THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

**Bridging the theory to evidence gap: a systematic review and analysis of individual × environment models of child development**


Peter Sheldon Rankin

B. Sc. (Hons)


*A thesis submitted for the degree of Doctor of Philosophy at*

*The University of Queensland in 2019*


Institute for Social Science Research

**Abstract**

A key theoretical principle of developmental psychology is that the response of individuals to developmental experiences can vary. This principle is exemplified in theoretical models of *individual* × *environment* interactions, including diathesis stress, differential susceptivity (biological sensitivity to context), and vantage sensitivity.

Despite a theoretical underpinning and growing empirical base, there is considerable variability in evidence for these theoretical interactions and the literature has not always delivered well with definitive findings towards the provision of evidence-based interventions tailored to meet the needs of individual children.

This thesis has therefore focused on understanding and bridging the gap between theory and evidence for specific *individual* × *environment* explanatory models of child behavioural outcomes. It examined the theory and evidence for moderation of parenting effects by child reactivity using a strategy of targeted review and empirical analysis. The thesis aimed to 1) systematically analyse the current literature to identify any patterns in the use of measurement and methods of analysis that might direct future research into *individual* × *environment* models. And, 2) conduct analyses using existing data (Longitudinal Study of Australian Children, Family Life Project) to explore effects of different measurements on *individual* × *environment* models; systematically examining effects of different measures of parenting, behavioural outcomes, and reactivity. A diagrammatic outline of the thesis is provided in Figure 1.

Findings from both the systematic review and empirical analyses revealed that the gap between theory and evidence is extensive. A review of 542 *individual* × *parenting* interactions found that 86 were described as statistically reliable. However, within- and between-studies, these reliable interactions inconsistently supported different theoretical models and sporadically varied with measures of parenting, behaviour, and individual characteristics.

Two empirical analysis chapters confirmed these inconsistencies. The first empirical study examined the interaction between parent-reported parenting (warmth and harshness) and parent-reported temperament (persistence, introversion, reactivity, difficult composite) predicting parent and teacher reported behavioural outcomes (internalizing, externalizing, pro-social) in a longitudinal sample of 1289 Australian children at ages 4 and 6, and across ages 4, 6, and 8. By transparently examining the full matrix of analytical decisions, the results showed that 13 (2.5%) of 512 interactions were statistically reliable and that temperament moderated the effects of parenting on behaviour. However, these reliable interactions did not consistently support a theoretical model

and varied with measures of parenting, temperament, and behaviour, the age of the child, and the reporter of behaviour.

The second empirical chapter extended this analysis by considering four additional measurements and examining an additional data set of children from the United States of America. Specifically, the analysis examined interactions between parent-reported parenting at 2 years (warmth, hostile) and parent-reported temperament at 1 year (approach, cooperation, irritability, difficult composite) predicting parent reported behaviour at 4 years (internalizing, externalizing, pro-social) using a longitudinal sample of 3062 Australian children. Additionally, the analysis explored interactions between observed parenting at 15 months and 2 years (sensitive composite, negative-intrusive composite, positive regard, negative regard, sensitivity, animation, stimulation, detachment, intrusiveness) and parent-reported temperament at 6 months (fear, distress to limitations, falling reactivity, negative affectivity composite, duration of orienting, approach) predicting parent reported behaviour at 3 years (social and emotional competence, internalizing, externalizing, pro-social) using a longitudinal sample of 1093 children from the United States of America. Utilising a transparent analytical approach, the results found 2 (3%) of 64 interactions were statistically reliable in the Australian sample, whilst there were no statistically reliable interactions (of 1620) in the sample from the United States of America. As with the earlier chapter, the two statistically reliable interactions supported different theoretical models and there was no consistency in interaction effects across measures of temperament, behaviour, or parenting. Thus, in total, these results and the systematic review demonstrate that theory and evidence do not consistently align for *individual × environment* models of development.

To better align theory and evidence for *individual × environment* models, the thesis concludes with several suggestions for future research and an integrative discussion of the results. Suggestions for future research include a detailed focus on measurement, extensive validation and exploration of analytical decisions, the use of optimum and more causal research designs to elucidate interaction effects, the use of simulation studies to understand the implications of design, and critically assessing and accurately portraying results.

*Individual × environment* interactions remain a key principle of developmental psychology. Though this thesis has demonstrated a gap between theory and evidence, improvement in research design and measurement may align theory with evidence and yield insights that can assist the behavioural development of children.

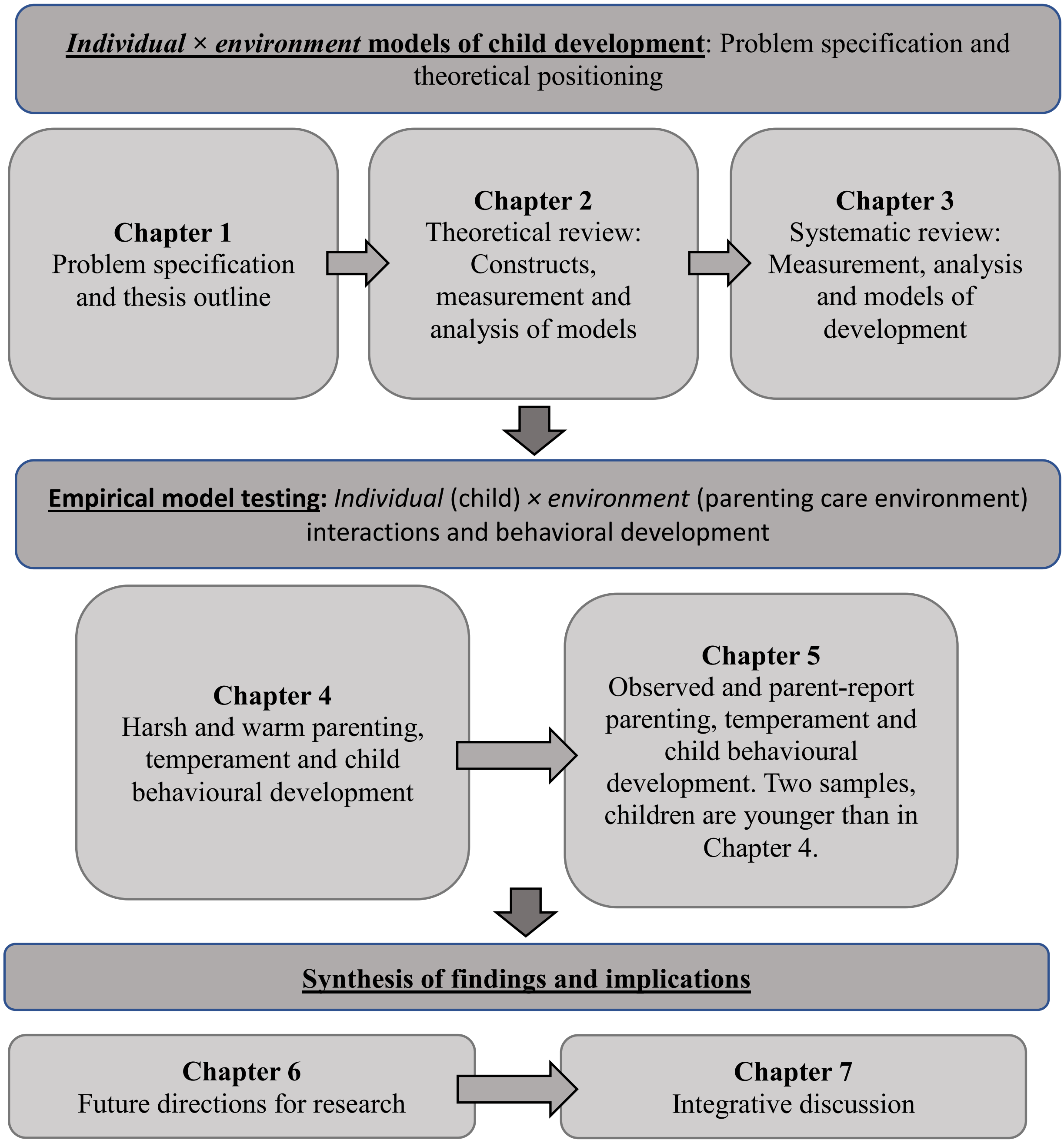



**Figure 1:** Outline of thesis structure.

**Declaration by author**

This thesis *is composed of my original work, and contains* no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted *to qualify for the award of any* other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

**Publications included in this thesis**

No publications included.

**Submitted manuscripts included in this thesis**

No manuscripts submitted for publication.

**Other publications during candidature**

**Peer reviewed scientific publications**

McClanahan, T. R., & Rankin, P. S. (2016). Geography of conservation spending, biodiversity, and culture. Conservation Biology, 30(5), 1089-1101.

Rankin, P. S., & Lemos, R. T. (2015). An alternative surplus production model. Ecological modelling, 313, 109-126

**Conference abstracts**

Rankin, P. S., Staton, S., Thorpe, K., & Smith, S. 2018. Actigraphy-based estimates of sleep duration in young children with and without a sleep log are comparable. Australasian Chronobiology Society 15th Annual Scientific Meeting, October 16, 2018, Brisbane, Australia.

Rankin, P. S., Staton, S., Beatton, T., & Thorpe, K. 2019. Variation and average teaching quality as a predictor of academic outcomes: extending CLASS. Society for Research in Child Development, Poster Symposium, Biennial Meeting, March 21 - 23, 2019, Baltimore, Maryland, USA.

Rankin, P. S., Western, M., Haynes, M., & Thorpe, K. 2019. Measurement, analyses and patterns of outcome: Systematic review of child by environment models of development. Society for Research in Child Development, Poster Symposium, Biennial Meeting, March 21 - 23, 2019, Baltimore, Maryland, USA.

Rankin, P. S., Western, M., Haynes, M., & Thorpe, K. (2018). Recommendations for assessing differential susceptibility with secondary data. Society for Research in Child Development, Paper Symposium: Studying Biosocial Developmental Processes with Secondary and Administrative Data. Conference on the Use of Secondary and Open Source Data in Developmental Science, October 4 - 6, 2018, Phoenix, Arizona, USA.

**Contributions by others to the thesis**

Professor Karen Thorpe, Professor Michele Haynes, and Professor Mark Western helped with critical revision of the thesis to contribute to the interpretation and presentation.

**Statement of parts of the thesis submitted to qualify for the award of another degree**

No works submitted towards another degree have been included in this thesis.

**Research Involving Human or Animal Subjects**

This project has been reviewed by the Office of Research Ethics and is deemed to be exempt from ethics review under the National Statement on Ethical Conduct in Human Research and University of Queensland policy. Ethics approval number 2019001822. See Appendix C.

**Acknowledgements**

I would like to sincerely thank and acknowledge the assistance of everyone who has helped, knowingly or otherwise, me to complete this project. I am indebted to Karen Thorpe, Michele Haynes, and Mark Western who were my supervisory team and offered unending and exceptional support. I am incredibly grateful for their generous investment in my development and for providing funding so I could attend conferences and visit international research institutes. I vastly appreciate the guidance and support from Sally Staton who has been a sounding board for many ideas and a source of encouragement and mentorship. I had a great time working with all the PhD candidates who came through the Institute for Social Science Research and I appreciate their companionship. Special mention must be given to Laetitia Coles and Angelina Tang who heard, experienced, and lived the existential wave of the PhD process and I am grateful for their support and friendship. The research staff at the Institute for Social Science Research and University of Queensland have also been an immense pleasure to work with. Many supported my development along the way, and I would like to particularly acknowledge Melanie Spallek, Simon Smith, Renee Zahnow, Francisco Perales, Jenny Povey, and Yangtao Huang for providing formal development and employment opportunities throughout my candidature. I would also like to thank Daniel Berry, Megan Gunnar, Ann Masten and the other exceptional researchers and PhD candidates at the University of Minnesota, Institute of Child Development for hosting me on two occasions and providing generous advice. Alyssa Varhol, Keira Leneman, and Alyssa Palmer also deserve special mention for their friendship and support. I am very appreciative to have been part of the Australian Research Council Centre of Excellence for Families over the Life Course and the opportunities that has provided. I would also like to thank the professional staff and graduate advisors at the Institute for Social Science Research for their generous support, with particular mention to Lisa Pope for her exceptional dedication to maximising the graduate experience. I would also like to thank the Australian citizens and government for their continued funding and support of tertiary education. Their assistance made this research possible. I also thank all those involved in the research presented herein for their time and effort in collecting, analysing, and sharing data and knowledge. I am indebted to the Australian, United States of America', and international citizens who agreed to participate in the research studies. Their willingness to contribute their time and share insights into their world is inspiring and incredibly valuable. Finally, my wife Bridie, my family, and my friends were integral sources of support to complete the project and I appreciate and love them greatly.

**Financial support**

This research was supported by an Australian Government Research Training Program Scholarship.

