & Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method.* London: Butterworths.
Lazarsfeld, P. F. (1959). Latent structure analysis. In S. Koch (Ed.). *Psychology: A Study of Science. Vol. 3.* (pp. 476–543). New York: McGraw-Hill.
Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent Structure Analysis.* New York: Houghton Mifflin Co..
LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications on Statistics, 1*, 277-330.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research, 25(3),* 313-334.

Leibniz, G. W. (1693). Supplementum geometriae dimensoriae, seu generalissima omnium tetragonismorum effectio per motum: similiterque multiplex constructio lineae ex data tangentium conditione. *Acta Eruditorum,* pp. 385-392.

Leonelli, S. (2014). Data interpretation in the digital age. *Perspectives on Science, 22(3),* 397-417.

Leonelli, S. (2015). What counts as scientific data? A relational framework. *Philosophy of Science, 82,* 810 – 821.

Levant, R. F., & Hasan, N. T. (2008). Evidence-based practice in psychology. *Professional Psychology: Research and Practice, 39(6),* 658.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist, 54(4),* 421-431.

Levitt, H. M., Motulsky, S. L., Wertz, F. J., Morrow, S. L., & Ponterotto, J. G. (2017). Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology, 4(1),* 2-22.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48(3),* 936-949.

Lin, A., Yung, A. R., Wigman, J. T., Killackey, E., Baksheev, G., & Wardenaar, K. J. (2014). Validation of a short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ) in adolescents and young adults. *Psychiatry Research, 215(3),* 778-783.

Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of" scale analysis" and factor analysis. *Psychological Bulletin, 45(6),* 507-529.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3(3),* 635-694.

Lord, F. M. (1950). *Notes on Comparable scales for Test Scores (RB-50-48).* Princeton NJ: Educational Testing Service.

Lord, F. M. (1952). A theory of test scores. *Psychometrika Monograph No. 7, 17* (4, Pt. 2).

Lord, F. M. (1980). *Applications Of Item Response Theory To Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA, USA: Addison-Wesley.

Lovie, P., & Lovie, A.D. (1995). The cold equations: Spearman and Wilson on factor indeterminacy. *British Journal of Mathematical and Statistical Psychology,* 48, 237-253.

Lovie, P., & Lovie, A. D. (1996). Charles Edward Spearman, FRS (1863-1945). *Notes and Records of the Royal Society of London, 50(1),* 75-88.

Lovie, S., & Lovie, P. (2010). Commentary: Charles Spearman and correlation: a commentary on ‘The proof and measurement of association between two things’. *International Journal of Epidemiology, 39(5),* 1151-1153.

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: resolution by maximum likelihood?. *Multivariate Behavioral Research, 41(4),* 499-532.

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology, 41(1)*, 79-87.

Luce, D., Krantz, D., Suppes, P., & Tversky, A. (1990). *Foundations of Measurement, Vol. III: Representation, Axiomatization, and Invariance*. San Diego: Academic Press Inc.

Lumsden, J. (1961). The construction of unidimensional tests. *Psychological Bulletin, 58(2)*, 122-131.

Luppino, F. S., van Reedt Dortland, A. K., Wardenaar, K. J., Bouvy, P. F., Giltay, E. J., Zitman, F. G., & Penninx, B. W. (2011). Symptom dimensions of depression and anxiety and the metabolic syndrome. *Psychosomatic Medicine, 73(3),* 257-264.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin, 109(3),* 502-511.

MacCallum, R. C., Browne, M. W., & Cai, L. (2007). Factor analysis models as approximations . In R. Cudeck, & R.C. MacCallum (Eds.). *Factor Analysis at 100* (pp. 153-175). Mahwah, NJ: Lawrence Erlbaum.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114(1),* 185-199.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4(1),* 84-99.

MacKenzie, D. (1976). Eugenics in Britain. *Social Studies of Science, 6(3-4),* 499-532.

Mackie, J. L. (1973). *Truth, Probability and Paradox: Studies in Philosophical Logic*. Oxford: Oxford University Press.

Maddy, P. (1997). *Realism in Mathematics.* Oxford: Clarendon Press.

Maddy, P. (2011). *Defending the Axioms: On the Philosophical Foundations of Set Theory.* Oxford: Oxford University Press.

Maddy, P. (2012). The philosophy of logic. *Bulletin of Symbolic Logic, 18(4),* 481-504.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur?. *Perspectives on Psychological Science, 7(6),* 537-542.

Mäki, U. (2005). Reglobalizing realism by going local, or (how) should our formulations of scientific realism be informed about the sciences?. *Erkenntnis, 63(2),* 231-251.

Maraun, M. D. (1996). Metaphor taken as math: Indeterminancy in the factor analysis model. *Multivariate Behavioral Research, 31(4),* 517-538.

Maraun, M. (2017). The object detection logic of latent variable technologies. *Quality & Quantity, 51(1),* 239-259.

Maraun, M. D., & Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas in Psychology, 31(1),* 32-42.

Mari, L. (2005). The problem of foundations of measurement. *Measurement., 38(4),* 259-266.

Markus, K. A. (2008). Constructs, concepts and the worlds of possibility: Connecting the measurement, manipulation, and meaning of variables. *Measurement, 6(1-2),* 54-77.

Markus, K. A. (2013). Correspondence without correspondence theory: Comment on Haig and Borsboom. *Theory & Psychology, 23(6),* 806-811.

Markus, K. A., & Borsboom, D. (2013a). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning.* New York, NY: Routledge.

Markus, K. A., & Borsboom, D. (2013b). Reflective measurement models, behavior domains, and common causes. *New Ideas in Psychology, 31(1),* 54-64.

Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J., & Beckmann, C. F. (2016). Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1(5)*, 433-447.

Marsh, T., & Boag, S. (2014). Unifying psychology: Shared ontology and the continuum of practical assumptions. *Review of General Psychology*, 18(1), 49-59.

Marsonet, M. (1994). On Rescher's conceptual idealism. *Idealistic Studies, 24(2),* 147-161.

Maul, A. (2013). On the ontology of psychological attributes. *Theory & Psychology, 23(6),* 752-769.

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15(2)*, 51-69.

Maydeu-Olivares, A., & McArdle, J. J. (2005). *Contemporary Psychometrics*. New York: Psychology Press.

Mayer, T., & Steneck, N. (2012). *Promoting Research Integrity in a Global Environment.* Singapore: World Scientific.

Maze, J. R. (1954). Do intervening variables intervene?. *Psychological Review, 61(4),* 226-234.

Maze, J.R. (1983). *The Meaning of Behavior*. London: Allen & Unwin.

Maze, J. R. (2001). Social constructionism, deconstructionism and some requirements of discourse. *Theory & Psychology*, *11(3)*, 393-417.

McClimans, L., Browne, J., & Cano, S. (2017). Clinical outcome measurement: Models, theory, psychometrics and practice. *Studies in History and Philosophy of Science Part A, 65,* 67-73.

McDonald, R. P. (1967). Factor interaction in nonlinear factor analysis*. *ETS Research Bulletin Series*, i-18.

McDonald, R. P. (1968). A unified treatment of the weighting problem. *Psychometrika, 33,* 351–381.

McDonald, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika, 39(2),* 203-222.

McDonald, R. P. (1977). The indeterminacy of components and the definition of common factors. *British Journal of Mathematical and Statistical Psychology, 30(2)*, 166-176.

McDonald, R. P. (1979). The simultaneous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology, 32(2)*, 212-228.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology,* 34(1), 100-117.

McDonald, R. P. (1987). *Factor Analysis and Related Methods.* Mahwah, NJ: Lawrence Earlbaum.

McDonald, R. P. (1996). Latent traits and the possibility of motion. *Multivariate Behavioral Research, 31(4),* 593-601.

McDonald, R. P. (1999). *Test Theory: A Unified Approach*. Mahwah, NJ: Lawrence Earlbaum.

McDonald, R. P. (2003). Behavior domains in theory and in practice. *Alberta Journal of Educational Research, 49(3),* 212-230.

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7(1),* 64-82.

McDonald, R. P., & Mulaik, S. A. (1979). Determinacy of common factors: A nontechnical review. *Psychological Bulletin, 86(2),* 297-306.

McKenna, F. P. (1985). Another look at the 'new psychophysics'. *British Journal of Psychology,* 76(1), 97-109.

McKinsey, J. C., & Suppes, P. (1955). On the notion of invariance in classical mechanics. *The British Journal for the Philosophy of Science, 5(20)*, 290-302.

McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science, 3(6),* 836-849.

Meehl, P. E. (1992). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. *Psychological Reports, 71,* 339-467.

Meissner, W. W. (1981). Metapsychology: Who needs it?. *Journal of the American Psychoanalytic Association*, 29, 921-938.

Meissner, W. W. (1984). Models in the mind: The role of theory in the psychoanalytic process. *Psychoanalytic Inquiry,* 4(1), 5-32.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58(4),* 525-543.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care,* S69-S77.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin, 89(3),* 575-588.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13-103). New York: Macmillan, New York.

Meyrick, J. (2006). What is good qualitative research? A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology, 11(5),* 799-808.

Michell, J. (1988). Some problems in testing the double cancellation condition in conjoint measurement. *Journal of Mathematical Psychology, 32(4),* 466-473.

Michell, J. (1994). Numbers as quantitative relations and the traditional theory of measurement. *The British Journal for the Philosophy of Science, 45(2),* 389-406.

Michell, J. (1999). *Measurement in Psychology: A Critical History of a Methodological Concept.* Cambridge: Cambridge University Press.

Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology,* 10(5), 639-667.

Michell, J. (2004a). Item response models, pathological science and the shape of error: Reply to Borsboom and Mellenbergh. *Theory & Psychology, 14(1),* 121-129.

Michell, J. (2004b). The place of qualitative research in psychology. *Qualitative Research in Psychology, 1(4),* 307-319.

Michell, J. (2008). Is psychometrics pathological science?. *Measurement, 6(1-2),* 7-24.

Michell, J. (2011). Qualitative research meets the ghost of Pythagoras. *Theory & Psychology, 21(2),* 241-259.

Michell, J. (2017). On substandard substantive theory and axing axioms of measurement: A response to Humphry. *Theory & Psychology, 27(3),* 419-425.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in crosscultural research. *International Journal of Psychological Research, 3(1),* 111-121.

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R.W. Lissitz (Ed.). *The Concept of Validity: Revisions, New Directions and Applications* (pp. 83-108). Charlotte, NC : Information Age Publishing.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement, 2(4), 201-218.

Molenaar, P., & von Eye, A. (1994). On the arbitrary nature of latent variables. In A. von Eye and C.C. Clogg (Eds.). *Latent Variable Analysis: Applications for Developmental Research.* Thousand Oaks: Sage.

Moore, G. H. (1988). The emergence of first-order logic. *History and Philosophy of Modern Mathematics, 11*, 95-135.

Motwani, J. (2001). Critical factors and performance measures of TQM. *The TQM Magazine, 13(4),* 292-300.

Moulines, C. U. (2006). Ontology, reduction, emergence: A general frame. *Synthese, 151(3),* 313-323.

Mulaik, S. A. (1985). Exploratory statistics and empiricism. *Philosophy of Science, 52(3),* 410-430.

Mulaik, S. A. (1986). Factor analysis and Psychometrika: Major developments. *Psychometrika, 51(1), 2*3-33.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22(3),* 267-305.

Mulaik, S. A. (2005). Looking back on the indeterminacy controversies in factor analysis. In A. Maydeu-Olivares, & J.J. McArdle. (Eds.). *Contemporary Psychometrics: A Festschift for Roderick P. McDonald* (pp. 173-206). New Jersey: Lawrence Erlbaum Associates.

Mulaik, S. A. (2010). *Foundations of Factor Analysis. New* York: Chapman and Hall/CRC.

Mulaik, S. A., & McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika, 43(2),* 177-192.

Muller, F. A. (2011). Reflections on the revolution at Stanford. *Synthese, 183(1),* 87-114.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1(1),* 0021.

Mundy, B. (1986). On the general theory of meaningful representation. *Synthese, 67(3),* 391-437.

Murphy, T., & O'Connell, V. (2013). Discourses surrounding the evolution of the IASB/FASB Conceptual Framework: What they reveal about the "living law" of accounting. *Accounting, Organizations and Society, 38(1),* 72-91.

Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data.

Journal of Econometrics, 22(1), 43-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49(1),* 115-132.

Muthén, B. O. (1993). Goodness of fit with categorical and other non-normal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 205–243). Newbury Park, CA: Sage.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour,* 3, 221-229.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation.* London : Routledge & Kegan Paul.

Narens, L. (2002). A meaningful justification for the representational theory of measurement. *Journal of Mathematical Psychology, 46(6),* 746-768.

National Research Council. (2002). *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct*. Washington: National Academies Press.

Nelder, J. & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, 135(3),* 370–384.

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69,* 511-534.

Nersessian, N. J. (1995). Should physicists preach what they practice?. *Science & Education, 4(3),* 203-226.

Nersessian, N. J. (2008). Mental modeling in conceptual change. In S. Vosniadou (Ed.). *International Handbook of Research on Conceptual Change* (pp. 391-416). New York: Routledge.

Newton-Smith, W. (1978). The underdetermination of theory by data. In *Rationality in Science* (pp. 91-110). Dordrecht: Springer.

Norton, B. (1979). Charles Spearman and the general factor in intelligence: Genesis and interpretation in the light of sociopersonal considerations. *Journal of the History of the Behavioral Sciences, 15(2),* 142-154.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science, 348(6242),* 1422-1425.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115(11),* 2600-2606.

Nunnally, J.C. (1978). *Psychometric Theory.* New York: McGraw Hill.

OECD (2010). *Investigating Research Misconduct Allegations in International Collaborative Research Projects*. Retreived from http://www.oecd.org/sti/inno/42770261.pdf

O'Neil, W.M. (1982). *The Beginnings of Modern Psychology.* Sydney: Sydney University Press.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7(6),* 657-660.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349(6251)*, aac4716.