**Keywords**

child development, child behaviour, parenting interactions, differential susceptibility, diathesis stress, biological sensitivity to context, vantage sensitivity, gene environment interaction and moderation, systematic review, temperament and parenting

**Australian and New Zealand Standard Research Classifications (ANZSRC)**

ANZSRC code: 010401, Applied Statistics 34%

ANZSRC code: 170102, Developmental Psychology and Ageing 33%

ANZSRC code: 170110, Psychological Methodology, Design and Analysis 33%

**Fields of Research (FoR) Classification**

FoR code: 0104, Statistics, 34%

FoR code: 1701, Psychology, 66%

## Table of contents

## List of Figures and Tables

**List of abbreviations**

5-HT 5-Hydroxytryptamine

5-HTT 5-Hydroxytryptamine Transporter gene

ADD Attention Deficit Disorder

ADHD Attention Deficit Hyperactive Disorder

ANS Autonomic Nervous System

APA American Psychological Association

ASD Autism Spectrum Disorder

BDNF Brain-Derived Neurotrophic Factor

CE1 one-way Contrastive Effects

CE2 Two-way Contrastive Effects

CO Crossover point

COMT Catechol-O-Methyltransferase

CRH Corticotrophin-Releasing Hormone

DA Diathesis Stress

DNA Deoxyribonucleic Acid

DS Differential Susceptibility

EDA Electrodermal Activity

FKBP5 FK506-binding protein 51

FLP Family Life Project

GABA Gamma-Aminobutyric Acid

HLE Home Learning Environment

HPA Hypothalamic–Pituitary–Adrenocortical

IBQ-R Infant Behavior Questionnaire Revised

LC Locus Coeruleus

LSAC Longitudinal Study of Australian children

MAOA Monoamine Oxidase A

M Mean

MOR μ-Opioid Receptor

NE Norepinephrine

OTR Oxytocin Receptor gene

OXT Oxytocin Coding gene

PA Proportion Affected

PNS Parasympathetic Nervous System

POI Proportion of the Interaction

POMP Percentage of Maximum Possible

ROS Regions of Significance

RSA Respiratory Sinus Arrhythmia

SCL Skin Conductance Level

SD Standard Deviation

SDQ Strengths and Difficulties Questionnaire

SEIFA Socio-Economic Indexes for Areas

SLE Stressful Life Events

SMC-FCS Substantive Model Compatible Fully Conditional Specification

SNPs Single Nucleotide Polymorphisms

SNS Sympathetic Nervous System

STSC Short Temperament Scale for Children

STSI Short Temperament Scale for Infants

TPH Tryptophan Hydroxylase

VNTR Variable Number of Tandem Repeats

VS Vantage Sensitivity

**Preface**

In 2015, I started a doctorate investigating how global climate change would shift the distribution of marine biological diversity.

Four years later, I am presenting a thesis in developmental psychology, not marine biology. My doctoral project clearly has not followed a linear trajectory. While it comes as some surprise to present the current thesis, that investigates the theory-evidence gap for *individual* × *environment* models in child development, the underlying principles of individual variation in response to the environment underpins my starting point.

Originally, the current thesis was inspired by the idea of resilience. I wanted to know why some children were unaffected by adversity whilst others succumbed and had developmentally poor trajectories. With access to large longitudinal data and the support of Mark Western and Michele Haynes to transfer to the Institute for Social Science Research, I aimed to identify the factors associated with diverging behavioural development in the face of adversity.

After two years and the writing of several chapters, I visited the renowned Institute of Child Development, University of Minnesota. I presented my research to prominent child development researchers (primarily Ann Masten, Megan Gunnar, and Daniel Berry) and absorbed as much knowledge as I could. The experience left me inspired and motivated. Indeed, I decided to re-write the entire thesis.

Fortuitously, I gained a greater understanding of resilience and picked up the latest differential susceptibility and individual differences literature on the research visit. Even better, returning to Australia coincided with Karen Thorpe joining the Institute for Social Science Research. Karen had published on the differential susceptibility and was happy to join my supervisory team as a substantive expert.

With a more nuanced approach to the topic area, I began further exploration of *individual* × *environment* models of child development. I reviewed the literature and learned the latest statistical criteria. Ultimately, I found diverging results and not the clear convergence of findings I had anticipated. My challenge became explaining the gap between a well-established biological theory and its application in *individual* × *environment* models of child development.

Initially, I thought the gap may have been from differences in the measurement approach between studies. Thus, I set out to evaluate how different types of measurement (genetic, endophenotypic, phenotypic, survey methods, and observation methods) might explain the theory-evidence gap. My focus in the systematic review and empirical chapters in this thesis are, therefore, is very much on different types of measurement.

The learning from this thesis, extended upon in the final chapters, is that it seems more likely that the theory-evidence gap is due to more foundational issues than those pertaining to measurement. Indeed, low power to detect interaction effects and validity and reliability of measurement are likely explanations for the gap between theory and evidence. In hindsight, I would have liked to follow the recommendations for future research I make in the final chapters – to focus on the validation and reliability of measurement, the adequacy (power) of existing literature to detect interactions, and undertaken empirical analysis with (more-) optimum samples.

Thus, this work has been a journey of discovery and learning. I am excited to start the next chapter and continue working towards bridging the gap between theory and evidence for *individual* × *environment* models of child development.

# Chapter 1: The research problem

## 1.1 The significance of understanding the effects of the environment on child development

Optimal developmental outcomes for children are in the interests of the child, their family and the nation's economy (Black et al. 2017). The Australian government specifically expresses this goal asserting *that every student in Australia should have the* [learning environment] *they need to be the best they can be* (https://www.education.gov.au/quality-schools-package, 29/05/2018). Likewise, most parents aim to provide their child with an environment that allows them to achieve their potential. This aim is mapped in public and private investment. Education programs represent a significant investment for governments and families. In the 2016–2017 financial year, Australian governments spent 91.271 billion dollars (AU) on education (Australian Bureau of Statistics 2018a). Alongside, Australian parents invested their time (e.g., >7 hours on average per day caring for children aged 0-2; Craig & Bittman 2008) and a significant proportion of their income (13% to 19% of family income for 1 child, 30% to 44% for 4 children; Henman et al. 2007) in supporting the health, learning and developmental experiences of their children.

A key focus for parents and policymakers, therefore, is understanding how the allocation of finite financial and personal resources might best support each child's development. This thesis sets out to test key theories or models of the effect of the environment on children's development. The focus is on understanding how neurological reactivity and its impact on the response of children to early care environments might guide optimal resource distribution to improve behavioural outcomes.

## 1.2 Models of the effect of the environment on child development

An early and basic model that guided, and continues to guide, investments in the learning and parenting environment to improve child development is presented in Figure 1.1a. This model proposes that as the quality of the environment improves, the average outcomes for children also improve in a uniform, linear manner. It reflects the theoretical model that child developmental vulnerabilities and advantages are caused by the exposure to risky and nurturing environments, respectively (Bronfenbrenner & Morris 2006).

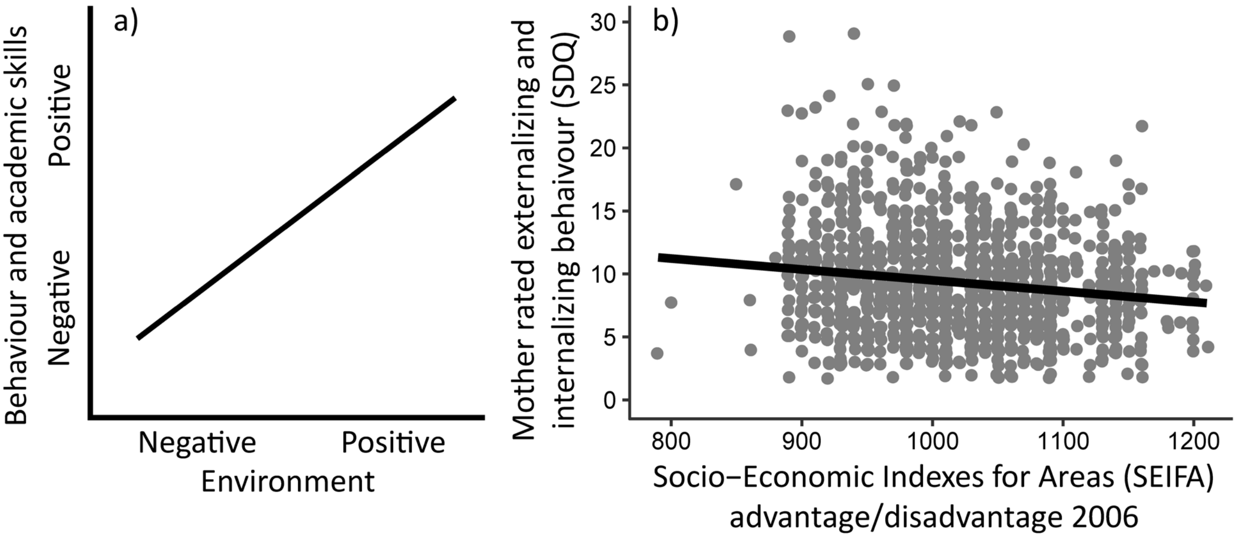



**Figure 1.1**: a) Linear effect of environmental quality on behavioural and academic outcomes. b) Linear relationship between neighbourhood socioeconomic status and mother reported behavioural outcomes (Strength and Difficulties Questionnaire; Goodman 1997) of 1238 children from the Longitudinal Study of Australian Children at age 4.

Although this model continues to guide policy and parental investments (e.g., sending a child to a high-performing or fee-paying Australian school *will* improve academic outcomes; Angus 2015; Smyth 2016; Vincent 2017) the match between investment and outcome can be relatively weak in predictive models (<10% of variance in academic achievement and externalizing and internalizing behaviour explained by parenting on average; Pinquart 2016a, 2016b, 2017). Figure 1.1b, illustrates this by comparing socio-economic status (higher indicates more resources) to behavioural difficulties (lower is better). There is clearly much variation in child outcomes not explained by variation in the socio-economic environment.

The weak association of environmental variation with child outcome may be explained in three ways. One explanation is that environments are inaccurately or poorly measured thus missing or minimising the full impact. A second is that characteristics of the children cause the developmental outcome irrespective of the quality of the chosen environment that was measured. In other words, vulnerabilities and advantages may be due to characteristics within the child, such as genetic and phenotypic traits. Finally, the third and more comprehensive view is that the poor average gains reflect differential response to a standard environment. A detailed field of scientific inquiry suggests that the environment can affect individual children differently than a uniform average trend line might suggest. In terms of policy or practice effects, the fit of the environment to individual children becomes critical.

## 1.3 What is individual variation to the environment?

Individual difference in response to a standard environment may occur when characteristics of the individual influence whether they are more or less sensitive to that environment and for more sensitive individuals the influence exerted on developmental outcomes is greater. This represents the theoretical model that vulnerabilities and advantages in child development are the outcome of child characteristics interacting with the environment. Studies of children provide empirical evidence that genetic, endophenotypic and phenotypic characteristics of children can moderate the influence of the environment on development (Boyce 2016). These studies document that children who have genetic, endophenotypic or phenotypic characteristics associated with higher neural reactivity are those more sensitive, or susceptible, to qualities of the environment and that variation in these qualities will have greater effect on developmental outcome for these individuals. Such findings have generated an ongoing set of hypotheses about the effect of *individual* × *environment* interactions on developmental outcomes. These hypotheses and the ways they are tested are the focus of the current thesis.

## 1.4 How does individual variation to the environment influence child development outcomes?

Social scientists have generally proposed three main models to explain the association of individual variation to the environment and child development outcomes. One focusses on "risk", one on "vantage" and the third on susceptibility to the environment indexed by "reactivity". Respectively, these models are termed diathesis stress (DA), vantage sensitivity (VS) and differential susceptibility (DS) and are illustrated in Figure 1.2. Note that each of these models has a 'strong' and 'weak' form (Belsky et al. 2013). In the strong form, the environment only has an effect conditional on sensitivity. That is, there is negligible effect of the measured environmental quality on non-sensitive individuals. The weak form of these models proposes environmental quality affects sensitive and non-sensitive individuals, but the environment affects sensitive individuals disproportionately more. Additionally, whilst biological or social circumstances experienced by an individual can moderate the effect of the environment, such as exposure to perinatal risk environments or neighbourhood dis/advantage, the focus in this thesis is on characteristics of the individual that index their endogenous neurological reactivity and behaviour.