Ostrowski, A. (1920). Über dirichletsche reihen und algebraische differentialgleichungen. *Mathematische Zeitschrift, 8(3-4),* 241-298.

Passmore, J. A. (1935). The nature of intelligence. *Australasian Journal of Psychology and Philosophy, 13,* 279–289.

Pearl, J. (1988). *Probabilistic Inference in Intelligent Systems.*

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82(4),* 669-688.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series, 5(50),* 157-172.

Pearson, K. (1904). On the laws of inheritance in Man. II. On the inheritance of the mental and moral characters in Man, and its comparison with the inheritance of the physical characters. *Biometrika , 3,* 131-190.

Pearson, K. (1907). *On Further Methods of Determining Correlation, Mathematical Contributions To The Theory Of Evolution - Biometric Series IV.* London: Dulau.

Pearson, K., & Moul, M. (1925). The problem of alien immigration into Great Britain, illustrated by an examination of Russian and Polish Jewish children. *Annals of Eugenics, 1(1),* 5-54.

Pearson, K., & Moul, M. (1927). The mathematics of intelligence. *Biometrika,* 246-291.

Perks, J. (2010). Commentary:'The next trick is impossible.'. *International Journal of Epidemiology, 39(5),* 1153-1155.

Petocz, A. (2015). The scientific status of psychoanalysis revisited. In S. Boag, L.A.W. Brakel & V. Talvitie (Eds.), *Philosophy, Science and Psychoanalysis* (pp. 145-192). London: Karnac Books.

Petocz, A., & Mackay, N. (2013). Unifying psychology through situational realism. *Review of General Psychology*, 17(2), 216-223.

Petocz, A., & Newbery, G. (2010). On conceptual analysis as the primary qualitative approach to statistics education research in psychology. *Statistics Education Research Journal*, 9(2), 123-145.

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature, 465(7299),* 721-727.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6(10),* 421-425.

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology, 52(2),* 126-136.

Ponterotto, J. G. (2006). Brief note on the origins, evolution, and meaning of the qualitative research concept thick description. *The Qualitative Report, 11(3),* 538-549.

Popper, K.R. (1935/1992). *The Logic of Scientific Discovery*. London: Routledge.

Popper, K. (1970). Normal science and its dangers. In *Criticism and the Growth of Knowledge,* (I. Lakatos & A. Musgrave, Eds.). Cambridge: Cambridge University Press, pp. 51-58.

Potter, M. (2004). *Set Theory and its Philosophy: A Critical Introduction*. New York:

Oxford University Press.

Pratt, J. W. (1976). FY Edgeworth and RA Fisher on the efficiency of maximum likelihood estimation. *The Annals of Statistics, 4(3),* 501-514.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Psillos, S. (2009). *Knowing the Structure Of Nature: Essays On Realism And Explanation.* Dordrecht: Springer.

Quine, W. (1969). *Ontological Relativity: and Other Essays*. New York: Columbia University Press.

Quine, W. V. (1970). On the reasons for indeterminacy of translation. *The Journal of Philosophy, 67(6),* 178-183.

Rabe-Hesketh, S., & Skrondal, A. (2008). Classical latent variable models for medical research. Statistical Methods in Medical Research, 17(1), 5-32.

Rao, C. R. (1961). Asymptotic efficiency and limiting information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics.* California: The Regents of the University of California.

Rao, C. R. (1992). RA Fisher: The founder of modern statistics. *Statistical Science, 7(1),* 34-48.

Rasch, G. (1960). *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*. Denmark: Nielsen & Lydiche.

Rasch, G. (1966). An item analysis which takes individual differences into account. British *Journal of Mathematical and Statistical Psychology, 19(1),* 49-57.

Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model?. *Structural Equation Modeling, 8(1),* 142-149.

Raykov, T., & Marcoulides, G. A. (2007). Equivalent structural equation models: A challenge and responsibility. *Structural Equation Modeling, 14(4),* 695-700.

Raykov, T, & Marcoulides, G.A. (2011) *Introduction to Psychometric Theory.* New York, NY: Routledge.

Raykov, T., & Penev, S. (1999). On structural equation model equivalence. *Multivariate Behavioral Research, 34(2),* 199-244.

Reichenbach, H. (1971). *The Theory of Probability*. California: University of California Press.

Reidy, J., & Keogh, E. (1997). Testing the discriminant and convergent validity of the Mood and Anxiety Symptoms Questionnaire using a British sample. *Personality and Individual Differences, 23(2),* 337-344.

Rescher, N. (1958). On prediction and explanation. *The British Journal for the Philosophy of Science, 8(32),* 281-290.

Rescher, N. (1968). *Topics in Philosophical Logic* (Vol. 17). Dordrecht: D. Reidel Publishing Company.

Rescher, N. (1969). *Many-valued Logic*. New York: McGraw Hill.

Rescher, N. (1977). *Methodological Pragmatism.* New York: New York University Press.

Rescher, N. (1979). *Cognitive Systematization.* Oxford: Blackwell.

Rescher, N. (1995). *Satisfying Reason: Studies in the Theory of Knowledge*. Dordrecht: Springer.

Rescher, N. (1997). *Objectivity: The Obligations of Impersonal Reason*. Michigan: The University of Michigan.

Rescher, N. (2005a). Pragmatism at the crossroads. *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy, 41(2*), 355-365.

Rescher, N. (2005b). *Reason and Reality: Realism and Idealism in Pragmatic Perspective*. New York: Rowman & Littlefield.

Rescher, N. (2016). Inference from the best systematization. *Mind & Society, 15(2),* 147-154.

Rescher, N., & Manor, R. (1970). On inference from inconsistent premisses. *Theory and Decision*, *1(2),* 179-217.

Revelle, W. (2015). Charles Spearman. In R. Cautin, R. and S. Lilienfeld (Eds.). *The Encyclopedia of Clinical Psychology.* London: Wiley-Blackwell.

Ricketts, T., & Levine, J. (1996). Logic and truth in Frege. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 121-175.

Roberts, F. S. (1980). On Luce's theory of meaningfulness. *Philosophy of Science, 47(3),* 424-433.

Roberts, F. S. (1985). Applications of the theory of meaningfulness to psychology. *Journal of Mathematical Psychology, 29(3),* 311-332.

Robinson, D. N. (2007). Theoretical Psychology What Is It and Who Needs It?. *Theory & Psychology*, 17(2), 187-198.

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology, 6,* 1715.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika*, *49(3),* 425-435.

Rozeboom, W. W. (1988). Factor indeterminacy: The saga continues. *British Journal of Mathematical and Statistical Psychology*, *41(2),* 209-226.

Russell, B. (1912). *The Problems of Philosophy.* Oxford: Oxford University Press.

Ryle, G. (1971). *Collected Papers: Volume II: Collected Essays (1929–1968).* London: Hutchinson.

Salmon, W. C. (1989). Four decades of scientific explanation. *Scientific Explanation, 13,* 3-219.

Savage, L.J. (1954). *The Foundations of Statistics.* New York: Wiley.

Sawyer, R. J., Testa, S. M., & Dux, M. (2017). Embedded performance validity tests within the Hopkins Verbal Learning Test–Revised and the Brief Visuospatial Memory Test–Revised. *The Clinical Neuropsychologist*, *31(1),* 207-218.

Schaeffer, J. (2009). On what grounds what. In D. Manley, D.J. Chalmers & R.Wasserman (Eds.). *Metametaphysics: New Essays on the Foundations of Ontology*. (pp. 347-383). Oxford: Oxford University Press.

Schaeffer, J. (2012). Grounding, transitivity, and contrastivity. In F. Correia & B. Schnieder (Eds.). *Metaphysical Grounding: Understanding the Structure of Reality*. (pp. 122-138). Cambridge: Cambridge University Press.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29(4),* 304-321.

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., &

Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology, 31(1)*, 43-53.
Schönemann, P. H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika, 36(1),* 21-30.
Schretlen, D. J., Peña, J., Aretouli, E., Orue, I., Cascella, N. G., Pearlson, G. D., & Ojeda, N. (2013). Confirmatory factor analysis reveals a latent cognitive structure common to bipolar disorder, schizophrenia, and normal controls. *Bipolar Disorders, 15(4),* 422-433.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6(2),* 461-464.
Schwartz, N. A. (2017). Nancy Nersessian's cognitive-historical approach. In *Springer Handbook of Model-Based Science* (pp. 355-375). Cham: Springer.
Shanin, N. A. (1997). Georg Cantor as the author of constructions playing fundamental roles in constructive mathematics. *Journal of Mathematical Sciences, 87(1),* 3183-3191.
Shapiro, S. (2004). Foundations of mathematics: metaphysics, epistemology, structure. *The Philosophical Quarterly*, 54(214), 16-37.
Shields, P. M., & Tajalli, H. (2006). Intermediate theory: The missing link in successful student scholarship. *Journal of Public Affairs Education*, 313-334.
Sila, I., & Ebrahimpour, M. (2002). An investigation of the total quality management survey based research published between 1989 and 2000. *International Journal of Quality and Reliablity Management, 19(7),* 902–930.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366.
Skow, B. (2014). Are there non-causal explanations (of particular events)?. *The British Journal for the Philosophy of Science*, 65, 445-467.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, And Structural Equation Models.* London: Chapman and Hall/CRC.
Slaney, K. (2017). *Validating Psychological Constructs: Historical, Philosophical, and Practical Dimensions.* Cham: Springer.
Slaney, K. L., & Maraun, M. D. (2008). A proposed framework for conducting data-based test analysis. *Psychological Methods, 13(4),* 376-390.
Slaney, & Racine. (2013). What's in a name? Psychology's ever evasive construct. *New Ideas in Psychology, 31(1),* 4-12.
Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.
Spearman, C. (1914). The theory of two factors. *Psychological Review, 21(2),* 101-115.
Spearman, C. (1916). Comments on Mr Thomson's paper. *British Journal of Psychology, 8,* 282–284.
Spearman, C. (1925). Some Issues in the Theory of``g"(including the Law of Diminishing Returns). *Nature, 116,* 436-439.
Spearman, C. (1927). *The Abilities of Man.* London: Macmillan.
Spearman, C. (1929). The uniqueness of "g.". *The Journal of Educational Psychology,*

212-216.
Spearman, C. (1930/1961). C. Spearman. In Murchison, C. (Ed.), *A History of Psychology in Autobiography Volume IV* (pp. 299-333). Worcester, MA: Clark University Press.
Spearman, C. (1933). The uniqueness and exactness of g. *British Journal of Psychology. General Section, 24(1),* 106-108.
Spearman, C. (1946). Theory of general factor. *British Journal of Psychology, 36(3),* 117-131.
Spearman, C., & Holzinger, K. (1925). Note on the sampling error of tetrad differences. *British Journal of Psychology. General Section, 16(2),* 86-88.
Staats, A. W. (1981). Paradigmatic behaviorism, unified theory, unified theory construction methods, and the Zeitgeist of separatism. *American Psychologist, 36,* 239 –256.
Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25(1),* 78-90.
Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika, 44(2),* 157-167.
Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of common factors. Handout for a presentation delivered at the meeting of the Psychometric Society, Iowa City, I A.
Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research, 21(3),* 309-331.
Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12(1)*, 53-74.
Stevens, S. (1951). Mathematics, measurement and psychophysics. In S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley.
Stewart, I. (1995). *Concepts of Modern Mathematics*. New York: Dover Publications.
Stigler, S. M. (1981). Gauss and the invention2004b of least squares. *The Annals of Statistics,* 465-474.
Stigler, S. M. (1989). Francis Galton's account of the invention of correlation. *Statistical Science*, 73-79.
Stoll, R. R. (1961). *Set Theory and Logic*. San Francisco: W.H. Freeman and Company.
Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F., Star, S.A., & Clausen, J.A. (1950). *Measurement and Prediction*. Princeton: Princeton University Press.
Struik, D. J. (Ed.). (1969). *A Source Book in Mathematics, 1200-1800 (Vol. 11)*. Harvard: Harvard University Press.
Suppes, P. (1954). Some remarks on problems and methods in the philosophy of science. *Philosophy of Science, 21(3),* 242-248.
Suppes, P. (1959). Measurement, empirical meaningfulness and three-valued logic. In C.W. Churchman and P. Ratoosh (Eds.). *Measurement: Definitions and Theories*. (pp. 129-143). New York: Wiley.
Suppes, P. (1961). A comparison of the meaning and uses of models in mathematics and the empirical sciences. In H. Freudenthal (Ed.). *The Concept and the Role of the Model in Mathematics and Natural and Social Science*s (pp. 163-177). Dordrecht: Springer.

Suppes, P. (1962/1969). Models of data. In *Studies in the Methodology and Foundations of Science* (pp. 24-35). Dordrecht: Springer.

Suppes, P. (1967). *Set-Theoretical Structures in Science.* California: Stanford University.

Suppes, P. (1969a). Measurement, empirical meaningfulness, and three-valued logic. In *Studies in the Methodology and Foundations of Science*. (pp. 65-80). Dordrecht: Springer.

Suppes, P. (1974). The structure of theories and the analysis of data. In F. Suppe (Ed.), *The Structure of Scientific Theories* (pp. 266-283). Urbana, IL: University of Illinois Press.

Suppes, P. (1975). Logics appropriate to empirical theories. In C.A. Hooker (Ed.). *The Logico-Algebraic Approach to Quantum Mechanics.* (pp. 329-340). Dordrecht: Springer.

Suppes, P. (1993). Representation theory and the analysis of structure. *Philsophia Naturalis, 25*, 254-268.

Suppes, P. (2001). Axiomatic theories. In N.J. Smelser and P.H. Baltes (Eds.). *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Elsevier.

Suppes, P. (2002). *Representation and Invariance of Scientific Structures.* Stanford, CA: CSLI Publications.

Suppes, P. (2003). From theory to experiment and back again. In M. C. Galavotti (Ed.). *Observation and Experiment in the Natural and Social Sciences* (pp. 1-41). Dordrecht: Kluwer Academic Publishers.

Suppes, P. (2006). Transitive indistinguishability and approximate measurement with standard finite ratio-scale representations. *Journal of Mathematical Psychology, 50(3),* 329-336.