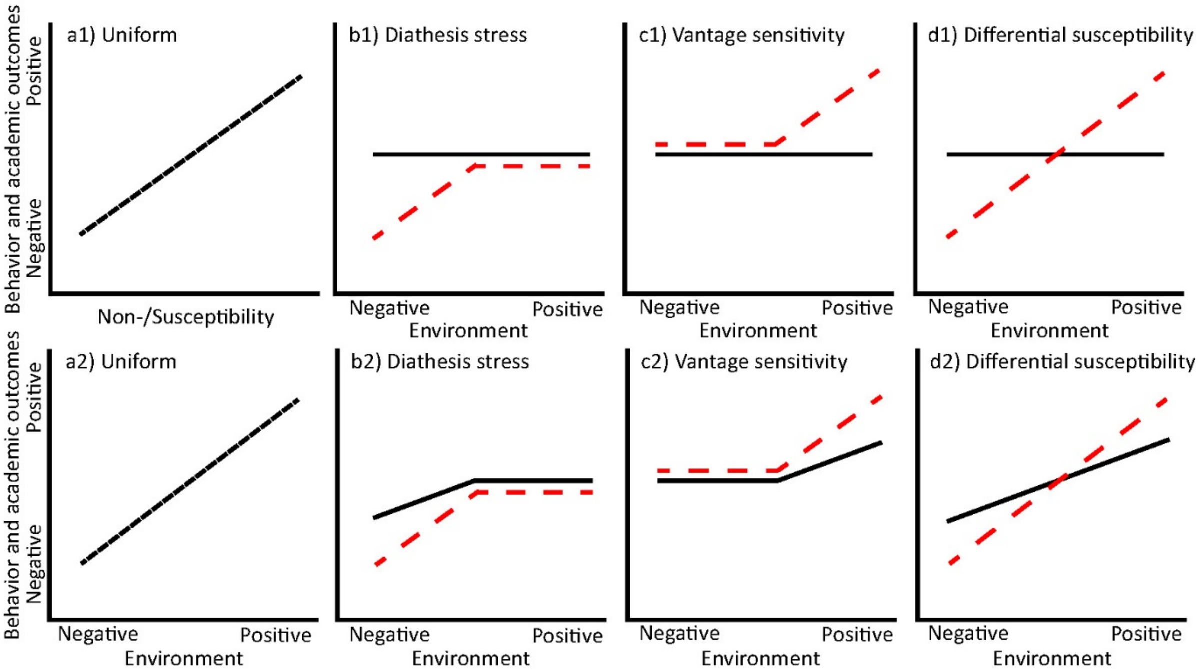



**Figure 1.2:** Four hypotheses for sensitivity to the environment in strong and weak forms: a) uniform, b) diathesis stress, c) vantage sensitivity, and d) differential susceptibility. Children sensitive to the environment are the dashed, red line. Although outcomes are on the *y-axis,* in practice these models are generally comparisons of the slope of the environment. Figure adapted from Pluess & Belsky (2013) and Ellis et al. (2011).

Risk based models are the most dominant model familiar to social scientists (Figure 1.2b1, 1.2b2). Multiple studies observed that some children have individual characteristics associated with vulnerability to developmentally poor environments. This model has variously been explained as diathesis-stress (Gottesman & Shields 1973; Monroe & Simons 1991; Zuckerman 1999), transactional/dual-risk (Sameroff 1983) and cumulative-risk (Rutter 1981; Evans et al. 2013), and suggests that the characteristics of some children cause their developmental path to negatively deviate from less-vulnerable peers when environmental conditions are poor.

Vantage sensitivity (Figure 1.2c) is a recently proposed mirror image to the diathesis stress hypothesis and aims to explain how some children possess characteristics associated with an advantage in positive environments (Manuck 2011; Pluess & Belsky 2013; Sweitzer et al. 2013). Specifically, it implies vantage-sensitive children benefit more from positive environments than vantage-insensitive counterparts, yet reductions in environmental quality do not cause vantage-sensitive individuals to underperform.

The third model is differential susceptibility (Figure 1.2d) and proposes variation associated with both poorer and superior environmental experiences. That is, susceptible children are hypothesised to possess characteristics that make them both more vulnerable to negative environments *and* more vantage-sensitive to positive environments (Belsky 1997, 2005; Belsky & Pluess 2009). Such bi-directional reactivity has also been termed biological sensitivity to context (Boyce et al., 1995; Boyce & Ellis 2005) and sensory processing sensitivity (Aron & Aron 1997; Aron et al. 2012). In these models, susceptible, or sensitive, children are associated with better outcomes in developmentally rich environments and worse outcomes in developmentally poor environments.

Testing the various hypotheses of *individual* × *environment* interactions, however, can only be theoretically meaningful if measurement can distinguish sensitive (responsive) and non-sensitive (non-responsive) individuals and measurement of the environment is sufficiently sensitive (Mitchell et al. 2013). It is important, therefore, to evaluate how measurement quality and type may influence support for these hypotheses.

### 1.5 Testing the theories of individual difference in reaction to environment: measurement of reactivity and statistical modelling

When it comes to evaluating the theories of individual difference to the environment a broad range of measurements have been used. Specific measures are provided in more detail in chapter 2, but it suffices to say that studies have examined genetic variability (e.g., polymorphisms in genes that affect neurotransmitter systems), endophenotypic variability (biological markers easier to detect than genetic sequences, such as heart rate variability and cortisol) and phenotypic variability (observable characteristics of an individual, such as temperament) in search of mechanisms of reactivity or sensitivity (Boyce 2016). These studies then primarily examine how mechanisms of reactivity moderate the effect of a range of parental and nonparental care environments on the behaviour of children and other child development outcomes. Specified in more detail in chapter 2, the aspects of the parental care environments that are measured varies between studies and can rarely be considered truly comparable. These environments range from deprivation (e.g., neglect and hostility) to daily experiences of environmental support (e.g., the home learning environment) and are measured using surveys or in situ observation. Behavioural measurements also differ between studies and use constructs such as externalising behaviours, internalising behaviours, social and emotional competence or a combination of these. Behaviours are primarily assessed via surveys and in some cases observations. Overall, measurement variability is a defining feature of this literature.

One potential consequence of this measurement variability is a divergence of findings. A recent review by Rabinowitz & Drabick (2017) clearly documents such divergence. Using a set of evaluative criteria approximating the state-of-the-art (Roisman et al. 2012; Widaman et al. 2012), they examined 27 empirical studies and 75 interactions that tested models of individual variation to the effects of the parenting environment on behavioural development. They report that 38 (51%) interactions found evidence for diathesis stress, 33 (44%) found evidence for differential susceptibility and 5 (6%) found evidence for vantage sensitivity. These numbers are somewhat optimistic, however, as nearly all 27 studies reported a higher proportion of statistically uncertain interactions not included in the evaluation (Rabinowitz & Drabick 2017). Moreover, the review also showed that there is much measurement variability between studies. Nearly all studies used a different measure of reactivity, a different measure of parenting, and a different outcome. There were some exceptions – e.g., similar parenting, behaviour, or reactivity measure – but few studies used identical measures across samples. Thus, the review demonstrated that there is much variability in the measurements used and, as a result, evidence for the three hypotheses is yet to converge.

Assessing the model of, and evidence for, individual variation to the environment for child development outcomes is challenging due to the inherent complexity of social systems and consequent measurement limitations which reduce the ability to robustly infer the nature of individual variation to the environment across studies that differ in methodological components. The extent of non-convergence in evidence for the theories of individual variation to the environment leaves much room for further exploration of the theoretical model of individual variation in response to environment and identification of the circumstances under which evidence for each emerges (Stoltz et al. 2017). The significant variability in measurement type and form also implies that a systematic construction of evidence will be required to progress the field. With so many moving components created by using different measurements identifying key repeatable findings and mechanisms of gene × environment effects is impeded. Building a systematic base for understanding individual variation to the environment is an important contribution to be made to this literature.

### 1.6 Thesis purpose: a systematic analysis of measurement and analysis of models testing individual × environment models of child development

Despite the evident feasibility and appeal of *individual × environment* models they have to date not delivered well on their promise to inform policy interventions to optimise child development (Belsky & van Ijzendoorn 2015). The field remains restricted by available data and the use of a diverse array of measurements. Therefore, this thesis aimed to

1. Systematically analyse the current literature to identify any patterns in measurement and analysis that might direct future research into *individual × environment* models

2. Conduct analyses using existing data to explore effects of different measurements on *individual* × *environment* models; systematically examining effects of different measures of parenting, behavioural outcomes, and reactivity.

### 1.7 Thesis structure

This thesis focuses on bridging the gap between theory and evidence for *individual* × *environment* explanatory models of child behavioural outcomes. It examines the theory and evidence for child reactivity moderating the effects of parenting using a strategy of targeted review and empirical analysis. The thesis consists of six chapters outlined in Figure 1.3.

**Chapter 2** is a narrative review that summarises key constructs and construct measures used to assess *individual* × *environment* interactions regarding parenting, child reactivity, and behavioural development. It examines the knowledge base and methodological barriers that arise for *individual* × *environment* explanations of child behavioural development.

**Chapter 3** presents a systematic review of studies that have applied *individual* × *environment* methodology to understand how child characteristics moderate the effects of parenting on behavioural development. The review quantifies the measurement and statistical methodology associated with findings consistent with diathesis stress, differential susceptibility and vantage sensitivity. In total, 542 *individual* × *parenting* interactions were examined and 86 were described as statistically reliable. However, within- and between-studies, these reliable interactions inconsistently supported different theoretical models and sporadically varied with measures of parenting, behaviour, and individual characteristics. This inconsistency within and between-studies suggests systematic selection and testing of measures in *individual* × *parenting* interactions may be useful to identify robust areas of evidence. Thus, chapter 4 and 5 present sequential empirical analyses of two longitudinal cohort studies to systematically evaluate multiple analytical choices with the aim of identifying consistent evidence of *individual* × *parenting* interactions.

**Chapter 4** is the first empirical study and examines the interaction between parent-reported parenting (warmth and harshness) and parent-reported temperament (persistence, introversion, reactivity, difficult composite) predicting parent and teacher reported behavioural outcomes (internalizing, externalizing, pro-social) in a longitudinal sample of 1289 Australian children at ages 4 and 6, and across ages 4, 6, and 8. By transparently examining the full matrix of analytical decisions, the results found 13 (2.5%) of 512 interactions statistically reliable and showed that temperament moderated the effects of parenting on behaviour. However, these reliable interactions did not consistently support a theoretical model and varied with measures of parenting, temperament, and behaviour, the age of the child, and the reporter of behaviour. The results demonstrate the utility of exploring a broad range of analytical decisions and highlight the need for additional research that systematically changes measurement to find consistent and robust *individual* × *parenting* interactions.

**Chapter 5** extends the first empirical analysis by making four changes to the measurements used (long and short temperament measures, parent and behavioural measures with more variation, parent-report and observed parenting measures, younger age of temperament assessment) and examining an additional data set of children from the United States of America. Specifically, the analysis examined interactions between parent-reported parenting at 2 years (warmth, hostile) and parent-reported temperament at 1 year (approach, cooperation, irritability, difficult composite) for predicting parent reported behaviour at 4 years (internalizing, externalizing, pro-social) using a longitudinal sample of 3062 Australian children. Additionally, the analysis explored interactions between observed parenting at 15 months and 2 years (sensitive composite, negative-intrusive composite, positive regard, negative regard, sensitivity, animation, stimulation, detachment, intrusiveness) and parent-reported temperament at 6 months (fear, distress to limitations, falling reactivity, negative affectivity composite, duration of orienting, approach) predicting parent reported behaviour at 3 years (social and emotional competence, internalizing, externalizing, pro-social) using a longitudinal sample of 1093 children from the United States of America. Utilising the same transparent analytical approach from chapter 4, the results found 2 (3%) of 64 interactions were statistically reliable in the Australian sample, whilst there were no statistically reliable interactions (of 1620) in the sample from the United States of America. The two statistically reliable interactions supported different theoretical models and there was no consistency in interaction effects across measures of temperament, behaviour, or parenting. Thus, the findings of chapter 4 and 5 aligned and demonstrate the need for additional research to systematically evaluate measurements and analytic choices to identify consistent and robust evidence of *individual* × *parenting* interactions. In unison with the systematic review, these results demonstrate a substantial gap between theory and evidence regarding *individual* × *parenting* interactions to which possible solutions are outlined in the final chapter.

**Chapter 6** outlines several recommended approaches for future research focused on design, analysis, and results. Approaches for future research include a detailed focus on measurement, extensive validation and exploration of analytical decisions, the use of optimum and more causal research designs to elucidate interaction effects, the use of simulation studies to understand the implications of design, and critically assessing and accurately portraying results.

**Chapter 7** concludes the thesis with an integrative discussion. It highlights that *individual* × *environment* interactions remain a key principle of developmental psychology. Though this thesis demonstrates a gap between theory and evidence, improvement in research design and measurement may align theory with evidence and yield insights that can assist the behavioural development of children.