Suppes, P. (2011). Future development of scientific structures closer to experiments: Response to F.A. Muller. *Synthese, 183(1),* 115-126.

Suppes, P. (2016). Qualitative axioms of uncertainty as a foundation for probability and decision-making. *Minds and Machines, 26(1-2)*, 185-202.

Suppes, P., & Winet, M. (1955). An axiomatization of utility based on the notion of utility differences. *Management Science, 1(3-4)*, 259-270.

Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible?. *Synthese, 48(2),* 191-199.

Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 1–76). New York: Wiley.

Tal, E. (2012). *The Epistemology of Measurement: A Model-Based Accoun*t. Canada: University of Toronto.

Tallon, A. (1997). *Head and Heart: Affection, Cognition, Volition as Triune Consciousness*. New York: Fordham University.

Tarski, A. (1933). The concept of truth in the languages of the deductive sciences. *Division III of Mathematics and Physics, 34,* 13–172.

Tarski, A. (1946). *Introduction to Logic and to the Methodology of Deductive Sciences.* New York: Dover Publications, Inc.

Tarski, A. (1956). The concept of truth in formalized languages. *Logic, Semantics, Metamathematics, 2,* 152-278.

Teo, T. (2006). *The Critique of Psychology: From Kant to Postcolonial Theory*. London: Springer Science & Business Media.

Teo, T. (2010). Ontology and scientific explanation: Pluralism as an a priori condition of psychology. *New Ideas in Psychology, 28(2)*, 235-243.

Teo, T. (2015). Critical psychology: A geography of intellectual engagement and resistance. *American Psychologist, 70(3)*, 243-254.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology, 8*, 271–281.

Thomson, G. H. (1934). Hotelling's method modified to give Spearman's g. *Journal of Educational Psychology, 25(5)*, 366-374.

Thomson, G. H. (1934b). On measuring g and s by tests which break the g hierarchy. *British Journal of Psychology. General Section, 25(2)*, 204-210.

Thomson, G. H. (1935). The definition and measurement of "g" (general intelligence). *Journal of Educational Psychology, 26*, 241-262.

Thomson, G. (1939). The factorial analysis of human ability. *British Journal of Educational Psychology, 9(2)*, 188-195.

Thomson, G. (1947). Charles Spearman, 1863-1945. *Biographical Memoirs of the Fellows of the Royal Society, 5(15)*, 373-385.

Thomson, G. H. (1951). *The Factorial Analysis of Human Ability (5th ed.).* London: University of London Press.

Thompson, P. (1988). Explanation in the semantic conception of theory structure. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association (Vol. 1988, No. 2.* (pp. 286-296). Chicago: Philosophy of Science Association.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. New York: American Psychological Association.

Thorndike, E. L., Lay, W., & Dean, P. R. (1909). The relation of accuracy in sensory discrimination to general intelligence. *The American Journal of Psychology, 20(3)*, 364-369.

Thurstone, L. L. (1924/1973). *The Nature of Intelligence.* London: Routledge.

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review, 41(1)*, 1-32.

Thurstone, L. L. (1935). *The Vectors of Mind: Multiple-factor Analysis for the Isolation of Primary Traits*. Chicago: University of Chicago Press.

Thurstone, L. L. (1937a). Current misuse of the factorial methods. *Psychometrika, 2(2)*, 73-76.

Thurstone, L.L. (1937b). *The Reliability and Validity of Tests*. Ann Arbor: Edwards.

Thurstone, L. L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple-factor Analysis; A Development and Expansion of The Vectors of Mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (Ed.). (1952). *Applications of Psychology*. Oxford: Harper & Brothers.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4), 401-419.

Townsend, J. T. (2008). Mathematical psychology: Prospects for the 21st century: A guest editorial. *Journal of Mathematical Psychology*, 52(5), 269-280.

Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry, 16(10)*, 837-851.

Troelstra, A. S. (1991). A history of constructivism in the 20th century. *University of Amsterdam, ITLI Prepublication Series ML-91-05*, 1-32.

Trusted, J. (1979). The Logic of Scientific Inference: An Introduction. London: Macmillan International Higher Education.

Vallor, S., Green, B., & Raicu, I. (2018). Ethics in Technology Practice. The Markkula Center for Applied Ethics at Santa Clara University. Retrieved from https://www.scu.edu/ethics/on 7 April 2019.

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences, 113(23)*, 6454-6459.

van der Maas, H. L., Kan, K. J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence, 2(1)*, 12-15.

van Rijn, P., & Rijmen, F. (2015). On the explaining-away phenomenon in multivariate latent variable models. *British Journal of Mathematical and Statistical Psychology, 68(1)*, 1-22.

van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. Journal of Experimental Social Psychology, 67, 2-12.

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology, 3(1)*, 1-5.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science., 13(4)*, 411-417.

Vessonen, E. (2017). Psychometrics versus representational theory of measurement. *Philosophy of the Social Sciences, 47(4-5)*, 330-350.

Von Mises, R. (1928/1957). *Probability, Statistics, and Truth.* London: George Allen and Unwin.

Wagner, R. (2017). *Making and Breaking Mathematical Sense: Histories and Philosophies of Mathematical Practice.* Princeton University Press.

Wai, J., & Halpern, D. F. (2018). The Impact of Changing Norms on Creativity in Psychological Science. *Perspectives on Psychological Science, 13(4)*, 466-472.

Wainer, H., & H. Robinson, D. (2007). Roderick P. McDonald. *Journal of Educational and Behavioral Statistics, 32(3)*, 315–332.

Wakefield, J. C. (1992). Freud and cognitive psychology: The conceptual interface. In (J.W. Barrow, M.N. Eagle & D.L. Wolitzky, Eds.), *Interface of Psychoanalysis and Psychology*, (pp. 77-98). Washington DC: American Psychological Association.

Walton, D. N. (1994). Begging the question as a pragmatic fallacy. *Synthese, 100(1)*, 95-131.

Wankhade, L., & Dabade, B. M. (2006). Analysis of quality uncertainty due to information asymmetry. *International Journal of Quality & Reliability Management, 23(2)*, 230-241.

Wardenaar, K. J., van Veen, T., Giltay, E. J., de Beurs, E., Penninx, B. W., & Zitman, F.

G. (2010). Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). *Psychiatry Research, 179(1),* 101-106.

Wardenaar, K. J., Vreeburg, S. A., van Veen, T., Giltay, E. J., Veen, G., Penninx, B. W., & Zitman, F. G. (2011). Dimensions of depression and anxiety and the hypothalamo-pituitary-adrenal axis. *Biological Psychiatry, 69(4),* 366-373.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology, 104(1),* 3-14.

Watson , D. , Clark , L. A. , Weber , K. , Assenheimer , J. S. , Strauss , M. E. , & McCormick , R. A. (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of Abnormal Psychology, 104(1),* 15–25.

Weber, E.H. (1965/1834). De pulsu, resorptione, auidtu et tactu: annotationes anatomicae et physiologicae. *Source Book in the History of Psychology* (R. Hernstein and E. Boring, Eds.). Cambridge: Harvard University Press, pp. 64-66.

Weber, E.H. (1965/1846). Der Tastsinn und das Gemeingefuhl. *Source Book in the History of Psychology* (R. Hernstein and E. Boring, Eds.). Cambridge: Harvard University Press, pp. 34-39.

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist, 5(3),* 78-83.

Weinberger, N. (2015). If intelligence is a cause, it is a within-subjects cause. *Theory & Psychology, 25(3),* 346-361.

Wells, R. S. (1951). Frege's ontology. *The Review of Metaphysics, 4(4),* 537-573.

Welton, J. (1896). *A Manual of Logic (Vol. 1)*. London: WB Clive.

Weyl, H. (1922). *Space--Time--Matter*. New York: Dutton.

Weyl, H. (1951). A half-century of mathematics. *The American Mathematical Monthly, 58(8),* 523-553.

Whitely, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93(1),* 179-197.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7*, 1832.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *Ariel, 165,* 220-14.

Wiecki, T. V., Poland, J., & Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. *Clinical Psychological Science, 3(3),* 378-399.

Wilson, E. B. (1928). The abilities of man, their nature and measurement. *Science, 67,* 244-248.

Wilson, E. B. (1929). Comment on Professor Spearman's note. *Journal of Educational Psychology, 20(3),* 217-223.

Wilson, E. B. (1933). On overlap. *Proceedings of the National Academy of Sciences of the United States of America, 19(12)*, 1039.

Wilson, E. B., & Gibbs, J. W. (1901). *Vector Analysis: A Text-Book For The Use Of Students Of Mathematics & Physics: Founded Upon The Lectures Of JW Gibbs*. Cambridge, MA: University Press.

Wilson, M. (2010). Frege's mathematical setting. In M. Potter and T. Ricketts (Eds.). T*he Cambridge Companion to Frege (*pp. 379–412). Cambridge: Cambridge University Press.

Wilson, M. (2013a). Seeking a balance between the statistical and scientific elements in psychometrics. *Psychometrika, 78(2),* 211-236.

Wilson, M. (2013b). Using the concept of a measurement system to characterize measurement models used in psychometrics. *Measurement, 46(9)*, 3766-3774.

Wisker, G. (2005). *The Good Supervisor: Supervising Postgraduate and Undergraduate Research for Doctoral Theses and Dissertations.* London: Palgrave Macmillan.

Witherington, D. C., Overton, W. F., Lickliter, R., Marshall, P. J., & Narvaez, D. (2018). Metatheory and the primacy of conceptual analysis in developmental science. *Human Development, 61(3),* 181-198.

Wold, H. (1954). Causality and econometrics. *Econometrica: Journal of the Econometric Society,* 162-177.

Wolfe, D. (1940). *Factor Analysis to 1940.* Psychometric Monographs. No. 3.

Woods, J., & Walton, D. (1982). The petitio: Aristotle's five ways. *Canadian Journal of Philosophy, 12(1),* 77-100.

Wright, S. (1918). On the nature of size factors. *Genetics, 3*, 367-74.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20,* 557–585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161-215.

Yang, I., & Becker, M. P. (1997). Latent variable modeling of diagnostic accuracy. *Biometrics,* 948-958.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12(6),* 1100-1122.

Yule, G. U. (1907). On the theory of correlation for any number of variables, treated by a new system of notation. *Proceedings of the Royal Society of London, 79(529),* 182-193.

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and loopholes. *European Journal of Psychological Assessment, 31*, 231–237.

Appendix A

Table 1

Comparison between conceptual framework, Spearman's empirical clarifications, and pre-registration processes

Conceptual Framework example			**Spearman**	**OSF social psychology pre-registration template**		**OSF General pre-registration template**	
U = Universe/ Domain	**A = Description**	**V = Variables**	**1904 paper**	**Location: https://osf.io/k5wns/**		**Location: https://osf.io/prereg/**	
Title Element						**1**	**Title (required)**
Universe in which title applies eg field of journal, university dept	The actual title of your study	Any variations on title (eg for different publications)				1.1.	Provide the working title of your study. It may be the same title that you submit for publication of your final manuscript, but it is not a requirement.
						1.2.	Example: Effect of sugar on brownie tastiness.
						1.3.	More info: The title should be a specific and informative description of a project. Vague titles such as 'Fruit fly preregistration plan' are not appropriate.
Researcher element							
Context of involvement in research project production (eg university affiliation)	Researcher name	Ideological, philosophical, institutional or other commitments				**2**	**Authors (required)**
Research situation element						**3**	**Description (optional)**
Field in which this study takes place including background (eg type of psychology)	Defining details of substantive research project situation, purpose, type of study	Features of the study situation (parameters that may otherwise have been different)	**Environmental conditions of study**			3.1.	Please give a brief description of your study, including some background, the purpose of the of the study, or broad research questions.
		Population parameters				3.2.	Example: Though there is strong evidence to suggest that sugar affects taste preferences, the effect has never been demonstrated in brownies. Therefore, we will measure taste preference for four different levels of sugar concentration in a standard brownie recipe to determine if the effect exists in this pastry.

Conceptual Framework example			Spearman	OSF social psychology pre-registration template		OSF General pre-registration template	
U = Universe/ Domain	**A = Description**	**V = Variables**	**1904 paper**	**Location: https://osf.io/k5wns/**		**Location: https://osf.io/prereg/**	
		Sample parameters - selcection, termination				3.3.	More info: The description should be no longer than the length of an abstract. It can give some context for the proposed study, but great detail is not needed here for your preregistration.
Theory element							
Context of theory eg field of study	Theory description	Variables identified for this theory					
Hypotheses element				**A. Hypotheses**		**4**	**Hypotheses (required)**
Theoretical or substantive context	Hypothesis/ Hypotheses	Variables as relevant to specific hypothesis, directionality, prediction		1	Describe the (numbered) hypotheses in terms of relationships between your variables.	4.1.	List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.
		Interactions		2	For interaction effects, describe the expected shape of the interactions.	4.2.	Example: If taste affects preference, then mean preference indices will be higher with higher concentrations of sugar.
		Manipulations		3	If you are manipulating a variable, make predictions for successful check variables or explain why no manipulation check is included.		
				4	A figure or table may be helpful to describe complex interactions.	**Design Plan**	
				5	For original research, add rationales or theoretical frameworks for why a certain hypothesis is tested.	**5**	**Study type (required)**
Model element				6	If multiple predictions can be made for the same IV-DV combination, describe what outcome would be predicted by which theory.	5.1.	Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.
Context of theory or hypothesis - is this a substantive, statistical, or data model	Definition of model	Variables relevant to model (eg as defined in phenomena or variables element)				5.2.	Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, ñnatural experiments,î and regression discontinuity designs.
						5.3.	Meta-Analysis - A systematic review of published studies.
Phenomenon element						5.4.	Other

Conceptual Framework example		
U = Universe/ Domain	**A = Description**	**V = Variables**
Context of phenomenon (eg research situation, theory, model)	Definition of phenomenon	Variables relevant to phenomenon (eg test items and calculation method, influence factors)
Variables element		
Context of application for variable	Definition of variable	Range of values and calculation method; manipulated versus non-manipulated, moderator or mediator
Data element		
Research situation, model, variable, or phenomena for which the data is collated.	Data description	Aspects of data eg type of information
		Aspects of data collection eg ESM, paid Mturk.