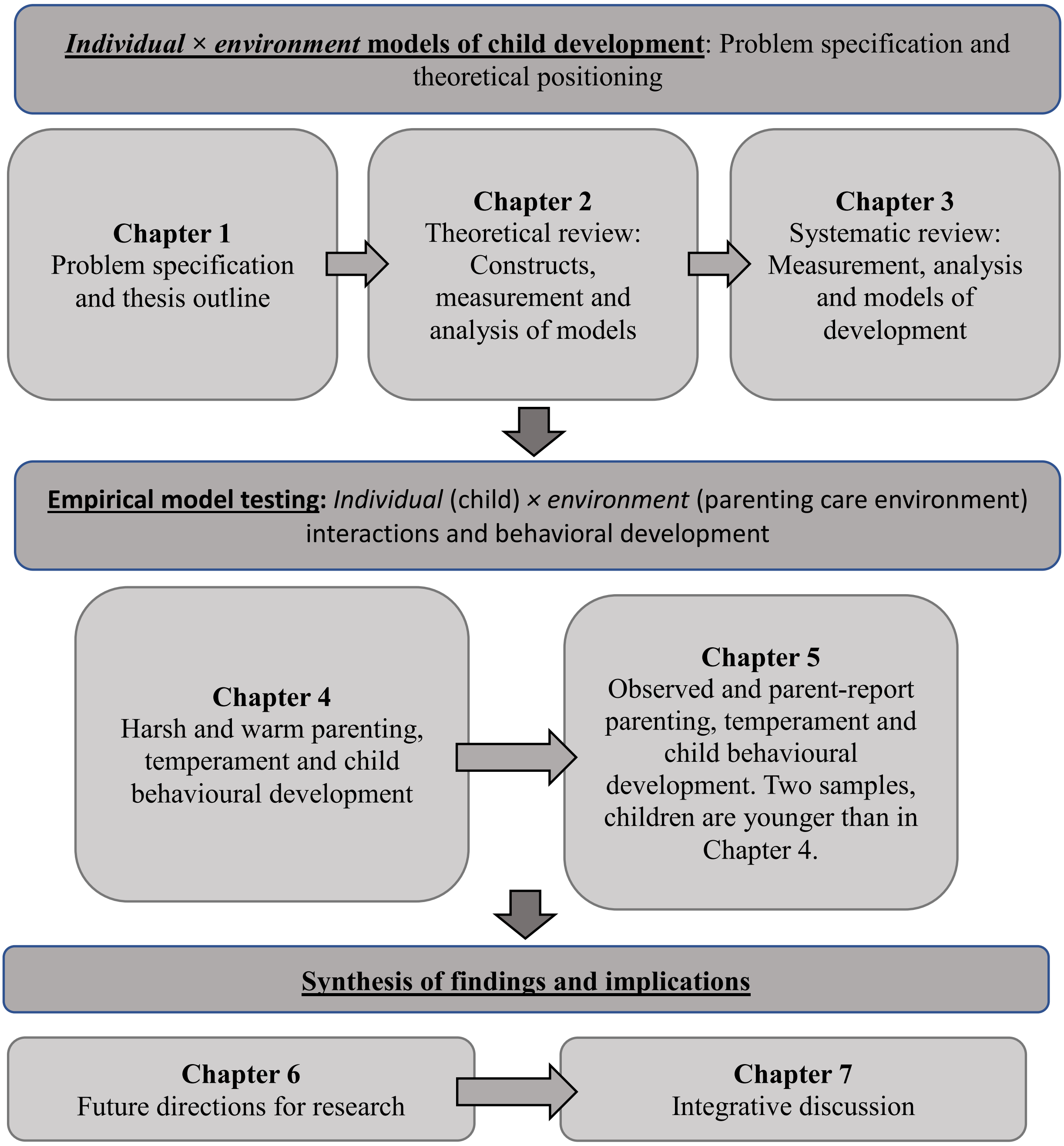



**Figure 1.3:** Outline of thesis structure.

# Chapter 2: A review of individual variation to the effects of parenting on behavioural development

## 2.1 Introduction

Adaption of biological organisms to their environment, both to survive and thrive, is an underpinning tenant of modern developmental science that recognises human development as a complex interplay of varied genetic predispositions with environmental exposures. Reflecting this understanding, research in child development has been moving beyond conceptualisations of uniform environmental or biological influences on child development to the incorporation of interactions between the environment and characteristics of the individual (Bronfenbrenner & Morris 2006; Zubrick et al. 2009; Bell 1968; Patterson 1982; Sameroff 1975; Wang & Liu 2017).

A specific push in this direction has focused on organismic specificity (Wachs & Grandur 1983) or "biological fit" to understand how children vary in their fit to the social environment (Ellis et al. 2011). Research has investigated how individual variation to the environment primes certain children to perform better or worse given a specific social context. A common analogy used to present this conceptualisation is that made to flower species (Boyce & Ellis 2005). Given a range of social contexts, some children are similar to *dandelions*; they perform adequately despite poor conditions. In contrast, some children are like *orchids*; they require specific conditions wherein they thrive, and experience deleterious outcomes in conditions that do not match these conditions.

## 2.2 Theoretical models of child variation in response to the social environment

Social scientists have primarily used three theoretical approaches to explain and explore individual variation in children's developmental response to the social environment. These include the diathesis stress, vantage sensitivity and differential susceptibility hypotheses.

Introduced in chapter 1, the logic of these models is as follows. First, for comparison, the uniform environment or biological model implies that improvements in environmental conditions or biological determinants improve child development outcomes. Diathesis stress, on the other hand, means that when environment conditions are poor, the characteristics of some children cause a negative deviation in their developmental path compared to the path of less-vulnerable peers. Vantage sensitivity mirrors diathesis stress and states that vantage-sensitive children possess characteristics that improve their development in positive environments. Finally, differential susceptibility (similar to biological sensitivity to context and sensory processing sensitivity) states that 'vulnerable' and 'vantage-sensitive' characteristics are susceptibility factors such that children have characteristics making them vulnerable to negative environments and vantage-sensitive to positive environments. In this model, susceptible children have better outcomes in developmentally rich environments and worse outcomes in developmentally poor environments, reflecting greater sensitivity to environment both positive and negative.

### 2.3 Evidence for the three models and the respective components in the models

Many studies have tested these theoretical models by investigating how child characteristics moderate the effects of pertinent environments on child development outcomes. There have been several special editions of high ranked journals (Ellis & Boyce 2011; Belsky & van IJzendoorn 2015) and reviews on individual variation to the environment (Belsky & Pluess 2009; Pluess & Belsky 2013; Boyce 2016; Rabinowitz & Drabick 2017; Bakermans-Kranenburg & van IJzendoorn 2011; van IJzendoorn & Bakermans-Kranenburg 2015). Ongoing, new research reflects the high credence given to the hypothesised models of development and their potential value in translation to intervention strategies.

To date, reflecting the relatively early stage in application of these hypotheses in developmental science, a proliferation of studies have applied a wide variation in the methodological approaches in empirical testing of the theory. Studies have assessed a range of child development outcomes (e.g., academic performance, behaviour, physical health, substance consumption and abuse), assessed a range of developmental environments (e.g., school, parenting, neighbourhood, stressful life events) and employed a range of child characteristics as the index moderator (e.g. temperament, genetic markers and epigenetic variation) (Moore & Depue 2016). There has also been robust discussion within the research literature, thus far, on the analytic decisions and methods to establish criteria for empirical testing of the theories (Belsky et al. 2007; Roisman et al. 2012; Widaman et al. 2012).

Given the large variations in analytical approach, this thesis seeks to contribute to substantive findings and methodological debate. The thesis commences (chapter 2 and chapter 3) with an analysis of the findings thus far and the methods used to arrive at these findings.

The contribution of the thesis focuses on how child characteristics may moderate the impact of care environments (parenting) on child behavioural outcomes (internalising, externalising and social-emotional competence). Parenting and behavioural outcomes are an exemplar of the field because initial work suggesting differential susceptibility regarded parenting (Belsky 1997) and there is a proliferation of subsequent work focussed on the interaction of care environments and child characteristics on behavioural outcomes. The focus on behavioural outcomes also has significant policy links with well-documented paths from behavioural problems to poorer trajectories across the life course (de Graaf et al. 2008; Fergusson et al. 2005; Leschied et al. 2008; Rutter 1989; Sanders 2008).

Several prior reviews of literature focussed on behavioural outcomes provide a strong base to understand the state of play in testing child by environment models (Belsky et al. 2009; Kiff et al. 2011; Rabinowitz & Drabick 2017; Slagt et al. 2016a). The following represents a summary and expansion of this literature.

## 2.4 Evidence of individual variation to the environment: a review of parenting and behavioural development

The aim of this review is to provide a foundational understanding for how child characteristics moderate the association between parenting and child behavioural development. The review introduces key constructs and structural links between these constructs. While the focus is specific to parenting and behavioural development, the mechanisms of sensitivity examined are generalised models of developmental science. The chapter structure introduces child behaviour, child characteristics and behaviour, parenting and child behaviour, how child characteristics moderate the effects of parenting on child behaviour and how methodology guides the understanding of these moderating effects.

## 2.5 Externalizing and internalizing behaviour and social and emotional competence

Social and emotional competence and externalising and internalising behaviours are constructs that categorise how children are responding to their external and internal environment. Specifically, the American Psychological Association (APA) defines behaviour as *an organism's activities in response to external or internal stimuli, including objectively observable activities, introspectively observable activities and nonconscious processes* (https://dictionary.apa.org/behavior 30/05/2018).

Social competence can be defined as behaviours that facilitate positive social interactions which may benefit others, (e.g. cooperation and helping) as well as undertaking behaviours that maintain peer relationships (e.g., sharing and compromise) and build peer rapport (Blair et al. 2015; Rose-Krasnor 1997; Rubin et al. 2012). Emotional competence, on the other hand, involves using identification and regulatory behaviours to manage emotions as well as managing emotions arising through social interactions with others (Camras & Halberstadt 2017; Denham 1998; Saarni 1990). During childhood and adolescence, the ability to form friendships and comply with behavioural expectations are indicators of social and emotional competence (Caldarella & Merrell 1997; Halberstadt et al. 2001; Matson et al. 1983; Rubin et al. 2012).

Externalising and internalising behaviours, on the other hand, are maladjustment of behaviour (Achenbach 1991; Bongers et al. 2004; Campbell 1995; Siddons & Lancaster 2004). Externalising behaviours are actions directed toward other people or property and are readily observable. These include aggression, defiance and antisocial conduct (Frick et al. 1993). Internalising behaviours are emotional and physiological responses and are difficult to observe (Campbell 1995). They include depression, anxiety and rumination (Achenbach 1991).

Externalising and internalising behaviours are occasionally equated with an absence of social and emotional competence, but they are conceptually distinct (Rabinowtiz & Drabick 2017). Regarding the distinction, there are differences in the genetic variation underlying externalising and internalising behaviours compared to that associated with social and emotional competence (Kendler et al. 2011a, 2011b). Moreover, an absence of behavioural problems is not conceptually equivalent to the presence of social competence (Goodman 1997). There are, however, several conceptual linkages. Children with externalising and internalising behavioural problems may have difficulties connecting with peers with attendant effect on development of social and emotional competence (Masten & Obradović 2006). Likewise, children with poor social emotional competence may be more likely to develop externalising and internalising behaviours due to issues with peers and a lack of social support (Masten & Obradović 2006). That is, a bidirectional relationship can occur, where a lack of social emotional competence can drive the development and escalation of externalising and internalising behaviours, while the presence of these behavioural problems may decrease opportunity to develop social competence (Masten & Obradović 2006). Finally, there may be a common cause of both the poor social emotional competence and increased externalising internalising behaviours, such as harsh parenting (Masten & Obradović 2006). Thus, the behaviours are likely associated with each other but are conceptually distinct.

Social and emotional competence and externalising and internalising behaviours are a key focus in developmental science as a substantial body of research identifies such behaviour problems as predictors of poor life course trajectories (Fergusson et al. 2007; Mannuzza et al. 2008; Nock et al. 2007; Odgers et al. 2008; Reef et al. 2011). Externalising and internalising behaviours are associated with lower academic performance, higher school dropout, substance abuse and peer rejection (Bongers et al. 2008; Henricsson & Rydell 2006; Kokko et al. 2006; Vaillancourt et al. 2013). In contrast, higher levels of social-emotional competence are associated with better academic performance and mental health, higher peer acceptance and lower risk-taking and substance abuse (Botvin & Griffin 2014; Domitrvoich et al. 2017; Forster et al. 2015; Herman-Stahl & Petersen 1996; Jones et al. 2015; Segrin et al. 2016). Whilst these associations are understood, prevention and remediation of behaviour problems occurs inconsistently across the population and directs attention to considering how individuals respond differently to different parenting environments and interventions (Belsky & van Ijzendoorn 2015; Domitrvoich et al. 2017; Ungar 2018).

**2.6 Child characteristics and behaviour**

Children's social-emotional competence and behaviours develop through genetic, endophenotypic and phenotypic characteristics interacting with the environment. The extent each characteristic contributes may vary. Child characteristics are conceptualised along a biological continuum with genes at the base (Figure 2.1; Gottesman & Gould 2003; Lenzenweger 2013). In Figure 2.1, phenotype is the observable attributes of a child such as morphological or biochemical features, whilst the endophenotype is biological markers, such as physiological responsivity, that are easier to observe than genetic sequences (Gottesman & Gould 2003; Lenzenweger 2013). In practice, the difference between the three concepts usually refers to scale of measurement. Endophenotype has close ties with genetic variation, whilst phenotype can also refer to larger scale, observable traits, such as temperament (Boyce 2016; Gottesman & Gould 2003; Rabinowitz & Drabick 2017). Throughout this thesis, phenotype will refer primarily to temperament, whilst endophenotype will refer to measures of observed biology that are not genetic sequences. Evidence suggests knowledge of the phenotype, endophenotype and genotype is important for understanding the development of child behaviour (Cicchetti & Dawson 2002). Each is discussed in turn below.