Spearman
1904 paper
Variables included and definition
Variables excluded

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	
B. Method	
Design	**List, based on your hypotheses from section A:**
1	Independent variables and all their levels
a.	whether they are within- or between-participants;
b.	the relationship between them (e.g., orthogonal, nested).
2	Dependent variables.
3	Third variables acting as covariates or moderators.
Planned sample	
4	If applicable, describe pre-selection rules.
5	Indicate where, from whom and how the data will be collected.

OSF General pre-registration template	
Location: https://osf.io/prereg/	
6	**Blinding (required)**
6.1.	Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.
6.1.1.	No blinding is involved in this study.
6.1.2.	For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
6.1.3.	Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as "double blind")
6.1.4.	Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.
7	**Is there any additional blinding in this study?**
8	**Study design (required)**

Conceptual Framework example		
U = Universe/ Domain	**A = Description**	**V = Variables**
		Transformation, exclusion, missing data
Quality control element		
Application of quality framework context (eg part of research situation, separate organisation)	Description of quality framework (eg reliability, validity, credibility, instrument checks)	Any variables related to quality uncertainty eg identity of raters, institutional commitments of raters, etc
Exclusion criteria element		
Data would be the likely domain for exclusion criteria	Description of exclusion features	Any variables related to exclusion criteria

Spearman
1904 paper
Reliability
Validity
Instrumental certainty

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	
6	Justify planned sample size.
7	Describe data collection termination rule.
Exclusion criteria	
8	Describe anticipated data exclusion criteria.
Some examples of exclusion criteria are:	
a.	missing, erroneous, or overly consistent responses;
b.	failing check-tests or suspicion probes;
c.	demographic exclusions;
d.	data-based outlier criteria;

OSF General pre-registration template	
Location: https://osf.io/prereg/	
8.1.	Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.
8.2.	Example: We have a between subjects design with 1 factor (sugar by mass) with 4 levels.
8.3.	More info: This question has a variety of possible answers. The key is for a researcher to be as detailed as is necessary given the specifics of their design. Be careful to determine if every parameter has been specified in the description of the study design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. For example, if the study design describes a complete factorial, 2 X 3 design and the treatments and levels are specified previously, you do not have to repeat that information.
9	**Randomization (optional)**
9.1.	If you are doing a randomized study, how will you randomize, and at what level?
9.2.	Example: We will use block randomization, where each participant will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.
9.3.	More info: Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.
Sampling Plan	

Conceptual Framework example			Spearman	OSF social psychology pre-registration template		OSF General pre-registration template	
U = Universe/ Domain	**A = Description**	**V = Variables**	**1904 paper**	**Location: https://osf.io/k5wns/**		**Location: https://osf.io/prereg/**	
				e.	method-based outlier criteria (e.g. too short or long response times).	In this section we'll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.	
				9	Set fail-safe levels of exclusion at which the whole study needs to be stopped, altered, and restarted.		
Study design element						**10**	**Existing data (required)**
Study type eg experiment, observation, meta-analysis	Describe design eg two-group, within/between subjects, cross sectional, blinding etc	Aspects of design relevant to particular variables eg cohort specifications		**Procedure**		10.1.	Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).
				10	Describe all manipulations, measures, materials and procedures including the order of presentation and the method of randomization and blinding (e.g., single or double blind), as in a published Methods section.	10.1.1.	Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.
Existing data element						10.1.2.	Registration prior to any human observation of the data: As of the date of submission, the data exist but have not yet been quantified, constructed, observed, or reported by anyone - including individuals that are not associated with the proposed study. Examples include museum specimens that have not been measured and data that have been collected by non-human collectors and are inaccessible.
The domain may be a variable, phenomenon, or data, as appropriate to project	Description of data	Variables relevant to the origins of the data set and any discrepancies for the existing project		**C. Analysis plan**		10.1.3.	Registration prior to accessing the data: As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution.

Conceptual Framework example		
U = Universe/ Domain	**A = Description**	**V = Variables**
Procedure element		
Research situation	Description of procedures for research situation	Any variables related to procedures

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	
Confirmatory analyses	**Describe the analyses that will test each main prediction from the hypotheses section. For each one, include:**
1	the relevant variables and how they are calculated;
2	the statistical technique;
3	each variable's role in the technique (e.g., IV, DV, moderator, mediator, covariate);
4	rationale for each covariate to be used, if any;
5	if using techniques other than null hypothesis testing (for example, Bayesian statistics), describe your criteria and inputs towards making an evidential conclusion, including prior values or distributions.

OSF General pre-registration template	
Location: https://osf.io/prereg/	
10.1.4.	Registration prior to analysis of the data: As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.
10.1.5.	Registration following analysis of the data: As of the date of submission, you have accessed and analyzed some of the data relevant to the research plan. This includes preliminary analysis of variables, calculation of descriptive statistics, and observation of data distributions. Please see cos.io/prereg for more information.
11	**Explanation of existing data (optional)**
11.1.	If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.
11.2.	Example: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.
11.3.	More info: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.

Conceptual Framework example			**Spearman**	**OSF social psychology pre-registration template**		**OSF General pre-registration template**	
U = Universe/ Domain	**A = Description**	**V = Variables**	**1904 paper**	**Location: https://osf.io/k5wns/**		**Location: https://osf.io/prereg/**	
				Specify contingencies and assumptions, such as:			
Contingencies element				6	method of correction for multiple tests;		
Domain as relevant (for example, research situation, variable, or data)	Description of contingency method or plan	Any variables related to contingencies		7	the method of missing data handling (e.g., pairwise or listwise deletion, imputation, interpolation);	**12**	**Data collection procedures (required)**
Data collection procedures element				8	reliability criteria for item inclusion in scale;	12.1.	Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that donÍt include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.
Domain as relevant (for example, research situation, variable, or data)	Description of data collection procedures	Any variables related to data collection procedures		9	anticipated data transformations;	12.2.	Example: Participants will be recruited through advertisements at local pastry shops. Participants will be paid $10 for agreeing to participate (raised to $30 if our sample size is not reached within 15 days of beginning recruitment). Participants must be at least 18 years old and be able to eat the ingredients of the pastries.
Exploratory analysis element				10	assumptions of analyses and plans for alternative/corrected analyses if each assumption is violated.	12.3.	More information: The answer to this question requires a specific set of instructions so that another person could repeat the data collection procedures and recreate the study population. Alternatively, if the study population would be unable to be reproduced because it relies on a specific set of circumstances unlikely to be recreated (e.g., a community of people from a specific time and location), the criteria and methods for creating the group and the rationale for this unique set of subjects should be clear.
Domain as relevant (for example, research situation or model)	Description of exploratory analysis	Any variables related to exploratory analysis					
						13	**Sample size (required)**

Conceptual Framework example		
U = Universe/ Domain	A = Description	V = Variables

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
13.1.	Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?
13.2.	Example: Our target sample size is 280 participants. We will attempt to recruit up to 320, assuming that not all will complete the total task.
13.3.	More information: For some studies, this will simply be the number of samples or the number of clusters. For others, this could be an expected range, minimum, or maximum number.
14	**Sample size rationale (optional)**
14.1.	This could include a power analysis or an arbitrary constraint such as time, money, or personnel.
14.2.	Example: We used the software program G*Power to conduct a power analysis. Our goal was to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha error probability.
14.3.	More information: This gives you an opportunity to specifically state how the sample size will be determined. A wide range of possible answers is acceptable; remember that transparency is more important than principled justifications. If you state any reason for a sample size upfront, it is better than stating no reason and leaving the reader to "fill in the blanks." Acceptable rationales include: a power analysis, an arbitrary number of subjects, or a number based on time or monetary constraints.
15	**Stopping rule (optional)**
15.1.	If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.
15.2.	Example: We will post participant sign-up slots by week on the preceding Friday night, with 20 spots posted per week. We will post 20 new slots each week if, on that Friday night, we are below 320 participants.