**Figure 2.1:** Biological continuum linking genes, endophenotype and phenotype.

**2.7 Genes and behaviour**

Genes refer to sections of deoxyribonucleic acid (DNA) located in chromosomes (a DNA molecule) of the genome (the entire DNA of an organism) which code for amino acid combinations that form proteins that make structural and functional components of biological organisms and contribute to the endophenotype and phenotype (Lenzenweger 2013; Plomin et al. 2008; Rutter 2006). This coding occurs through two strands of alternating matched nucleotides adenine (A) and thymine (T) and cytosine (C) and guanine (G) that determine, based on the central dogma, the functional expression of the gene (Rutter 2006).

Though human genomes are very similar (~99%), there are variations in genes between individuals known as alleles and polymorphisms which constitute the genotype (The 1000 Genomes Project 2015). Alleles refer to the variations (at least two) in a gene that occur at the same relative location (locus) of the chromosome and polymorphisms are the presence of different DNA sequences in the population (Moore & Depue 2016). For example, one individual may have CC**T**AG, whilst another has CC**A**AG, where the bold text indicates the allelic variant. Single nucleotide polymorphisms (SNPs) are differences occurring in a single nucleotide at a locus. These are the most common polymorphism between people and feature heavily in candidate gene studies (Hattersley & McCarthy 2005; Weeland et al. 2015). There are also variable number of tandem repeats (VNTR) polymorphisms which refers to an allele defined by a sequence being repeated a certain numbers of times. For example, the DRD4 7-repeat allele identifies some individuals repeating the same nucleotide sequence seven times, compared to four repeats in another – plus other various repeat lengths (Plomin & Rutter 1998). Moreover, because humans are diploid and have two copies of each chromosome, individuals can be homozygous (both alleles the same) or heterozygous (one copy of each allele) (Plomin et al. 2008). These allelic variations and polymorphisms constitute the genotype of an individual.

As the genotype of an individual can determine variation in structural and functional components of the endophenotype and phenotype, researchers have examined whether genetic variation can explain social and emotional competence and externalizing and internalizing behaviour, as well as other child development outcomes (McAdams et al. 2014; Plomin & Rutter 1998). Researchers have examined genetic variability associated with the dopamine, opioid and oxytocin, serotonin, gamma-Aminobutyric acid (GABA) corticotrophin-releasing hormone (CRH) and norepinephrine (NE) systems as well as the Brain-Derived Neurotrophic Factor (BDNF) and FK506-binding protein 51 (FKBP5), amongst others (Rabinowitz & Drabick 2017; Rhee et al. 2015; Rutter et al. 1999; Weeland et al. 2015). If genetic variation changes the functional structure of these systems, it could influence the endophenotype and phenotype of an individual and influence their behaviour. Though not comprehensive, the overview below outlines research that has focused on examining how candidate genetic markers and the functional components they code for are associated with the behavioural development of children.

Dopamine is a neurotransmitter that is important for motivation, exploration, goal-directed behaviour, positive emotions and, in some cases, the avoidance of negative outcomes. Four main genes have been examined to link dopamine related genetic variability to child behavioural outcomes. These include the 48-basepair variable number of tandem repeats (VNTR) on exon 3 of the DRD4 gene (ranging from 2 to 11 repeats, with the 7-repeat associated with lower affinity for dopamine), the Taq1 polymorphism of the DRD2 gene (A1 allele of rs1800497, decreased dopamine binding and receptor availability), the dopamine transported DAT1 gene SLC6A3 variant (variable repeat in intron 8 rs3836790, 10 repeat associated with lower dopamine binding thus a longer time period for dopamine be available) and the catechol-O-methyltransferase (COMT) val168met gene (higher COMT increases dopamine metabolism, both val/val and met/met have been associated with differential reactivity). The DRD4, DRD2 and DAT1 variants mentioned above have been associated with externalizing behaviours (Kirley et al. 2004; Mill et al. 2005; Young et al. 2002; Zai et al. 2012; Arcos-Burgos et al. 2012). Whilst, COMT val/val has been linked to internalizing behaviours (Egan et al. 2001; Lee & Prescott 2014; Sheikh et al. 2013), yet the met/met allele has also been associated with and internalizing behaviour (Klein 2016; Woo et al. 2004).

The opiate and oxytocin systems are important for social interaction, social bonding, trust and caring. β-endorphin constitutes the strongest opioid peptide and has a high affinity for the μ-opioid receptor (MOR). Oxytocin is a hormone with receptors located throughout the brain, brain stem and spinal cord (Stoop 2012). Links to the opiate system have been made using the OPRM1 gene (rs1799971) associated with MOR. Although the functional implications remain uncertain, individuals with the G/G alleles are implicated as gaining additional MOR function and show more internalizing behaviours and social withdrawal compared to A/A or A/G variants (Bertoletti et al. 2012; Way et al. 2009), perhaps because social losses are registered heavily for G/G due to enhanced MOR function (Moore & Depue 2016). The oxytocin receptor gene (OTR) and several polymorphisms has been a primary focus for linking oxytocin genetic variation to behaviour. However, the functional implications of those polymorphisms remain uncertain and it is difficult to make strong conclusions regarding their effects on behaviour (Bakermans-Kranenburg & van IJzendoorn 2014). Nonetheless, research has also examined an oxytocin coding gene (OXT) and found single nucleotide polymorphisms, such as the C allele of rs4813625, were associated with more internalizing behaviours in children (Francis et al. 2016) and adult females (Love et al. 2012).

Corticotrophin-Releasing Hormone (CRH) is a hormone that effects the regulation of the hypothalamic–pituitary–adrenocortical (HPA) axis, responses to stress, anxious behaviour, behavioural activation, learning and the formation of memory and emotional memories (Moore & Depue 2016). Though the functionality remains unknown, CRHR1 gene SNPs (e.g., T allele of rs6159, G allele of rs1876828 and rs242939, T allele of rs242941) have been associated with more internalizing behaviours (Smoller et al., 2005; Liu et al., 2006).

Serotonin (5-Hydroxytryptamine; 5-HT) is a vasoconstrictor and neurotransmitter that effects mood, emotion, cognition and circadian rhythms (Heils et al. 1996). Several genes and polymorphisms have been linked to 5-HT and human behaviour. The 5-HT Transporter (5-HTT) gene and the associated 5-HTTPLR polymorphism is commonly examined in studies of individual variation to the environment. The short allele (s/s) of 5-HTTLPR is implicated as lowering serotogenic functioning and increased internalizing behaviours compared to s/l or l/l variants (Caspi et al. 2010; Furmark et al. 2004; Lesch et al. 1996). Likewise, monoamine oxidase A (MAOA) is an enzyme that degrades biogenic amines (e.g., 5-HT and norepinephrine) in the pre-synapse such that higher activation results in lower pre-synaptic concentrations of amines. Regarding the MAOA VNTR gene, 3.5 to 4 repeats show higher activation compared to 3 repeats, whilst activation of 2, 5 and 6 repeats remains ambiguous (Brummett et al. 2007). Some research has found people with low activity variants of MAOA VNTR have more internalizing behaviours (Yu et al. 2005), whilst other authors found high activity variants had more internalizing behaviours (Gutiérrez et al. 2004; Rivera et al. 2009). Another genetic variant associated with serotonin function and behaviour is tryptophan hydroxylase (TPH1 and TPH2). TPH is the rate-limiting enzyme in the biosynthesis of serotonin. Although the functional implications of polymorphisms in TPH1 and TPH2 are uncertain, the T allele of TPH1 rs2108977 and TPH2 rs11178997 were associated with increased post-traumatic stress symptoms (Goenjian et al. 2012). Further, the A allele of TPH2 G1463A was associated with ~80% decrease in 5-HT production in vitro and linked to internalizing behaviour (Zhang et al. 2005).

Norepinephrine (NE) is a stress hormone and neurotransmitter that stimulates adrenergic receptors and neural responses to emotionally relevant cues, including playing a role in memory consolidation of emotional situations (Kravets et al. 2015). The locus coeruleus (LC) is the primary source of NE within the brain, and so the terminology of the LC NE system is often used. Genetic polymorphisms of the α-receptor genes of the NE system are suggested to link with behaviour. The functional role of this gene is uncertain, but a deletion variant in ADRA2B is associated with lower NE concentration and greater emotional memories (de Quervain et al. 2007; Todd et al. 2013). Alpha-receptor gene variants have been associated with externalizing behaviours such as Attention Deficient Hyperactive Disorder (ADHD; Schmitz et al. 2006; Comings et al. 2000; Hawi et al. 2013; Banaschewski et al. 2010) and suggested to be more important for inattention compared to hyperactivity (Schmitz et al. 2006).

Gamma-Aminobutyric acid (GABA) is the primary inhibitory neurotransmitter in the mammalian central nervous system (Le Magueresse & Monyer 2013). Deficits of GABA have been linked to Autism Spectrum Disorders (ASD), internalizing behaviours and externalizing behaviour (Brambilla et al. 2003; Coghlan et al. 2012; Prosser et al. 1997). Also, genetic polymorphisms such as the GABRA2 gene minor haplotypes have been linked to externalizing behaviours (Dick et al. 2006, 2009 Villafuerte et al. 2012, 2013). However, a gene-wide association study did not find links of relevant GABA genes to ASD or ADHD symptoms (Naaijen et al. 2017).

Other potential genetic influences on behaviour relate to the Brain-Derived Neurotrophic Factor (BDNF) and FK506-binding protein 51 (FKBP5). By affecting hippocampal long-term potentiation, BDNF indirectly influences memory and learning (Bueller et al. 2006; Cunha et al. 2010; Panja & Bramham 2014). Met/met homozygous individuals for the val66met polymorphism have lower BDNF secretion, and this has been associated with more internalizing behaviour (Bueller et al. 2006; Chen et al. 2004, 2006; Enoch et al. 2008; Frodl et al. 2007). On the other hand, Alexander et al. (2010) found the val allele was associated with increased internalizing behaviours. The FKBP5 gene codes for FK506-binding protein 51 that regulates glucocorticoid-receptor (GR) sensitivity (Binder 2009) by reducing the GR affinity for cortisol (Wochnik et al. 2005). Appel et al. (2011) and Ising et al. (2008) found FKBP5 gene polymorphisms (e.g., T/T alleles of rs1360780) were associated with excess HPA axis functioning and higher internalizing behaviours.

### 2.8 Endophenotype and behaviour

The endophenotypic characteristics of children can also influence phenotypic expression and behavioural development. Two main systems of interest include the adrenocortical, hypothalamic–pituitary–adrenocortical (HPA) axis and autonomic nervous system (ANS) constituting the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). A range of measures can be used to index these endophenotypic characteristics and some primary measures are discussed below.

The HPA axis is one of the main stress response systems of humans that can influence behaviour (Chrousos et al. 1992; Koss & Gunnar 2018). It largely controls the release of cortisol and subsequent inhibition of that release. Cortisol receptors are found throughout the human body and, as a result, cortisol can have a large effect on developmental processes and behaviour (Bauer et al. 2002). Stress responses, and thus HPA axis activation, in humans are primarily caused by threats of imminent physical harm and threats to social identities such as relationships and self-esteem (Dickerson & Kemeny 2004; Koss & Gunnar 2018). However, cortisol is also produced in non-stress environments and regulates learning, memory and emotion. Additionally, cortisol levels have a diurnal pattern that is a key regulator of sleep (Gunnar & Vazquez 2001). When examining the effects of cortisol it is important to consider the cortisol profile in response to stress and challenges. The benefit profile of cortisol is said to follow an inverted U, with moderate cortisol reactivity in response to moderate challenges associated with improved attention and effortful control (Blair et al. 2008; Wagner et al. 2017). On the other hand, low or high cortisol reactivity in response to moderate challenges or low and high baseline cortisol levels are associated with regulatory problems and increased externalising and internalizing behaviours (Blair et al. 2005; McBurnett et al. 2000; Mills-Koonce et al. 2015; Koss & Gunnar 2018). Thus, through its role in cortisol regulation the HPA axis is a component of child behaviours.

The autonomic system consists of the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The sympathetic nervous system is considered the “fight or flight” system and initiates arousal and physiological responses to stress, such as enhanced blood flow to the lungs and skeletal muscles, sweating and increased heart rate (Furness 2006; Gabella 2001; Kreibig 2010). Acting as the complementary antagonist or “brake” on the SNS, the parasympathetic nervous system is considered the “rest and digest” system and invokes relaxation, enhanced digestion, increased blood flow to internal organs and lowered heart rate (Furness 2006; Gabella 2001; Kreibig 2010).

The SNS can be measured using Skin Conductance Level (SCL) which has been linked to child behaviour (El-Sheikh 2007; El-Sheikh et al. 2010; Lidberg & Wallin 1981). Skin conductance level (SCL) is a measure of electrodermal activity (EDA) that works by tracking how differences in sweat gland activation change the rate that electricity passes across the skin (Kochanska, Brock et al. 2015). As EDA is largely determined by the SNS, SCL is assumed to reflect SNS functioning. When examining links to child behaviour, higher EDA is associated with fear and anxiety whilst lower EDA indicates fearlessness and impulsivity (Block 1957; Fowles et al. 2000). Subsequently, children with low have SCL been associated with more externalising behaviours (Crowell et al. 2006; Lorber 2004; Posthumus et al. 2009; Raine 2002). Thus, measures of SNS are linked to behaviour of children.

Activity of the PNS systems is linked to child behaviour and can be measured with Respiratory Sinus Arrhythmia (RSA; Beauchaine 2001; Van der Graaff et al. 2016). RSA measures heart rate variability and indexes how heart rate changes over the respiratory cycle of inhalation and exhalation. This is a measure of the PNS because heart rate is primarily determined by the vagus nerve which is a component of parasympathetic control (Berntson et al. 1997). It is proposed that differences in RSA during inhalation and exhalation measure the ability to respond to changes in the environment (Porges 1995; Thayer & Lane 2000). Lower RSA is thought to indicate self-regulation difficulties and consequently is associated with a higher number of externalising and internalising behaviours, whilst competent self-regulation and fewer behavioural difficulties are observed for individuals with higher RSA (Porges 1992; Beauchaine 2001). PNS activity is therefore a component of child behaviours.

Another endophenotypic trait linked to behaviour in children is testosterone. Testosterone is an anabolic–androgenic steroid that is a key sex hormone in males and promotes development of reproductive tissues and is associated with increases in muscle and bone mass (Hines et al. 2015; West & Phillips 2010). Higher levels of testosterone have been associated with more externalising, aggressive behaviours, with small effect sizes ranging from 0.08 to 0.14 (Archer et al. 2005; Book & Quinsey 2005; Book et al. 2001).

### 2.9 Temperament and behaviour

Temperament is a measure of typical behavioural patterns of an individual child and reflects underlying genetics (traits) and neural reactivity. Temperament has been variously defined (Buss & Plomin 2014; Zentner & Bates 2008), but generally captures *constitutionally based individual differences in reactivity and self-regulation in the domains of affect, activity and attention* (Rothbart & Bates 2006, pg. 100). Constitutional in this definition refers to the biological underpinning of temperament, in the form of genetic expression, which interacts with a child's environment to determine temperament. Reactivity encapsulates how responsive children are to changes in the internal and external environment. Reactivity can be both specific (e.g., emotion in response to negative events) and general (e.g., fear), and is measured by the length, duration and intensity of a child's motor, affective and attentional reactions to stimulus (Rothbart & Derryberry 1981). Self-regulation, on the other hand, refers to the processes used to control and regulate a child's reactivity.

Zenter & Bates (2008) outline several criteria that capture the theoretical substance of temperament. First, temperament captures behavioural differences between individuals in affect, activity, attention and sensory sensitivity. Second, the behavioural differences are measureable such as the length, duration and intensity of affective and attentional reactions to stimuli. Third, these behaviours appear early in life and are fully expressed, though not necessarily stable, within the first several years. Fourth, primates and some social mammals should exhibit similar behaviours, reflecting biological underpinnings (Jones & Gosling 2005; Réale et al. 2007). Fifth, there are, even if complex, links to biological processes and systems (e.g., genes and endophenotype). And, sixth, the behaviours are enduring on a relative time scale and consistent predictors of outcomes.

Though the debate is open, several main, identified temperament components variously meet these theoretical criteria (Nigg 2006; Zenter & Bates 2008). These components include *behavioural inhibition* or *fear* of the child when experiencing new environments and meeting new people. *Irritability* or *frustration* or *negative emotionality* in response to the child experiencing negative or painful events. *Approach* or *positive emotionality* when the child experiences positive emotions when investigating environments and people or anticipating outcomes. *Activity level* or *energy* of the child measured by the frequency and intensity of movement and activity throughout the day and in response to imposed stillness. *Persistence* or *attention* or *effortful control* when the child is trying to complete tasks, activities or challenges. That is, the self-regulatory capacity of the child to persist voluntarily with difficult tasks. And, *sensory sensitivity* or *threshold* of the child to react and experience, or be negatively overwhelmed by, sensory stimuli (e.g., sound, taste, sight and touch).

These individual components have also been summarised into an *easy vs difficult* overall temperament construct. Children who are behaviourally inhibited, irritable, low in approach, high in activity, low in persistence and high in sensory sensitivity are *difficult*, whilst children with the inverse for these components are *easy* (Carey & McDevitt 1978). Children with other combinations of temperament are *slow-to-warm-up* (low activity and approach and behaviourally inhibited) or *intermediate* types (Carey & McDevitt 1978).

Each of these temperament components can be measured using survey questionnaires, observation and laboratory tasks (Zentner & Bates 2008). The main benefit of laboratory tasks and observation is that researchers can objectively evaluate and compare children's reaction to similar stimuli (Sanson et al. 2004). However, these tasks are time and resource intensive and offer a limited set of experiences to derive a child's temperament profile. On the other hand, surveys of parents are beneficial as they are relatively quick to implement and can capture a child's temperament in response to extensive and hard to observe experiences. Nonetheless, surveys may be biased by parents reporting desirable temperamental traits and by not having an objective comparison to derive how their child would behave (Sanson et al. 2004; Seifer et al. 2004; Zentner & Bates 2008). Surveys of external caregivers can help mitigate the personal biases of parents, but may be limited if external caregivers have not observed the child respond to a wide range of challenging experiences (Achenbach et al. 1987; Sanson et al. 2004; Seifer et al. 2004). Thus, surveys, observations and laboratory tasks can all be used to measure temperament and each method has advantages and disadvantages.

Temperament consistently links to the behavioural development of children. Children who are behaviourally inhibited, irritable, low in approach, high in activity, low in persistence or high in sensory sensitivity or classified as having a difficult temperament have been found at risk for developing more externalizing and internalizing behaviours and lower social and emotional competence (Brock & Curby 2016; Caspi et al. 1995; De Pauw & Mervielde 2010; Graham et al. 1973; Muris & Ollendick 2005; Rubin et al. 2002; Sanson et al. 2004; Thomas 1968; Wiggins et al. 2014).

## 2.10 Parenting and behaviour

Although the behaviour of children is associated with phenotypic, endophenotypic and genetic characteristics of the child, empirical evidence has widely shown that parent behaviours are a potent influence on child development. Components of parenting – including parental behavioural and psychological control, emotional quality of parent-child relationships and parent responsiveness to children's emotions and needs – consistently link to the expressed behaviour of children (Fraley & Roisman 2015; Kiff et al. 2011; Maccoby 2000; Pinquart 2016a, 2017; Roisman & Fraley 2012).

Parental control consists of two main components, behavioural control and psychological control. Behavioural control refers to the strategies parents use to manage the behaviour of their children. Examples of these strategies include monitoring what a child has been doing, punishing and rewarding behaviour, setting rules and limits for acceptable behaviour and how consistently parents apply the control strategies (Baumrind 1971; Kiff et al. 2011; Stattin & Kerr 2000). Children of parents who try to control their behaviour using harsh, hostile or coercive tactics and physical punishment, as well as an inconsistent application of control strategies, tend to develop more externalising and internalising behaviours and lower social and emotional competence (Gershoff 2002; Piko & Balázs 2012; Pinquart 2016a, 2017), though an absence of behavioural control can also lead to the development of behavioural problems (Steinberg et al. 1994; Jewell & Stark 2003).

Psychological control refers to strategies used by parents to constrain, manipulate and invalidate the self-expression and emotional experiences of children (Barber 1996; Barber & Harmon 2002; Bindman et al. 2015; Soenens & Vansteenkiste 2010). Measures of the psychological control construct include low autonomy granting, high intrusiveness, excessive negative control and manipulation and over-control. Children of parents who use the psychological control behaviours in excess tend to have more internalizing behaviours (Möller et al. 2016; Pinquart 2016a, 2017; Pettit et al. 2001; Symeou & Georgiou 2017), and in some cases more externalizing behaviours (Pettit et al. 2001; Symeou & Georgiou 2017). The negative effects of parental psychological control can also be larger in later developmental stages when children require more autonomy to function adequately within social structures (Barber & Harmon 2002; Luyckx et al. 2007; Soenens et al. 2005).

Emotional quality of the parent-child relationship refers to the climate of support and responsiveness parents provide their children. This is typically conceptualised in terms of warmth, closeness and acceptance compared to negativity and rejection (Clark & Ladd 2000; Maccoby 1992; Spera 2005). The warmth component measures affection, positive emotions, involvement and admiration of parents with their children (MacDonald 1992; Davidov & Grusec 2006). Parenting environments high in warmth and acceptance are associated with fewer externalizing and internalizing behaviours and greater social and emotional competence in children (Pinquart 2016a, 2017; Waller et al. 2015; Zhou et al. 2002). Conversely, negativity, rejection and a lack of emotional support from the parents for their children captures the opposite side of parenting emotional quality (Rohner 2004) and is associated with an increase in internalizing and externalizing behaviour and reduced social and emotional competence (Pinquart 2016a, 2017; Rothbaum & Weisz 1994; Waller et al. 2015).

Parent responsiveness to child's emotions and needs refers to how parents respond to child cues, actions, behaviours and affect (Bornstein et al. 2008; Feldman 2007; Kiff et al. 2011). Measures, for example, capture the frequency that a parent provides affirmations, questions and exploratory prompts to a young child playing with toys (Bornstein et al. 1996, 2008). Further, measures include how parents incorporate and respond to the perspectives of children (Ainsworth et al. 1978; De Wolff & van Ijzendoorn 1997). Children of parents who are responsive, sensitive, synchronised and attentive to the child's needs tend to have fewer externalizing and internalizing behaviours and higher social and emotional competence (Belsky & Fearon 2002; Fraley et al. 2013; Gardner 1994; Haltigan et al. 2013; Kochanska et al. 2008; Mesman et al. 2012; Stams et al. 2002).

These parenting measures differ between studies such that there is lower comparability between investigations.

## 2.11 Child characteristics as reactivity mechanisms that moderate the association between parenting and children behaviour

### 2.11.1 Mechanism

Empirical evidence shows that child characteristics including genetic, endophenotypic and phenotypic measures, described above, can moderate the effects of parenting on behavioural development (Boyce 2016). This is not surprising given the extensive data from over 30 years of behavioural genetic research shows that children not only receive an environment but create one through the behavioural (including parenting) responses they elicit and the environments and activities they self-select (Avinun & Knafo 2014; Plomin et al. 2008; Rutter 2006).

The underlying rationale for each moderating mechanism is that child variability in these measured characteristics affects sensitivity to experience, central among these experiences being the parenting environment. The mechanisms are multiple. Assuming the relevance and magnitude of the experience is similar, sensitivity to experience would be present in 1) the differences in individual thresholds at which behavioural responses are neurologically triggered, and 2) the extent that selective attention is focused to heightened processing of emotions, positive and negative incentives and understanding experiences along with forming stronger associations and memories between emotions, incentives and triggering experiences (Moore & Depue 2016). Moore & Depue (2016), present an example of this reactivity framework suggesting that sensitivity of biochemical systems including dopamine (incentives and motivation), oxytocin and opiate (rewards and social cues) and Corticotrophin-Releasing Hormone (CRH; response to uncertainty and risk) all present potential to affect developmental outcomes under conditions of warm and responsive parenting environments (positive) compared to a rejecting and unresponsive parenting environment (negative).

### 2.11.2 Parenting warmth

Measures of parenting warmth and responsiveness, in the context of observational studies where a child plays with toys (e.g., Bornstein et al. 2008), relate to the three biological systems in several ways. First, warm and co-ordinated parenting behaviours that facilitate infant exploration of new environments and toys could activate the incentive system (dopamine). Second, comforting and affectionate embraces in response to observed distress in the infant whilst playing with toys could activate the reward system (oxytocin and opiate). Third, parenting that allows and encourages the child to play freely until distressed, at which point the parent appropriately intervenes and supports the child, could reduce the activation of the stress response system (CRH). Conversely, rejecting and unresponsive parenting may not encourage lengthy exploration or play (dopamine system), distress in the child may cause a withdrawal of social support (oxytocin and opiate) and the parent may inappropriately intrude and inhibit play and be unable to provide warmth and support that could regulate infant distress thus activating the stress response system (CRH).