Conceptual Framework example		
U = Universe/ Domain	A = Description	V = Variables

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
15.3.	More information: You may specify a stopping rule based on p-values only in the specific case of sequential analyses with pre-specified checkpoints, alphas levels, and stopping rules. Unacceptable rationales include stopping based on p-values if checkpoints and stopping rules are not specified. If you have control over your sample size, then including a stopping rule is not necessary, though it must be clear in this question or a previous question how an exact sample size is attained.
Variables	
In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.	
16	Manipulated variables (optional)
16.1.	Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.
16.2.	Example: We manipulated the percentage of sugar by mass added to brownies. The four levels of this categorical variable are: 15%, 20%, 25%, or 40% cane sugar by mass.
16.3.	More information: For any experimental manipulation, you should give a precise definition of each manipulated variable. This must include a precise description of the levels at which each variable will be set, or a specific definition for each categorical treatment. For example, "loud or quiet," should instead give either a precise decibel level or a means of recreating each level. 'Presence/absence' or 'positive/negative' is an acceptable description if the variable is precisely described.
17	**Measured variables (required)**
17.1.	Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.

Conceptual Framework example		
U = Universe/ Domain	**A = Description**	**V = Variables**

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
17.2.	Example: The single outcome variable will be the perceived tastiness of the single brownie each participant will eat. We will measure this by asking participants 'How much did you enjoy eating the brownie' (on a scale of 1-7, 1 being 'not at all', 7 being 'a great deal') and 'How good did the brownie taste' (on a scale of 1-7, 1 being 'very bad', 7 being 'very good').
17.3.	More information: Observational studies and meta-analyses will include only measured variables. As with the previous questions, the answers here must be precise. For example, 'intelligence,' 'accuracy,' 'aggression,' and 'color' are too vague. Acceptable alternatives could be 'IQ as measured by Wechsler Adult Intelligence Scale' 'percent correct,' 'number of threat displays,' and 'percent reflectance at 400 nm.'
18	**Indices (optional)**
18.1.	If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.
18.2.	Example: We will take the mean of the two questions above to create a single measure of 'brownie enjoyment.'
18.3.	More information: If you are using multiple pieces of data to construct a single variable, how will this occur? Both the data that are included and the formula or weights for each measure must be specified. Standard summary statistics, such as "means" do not require a formula, though more complicated indices require either the exact formula or, if it is an established index in the field, the index must be unambiguously defined. For example, "biodiversity index" is too broad, whereas "Shannon's biodiversity index" is appropriate.
Analysis Plan	
You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.	

Conceptual Framework example		
U = Universe/ Domain	A = Description	V = Variables

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.	
19	**Statistical models (required)**
19.1.	What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions, subgroup analyses, pairwise or complex contrasts, or follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Remember that any test not included here must be noted as an exploratory test in your final article.
19.2.	Example: We will use a one-way between subjects ANOVA to analyze our results. The manipulated, categorical independent variable is 'sugar' whereas the dependent variable is our taste index.
19.3.	More information: This is perhaps the most important and most complicated question within the preregistration. As with all of the other questions, the key is to provide a specific recipe for analyzing the collected data. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? Be aware for instances where the statistical models appear specific, but actually leave openings for the precise test. See the following examples:
19.3.1.1.	If someone specifies a 2x3 ANOVA with both factors within subjects, there is still flexibility with the various types of ANOVAs that could be run. Either a repeated measures ANOVA (RMANOVA) or a multivariate ANOVA (MANOVA) could be used for that design, which are two different tests.
19.3.1.2.	If you are going to perform a sequential analysis and check after 50, 100, and 150 samples, you must also specify the p-values you'll test against at those three points.
20	**Transformations (optional)**

Conceptual Framework example		
U = Universe/ Domain	**A = Description**	**V = Variables**

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
20.1.	If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.
20.2.	Example: The "Effect of sugar on brownie tastiness" does not require any additional transformations. However, if it were using a regression analysis and each level of sweet had been categorically described (e.g. not sweet, somewhat sweet, sweet, and very sweet), 'sweet' could be dummy coded with 'not sweet' as the reference category.
20.3.	More information: If any categorical predictors are included in a regression, indicate how those variables will be coded (e.g. dummy coding, summation coding, etc.) and what the reference category will be.
21	**Inference criteria (optional)**
21.1.	What criteria will you use to make inferences? Please describe the information youÍll use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will your account for this?
21.2.	Example: We will use the standard p<.05 criteria for determining if the ANOVA and the post hoc test suggest that the results are significantly different from those expected if the null hypothesis were correct. The post-hoc Tukey-Kramer test adjusts for multiple comparisons.
21.3.	More information: P-values, confidence intervals, and effect sizes are standard means for making an inference, and any level is acceptable, though some criteria must be specified in this or previous fields. Bayesian analyses should specify a Bayes factor or a credible interval. If you are selecting models, then how will you determine the relative quality of each? In regard to multiple comparisons, this is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.
22	**Data exclusion (optional)**

Conceptual Framework example		
U = Universe/ Domain	A = Description	V = Variables

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
22.1.	How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled? Will you use any awareness check?
22.2.	Example: No checks will be performed to determine eligibility for inclusion besides verification that each subject answered each of the three tastiness indices. Outliers will be included in the analysis.
22.3.	More information: Any rule for excluding a particular set of data is acceptable. One may describe rules for excluding a participant or for identifying outlier data.
23	**Missing data (optional)**
23.1.	How will you deal with incomplete or missing data?
23.2.	Example: If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.
23.3.	More information: Any relevant explanation is acceptable. As a final reminder, remember that the final analysis must follow the specified plan, and deviations must be either strongly justified or included as a separate, exploratory analysis.
24	**Exploratory analysis (optional)**
24.1.	If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.
24.2.	Example: We expect that certain demographic traits may be related to taste preferences. Therefore, we will look for relationships between demographic variables (age, gender, income, and marital status) and the primary outcome measures of taste preferences.
Other	
25	**Other (Optional)**

Conceptual Framework example		
U = Universe/ Domain	A = Description	V = Variables

Spearman
1904 paper

OSF social psychology pre-registration template	
Location: https://osf.io/k5wns/	

OSF General pre-registration template	
Location: https://osf.io/prereg/	
25.1.	If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.