To summarise potential effects of the warm and responsive parenting environment, the dopamine and incentive motivation system could lean towards productive exploration and play with toys. Likewise, affectionate and responsive support could activate social bonding systems (opiates and oxytocin) that encourage secure attachments. Moreover, because responsive, caring parents could quickly detect distress and intervene with appropriate calming responses the infant may encode stressors as controllable and parents as a dependable stress reducing resource (CRH). As instances of stress are appropriately resolved repeatedly, this kind of parenting could encourage the infant's biological systems towards exploration, self-regulation and solving problems. The result is that the brain has heightened activation of the dopamine and oxytocin and opiate insensitive systems, the brain is primed to engage these systems in context of positive social experiences and there is lower anxiety and stress reactivity from the absence of sustained stress.

### 2.11.3 Harsh parenting

In a harsh and rejecting parenting environment the dopamine system could be activated more to achieve short-term goals before parent interference with child play or due to a lack of support and encouragement when exploring new or complicated toys. Likewise, the oxytocin and opiate system may activate more in the context of avoiding withdrawal of social support if the child became distressed or upset the parent. The lack of parenting responsiveness to needs of the child could mean the stress response system could activate for longer if the child becomes distressed and the child may view parents as unreliable resources for stress-regulation (CRH). Additionally, a rejecting and hostile emotional climate may exacerbate child distress and of itself become a source of stress for the child. In this way, over time the neurological circuitry of the child becomes more sensitive to short-term goals and avoids long-term challenge (dopamine), views social resources and support as scarce (oxytocin and opiates) and reacts more intensely to distress, frequently encodes the environment as dangerous and lowers approach and achievement behaviour from heighted uncertainty and anxiety (CRH).

As can be implicated from the above, and extended to multiple outcome domains e.g., socialisation and schooling, the neurological threshold reactivity of a child regarding the dopamine, oxytocin and opiate and CRH systems would influence how often, and the strength at which, each system is activated given differing levels of warmth and responsive parenting and rejecting and unresponsive parenting. Heightened sensitivity in any, or all, of the systems would exacerbate the effects of the parenting environment for better (warm and responsive) or worse (rejecting and unresponsive). Over time, this interaction between sensitivity to parenting and the association of parenting with the biological systems and behavioural development would be consistently strengthened and lead to divergent trajectories of child development outcomes.

These reactivity thresholds relate to the theoretical models in obvious ways. When a child's neurological circuitry aligns towards heightened stress responses, low social integration and short-term incentive behaviour it may represent a vulnerability for the child such that they develop more externalising and internalising behaviours in harsh environments (diathesis stress). Likewise, a strengthened motivation and reward system, along with downregulated stress responses may allow the child to achieve higher social and emotional outcomes and competence in positive environments where social resources are plentiful (vantage sensitivity). Finally, the differential susceptibility framework would be realised when these children with heightened sensitivities have either the best outcomes when in positive environments or the worse outcomes when in negative environments. Thus, neurological sensitivity could reasonably underlie each theoretical model.

## 2.12 Methodological approaches to assessing *individual* (child characteristics) × *environment* (parenting) models of behavioural development

### 2.12.1 Measurement

Measurement of 'reactivity' or 'sensitivity' to the environment is central to testing *individual* × *environment* models. Researchers have variously measured genetic, endophenotypic and phenotypic variation as characteristics of sensitivity. Genetic measures of sensitivity have focused on single nucleotide polymorphisms (SNPs) and multiple allelic variants (haplotypes) as well as in epigenetic marks that regulate gene expression without altering the genotype (Boyce & Kobor 2015; Boyce 2016; Mitchell et al. 2016). Additionally, recent cases combine several alleles into polygenic scores that index multiple polymorphisms (e.g., Stocker et al. 2017). Endophenotypic measures have focused on heart rate variability (RSA), biochemical and physiological markers such as cortisol, testosterone and skin conductance level (SCL). Measurement of these markers occurs at baseline (resting) and in some cases in response to stress to capture system reactivity. Temperament, phenotypic measures used to index reactivity include surveys and observations of temperament components e.g., negative emotionality, impulsivity, effortful control, inhibition and the difficult temperament composites, as well as surveys measuring sensory processing sensitivity (Wachs & Grandour 1983; Belsky 2005; Belsky & Pluess 2009).

To date, studies of *individual* × *environment* models of behavioural development predominantly use *single* measures of reactivity. With small exceptions that include measures of both genetic variation and temperament (e.g., Davies et al. 2015; Kochanska, Boldt et al. 2015), there has been a lack of multisystem reactivity classification. Studies using genetic, endophenotypic and phenotypic measures of reactivity limit their evaluation to these domains and characteristics (Moore & Depue 2016). Moreover, within these domains studies focus on either generalised or singular constructs of neural reactivity with varying frequency.

The ideal approach to delineating sensitivity, however, would be multivariate and include a mixture of genetic, endophenotypic and phenotypic measures (Bauer et al. 2002). Though stretched along a continuum of measurement type, these measures share a common biological linkage. Genes code for the proteins that create structural and functional components of the human endophenotype that interact with the environment to create the phenotypical temperament of a child (Figure 2.1). Therefore, research may better evaluate these hypotheses by defining and measuring sensitivity in multiple biological systems and with multiple indicators within each system (Cicchetti & Dawson 2002).

The empirical results of chapter 3 provide examples of the measurements used to examine how reactivity moderates the effects of parenting on child behaviour.

### 2.12.2 Statistical method

The primary approach to assessing evidence for child characteristics moderating the effect of the environment on child behavioural outcomes has been to fit a linear regression model to the data and include interaction/moderation terms. In these models, the interaction term is the inclusion of a covariate that represents a characteristic of the environment (e.g., a scale of responsive parenting measured via self-report or observation) multiplied by the moderating characteristic (e.g., parent reported child temperament).

After demonstrating that there is a low probability (e.g., $p<0.05$) that the interaction parameter is equal to zero, the interaction coefficient is explored to interpret implications for child development outcomes.

The historical approach to this exploration was to calculate the simple slopes for the effect of the environment and behaviour at several values of the moderating variable e.g., +/- 1 standard deviation (Aiken et al. 1991; Roisman et al. 2012). After establishing which simple slopes were significant ($p<0.05$), visual comparison of the slopes would then determine the support the proposed theoretical models (Figure 2.2a).

Recent critiques, however, have advocated regions of significance testing (Dearing & Hamilton 2006; Preacher et al. 2006; Roisman et al. 2012). Specifically, for example, simple slope analysis indicates that there is a statistically significant difference in child development outcomes for children with different levels of reactivity (e.g., +/- 1 SD) at some level of parenting. However, it does not indicate the level of parenting where children with different levels of reactivity have a statistically significant difference in child developmental outcomes. The regions of significance (ROS) test overcomes this limitation by testing for an association (a difference) between the child development outcome and reactivity at many values of the parenting environment. By doing this test, a ROS test can successfully indicate when children with different reactivity have a statistically significant difference in child development outcomes given the parenting they experience (Dearing & Hamilton 2006; Preacher et al. 2006; Roisman et al. 2012). Thus, combined with visual interpretation the ROS test can provide a better indication of support for the theories (Figure 2.2b).

Several other criteria have also been proposed to enhance interpretation of the interaction model in context of the theories. First, by adjusting the p-value significance cut-off criteria for multiple comparisons researchers can reduce the chance of mistaking a chance association as a real interaction (Benjamini & Hochberg 1995; Roisman et al. 2012). Specifically, as more interactions are tested it becomes more likely that a study may find spurious interactions. These interactions may only be random error and not 'real' interaction effects. By adjusting for how many interactions are tested researchers can reduce the likelihood of making these 'false-discoveries'. Second, the Proportion of the Interaction (POI) index shows how much of the environmental variables range, within a specified criteria e.g., +/- 2 SD, is above compared to below the point where the simple slope for the moderator at +/- 2 SD the environment crosses (Figure 2.2c). Roisman et al. (2012) suggest 0.4 to 0.6 is required for differential susceptibility, though Del Giudice (2017) suggests 0.2 to 0.8. Third, the Proportion Affected (PA) index is a measure of how many individuals are located above or below the score for the environment that the interaction line crosses over irrespective of environmental range (Figure 2.2c). Roisman et al. (2012) recommend differential susceptibility occurs if PA is greater than 0.16. Fourth, the model should include non-linear environmental or non-linear *environment* × *moderator* effects as they may better represent the data than an interaction with a crossover point (Figure 2.2d). Differential susceptibility is observed when a linear interaction remains significant when non-linear effects are added (Roisman et al. 2012). Finally, the regression model can be re-parameterised so that it estimates the point at which the interaction crosses over and provides a confidence interval for that point that is in effect similar to ROS testing (Widaman et al. 2012).

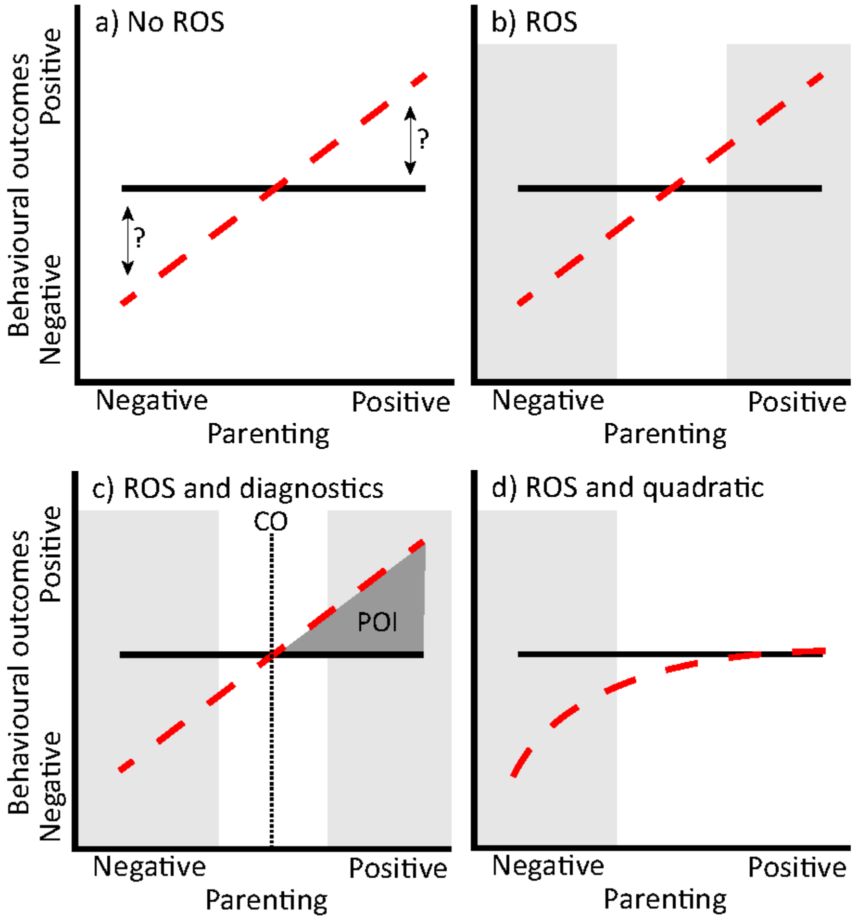



**Figure 2.2:** Updating the exploration of interactions with Regions of Significance (ROS) testing. a) Based on simple slope analysis the dashed line is significant and indicates that children scoring 1 SD higher than the mean reactivity measure have significantly different behaviour to children scoring 1 SD below the reactivity average (solid line) at an unknown level of parenting. For example, the lower or upper question mark. b) The grey polygons indicate areas of parenting where reactive and less-reactive children have significantly different behaviour given parenting. There is higher certainty that this interaction supports differential susceptibility. c) The crossover (CO) point is a reference to a range of criteria. This includes the Proportion of Interaction (POI) metric which is the proportion of the interaction to the right of the CO line (compared to the left) that falls between the dashed and solid line using a specified criteria of parenting e.g., +/- 2 SD. The Proportion Affected (PA) index is the proportion of children who experience parenting above (to the right of) the CO line calculated irrespective of parenting range. d) Including a reactivity by quadratic parenting interaction term, in this case, shows that the initial results of differential susceptibility in b) and c) were actually from not accounting for the quadratic interaction shape.

Although these modifications have improved the exploration of interaction effects, there are some limitations and assumptions involved in their application. First, these methods can be complicated when the form of statistical analysis does not specifically allow these comparisons. For example, Trucco et al. (2016) compare the odds of being included in a latent class associated with high externalizing behaviours compared to the odds of being in a latent class with low or decreasing externalizing behaviours as a function of a genetic polymorphism and parental monitoring. Although this establishes significant difference in class membership, there are no available, to the author's knowledge, equivalents for crossover points and proportional metrics of exposure. Second, these methods assume that the interactions are reliable and not an outcome of chance. Although a researcher may have adjusted for multiple comparisons, there could have been other decisions that influenced the estimated interaction. For example, coding decisions for creating index variables, outcome measures tested and reactivity measures included as moderators. Finally, data issues common to child development research limit these improvements. For example, deeper interaction evaluation does not overcome selection effects or bias from unobserved or non-modelled variables, it does not indicate if the model is appropriate for the data, it does not indicate that variables included in the models represent the construct of interest and are not the result of measurement error and it does not solve issues of observing a limited range of experiences or outcomes (Roisman et al. 2012). The flow chart in Figure 2.3 further summarises the generalised workflow and assumptions of researchers exploring *individual* × *environment* interactions. As can be seen, the exploration of the interaction remains a last and important step. This exploration, however, entirely depends on the preceding data collection, variable choice and modelling strategy being adequate. Without adequate study design and evaluation, deeper exploration of interactions is open for inaccuracies. Thus, overall, deeper exploration of interactions has improved the quality of interaction probing, but study design and evaluation remain very important factors.

| Step | Assumptions |
|---|---|
| Collect data<br>1) Outcome<br>2) Environment<br>3) Moderator | Accurate measurements<br>Representative of *enough* experience<br>Limited selection bias (cross-sectional) |
| Choose variables to include in model | Results independent of variable and processing choices<br>Relevant confounders included |
| Run linear regression with interaction | Model is appropriate |
| Explore interaction | Criteria of Roisman et al. 2012 |

**Figure 2.3:** Flow chart of standard steps in the analysis of *individual* × *environment* interaction effects.

## 2.14 Summary

This chapter provided an overview of theory and evidence for individual variation of children in response to the environment. The chapter summarised child behaviour, the associations between child genetic, endophenotypic and phenotypic characteristics and behaviour, the associations between parenting and child behaviour and how child characteristics may act as mechanisms to moderate the effect of parenting on behaviour. This summary demonstrated that interactive theories of *individual* × *environment* effects on child behavioural development have biological and psychological validity. However, it also demonstrated that improvements in statistical methodology have shaped the evidential concerns of the field. The *individual* × *environment* theories are decades old, yet methods of measurement and evaluative testing are still emerging.

Chapter 3 presents evidence for how often and when child reactivity (indexed using child genotype, endophenotype, and phenotype) is shown to moderate the effects of parenting on child behavioural development consistent with diathesis stress, vantage sensitivity, and differential susceptibility.

## Chapter 3: Evidence for reactivity mechanisms moderating the effects of parenting on children behaviours consistent with the theoretical models

This chapter provides a systematic review of current measurement and methodologies in assessing *individual* × *environment* models. Evidence for reactivity mechanisms that moderate the effect of parenting on child behaviour are the substantive focus. Empirical support for each of the *individual* × *environment* models are systematically examined in the light of methodological strategies of measurement and analysis and an overview of the association between methodological approach findings provided.

### 3.1 Method

This study builds on the systematic review of Rabinowtiz & Drabick (2017) that examined studies reporting child reactivity moderating the effects of parenting on child behavioural development. Using articles identified by Rabinowtiz & Drabick (2017) as the base, the search terms were duplicated to extend to May 1, 2018. The process of literature selection and exclusion is presented in Figure 3.1. This chart highlights all the major decisions used to select the sample of 30 articles that constitute this review.

Specifically, Rabinowtiz & Drabick (2017) searched empirical articles from Psychinfo and Google Scholar. The search criteria included diathesis stress OR dual risk OR differential susceptibility OR vantage sensitivity OR biological sensitivity to context. The date range was from 1997 to November 8, 2016 to coincide with the seminal differential susceptibility publication of Belsky (1997).

Article selection was then refined based on outcomes, parenting and the reactivity moderator. Articles were included if the outcomes indexed social and emotional competence, defined as prosocial behaviours, emotion- or self-regulation, social skills, moral internalization, empathy, and behavioural compliance. Internalizing symptoms such as anxiety and depressive symptoms. Or, externalizing symptoms (e.g., aggression, self-regulation, noncompliance, antisocial or delinquent behaviours). Parenting measures of included articles had to assess parent-child interactions and relationships directly (e.g., questionnaires or observations of parent-child interactions). Articles were excluded if they measured characteristics of the parent (e.g., maternal negative emotionality), broader family processes (e.g., family cohesion), or extreme and non-normative parenting behaviour (e.g., maltreatment); used a composite of parenting and other factors (e.g., parenting and neighbourhood disadvantage); or included behaviours by non-parents (e.g., non-parental childcare). Finally, the reactivity moderator measure had to be genetic, endophenotypic, or phenotypic features of the child.

Other criteria included that the children were between the ages of 0 to 18; the paper was available in English; the study excluded youth and caregivers with psychiatric problems (e.g., attention-deficit/hyperactivity disorder, maternal depression); the article excluded youth from clinic-based samples or youth with potentially restricted generalizability (e.g., youth in foster care); the research did not evaluate the efficacy of interventions; and, the investigation used a moderating interaction between parenting and child reactivity to predict behavioural outcomes during the developmental period of interest (e.g., up to age 18).

Further criteria included that at a minimum, the study tested whether interaction effects were consistent with the diathesis stress and differential susceptibility hypothesis a priori; and, the findings were not from meta-analyses as articles included often overlapped with those reviewed. Finally, the study had to use a test (e.g., regions of significance; ROS) to identify whether the effects of positive and negative parenting on behavioural outcomes were different based on youth characteristics. This could be quantified further with simple slope analysis or by the Roisman et al. (2012) and Widaman et al. (2012) approaches, but was not necessary. Studies that only used simple slopes analyses were excluded.

The current systematic review then added additional empirical articles by searching Psycinfo as Google Scholar did not provide sufficient search capabilities. The search criteria included diathesis stress OR dual risk OR differential susceptibility OR vantage sensitivity OR biological sensitivity to context. The search covered dates from January 1, 2016 to May 1, 2018.

Specific inclusion criteria were the same as above; however, studies were excluded if the parenting measures were child attachment security or the child's relationship with the parent; studies were excluded if interactions were not fully evaluated for one of the theoretical models within the study (e.g., an interaction for which the regions of significance was not evaluated); and, studies were excluded if the regions of significance results for an interaction had a p-value greater than 0.05.

In total, using all criteria, 22 studies were obtained from Rabinowtiz & Drabick (2017) and 8 additional studies were identified. Thus, 30 articles were selected for detailed analysis.

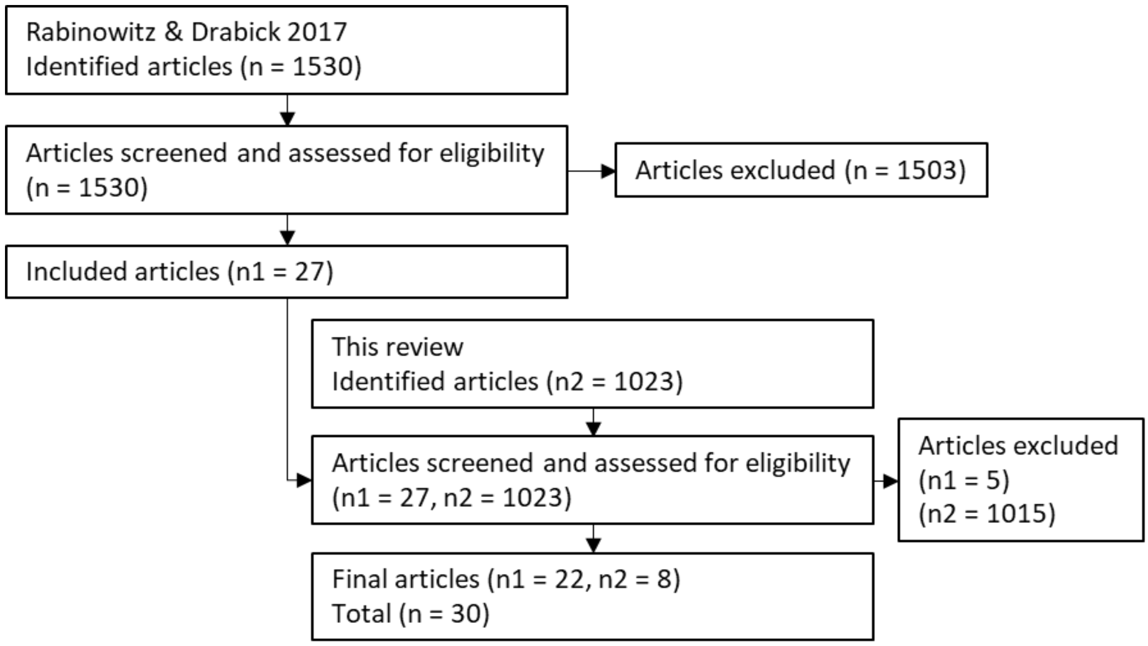



**Figure 3.1:** Article selection process.

For each of the 30 identified articles analysis was undertaken of methodology including the measures of parenting, behaviour, and child characteristic, sample size, gender ratio, and age. The key focus when assessing the statistical procedure and reported results was the theoretical model supported. Preference fell to the reviewer when there was a discrepancy between the original author's interpretation and that of the reviewer. Table 3.1 shows these discrepancies (n=8).

Additionally, the Newcastle-Ottawa scale (Wells et al. 2015), an analytical method for rating quality of reviewed articles, was applied to develop criteria for assessing reported results. Seven criteria were set to grade quality as follows;

1. Independence of measurement: precedence was given to interactions that had independence between measures with the following scores (2=three sources, 1=two sources, 0=one information source). Objective, validated observation measures undertaken by trained researchers were coded as independent sources of information.
2. Adjustment for multiple comparisons: studies adjusting for multiple comparisons were assigned a quality score (1=yes, 0=no).
3. Regions of significance: if an ROS test was performed the interaction scored higher (1=yes, 0=no).
4. Sensitivity analysis: interactions in studies that explored the sensitivity of the result to different specification of their variables scored higher (2=extensive, 1=some, 0=none).
5. Detail of visual summaries: studies scored higher if they plotted raw data or partial residuals (Fox & Weisberg 2017) with, or separate to, summary graphs (1=yes, 0=no).

6. Replicability: the interaction effect scored higher if it was consistent across different ages, outcomes, child characteristics or parenting environments examined in the study (1=yes, 0=no). This was applied quite conservatively, requiring extensive consistency for a 'yes'.
7. Approach to missing data: if studies performed a sensitivity analysis for the method used to account for missing data they scored higher (1=yes, 0=no).

Some studies only assessed one *child* × *parenting* interaction predicting behaviour or no had missing data. These interactions were not scored for replicability or missing data.

**3.2 Results**

**3.3 Description of reviewed studies**

Table 3.1 lists the studies examined in the review. The analysis of studies detailed in Table 3.1 focuses on the interactions claimed as evidence for one of the 3 *individual* × *environment* theories.

In total, 86 interactions were identified by the reviewer to contain adequate evidence. Note, however, three of these interactions provided evidence for a fourth, unique model of one-way contrastive effects (Figure 3.2). One-way contrastive effects occur when the moderated effect of the environment is consistent with diathesis stress for one group e.g., children with a difficult temperament, whilst it is consistent with vantage sensitivity for the other group, e.g., children with an easy temperament. Although these one-way contrastive effects might be reported as separate instances of diathesis stress or vantage sensitivity, they are reported together in this review. Additionally, four interactions from Sulik et al. (2015) came from two, three-way interactions between *gender* × *reactivity* × *parenting*. Specifically, each of these three-way interactions counted as two separate interactions because significant two-way components were estimated.

Some interactions reported had a p-value less than 0.05 yet were non-significant after adjusting for multiple comparisons. These adjusted non-significant interactions were excluded as per the interpretation within each paper.

**3.3.1 Child behavioural outcomes**

Nearly all significant interactions (78) used survey measures of behavioural outcome, whilst only 6 used observations and 2 used surveys and observations. Respondents for the surveys, in findings of significant interactions, were either the parents, teacher or child. Some studies combined ratings from multiple respondents.

Significant interactions were most frequent (n=42) when the dependent variable was externalizing behaviours. Measures of externalizing included proactive and reactive aggression (6), aggression

(1), antisocial behaviours (2), self-regulated compliance (1), cooperation with parental monitoring (1), effortful control (5), callous and unemotional behaviour (1), non-compliance (4), hostility (1), inhibitory control (2), behaviour regulation (1), conduct problems (1) and general externalizing behaviour composites (17).

Significant interactions were also found when pro-social behaviours were the dependent variable (n=19). Measures of prosocial behaviour included social competence (11), internalisation of adult values (1), moral internalisation (1), empathic concern (2), empathetic prosocial behaviour (1) and general prosocial behaviours or skills (3).

When internalising behaviours were the dependent variable, 18 significant interactions were observed. Measures included depression (6), anxiety (2), social anxiety (3) and general internalizing behaviour composites (7).

Some significant interactions also were reported for composites of maladjustment that combined externalising and internalising behaviours into one scale (7).

### 3.3.2 Parenting measures

Over half of the parenting measures in significant interactions derived from observation (48), whilst 36 used survey report measures and 2 used both survey and observation.

Multiple measures of parenting were operationalised in the interactions. Measures of a positive parenting environment included maternal sensitivity (18), maternal warmth (1), maternal affection (1), parental/maternal responsiveness (6), a combination of maternal sensitivity and warmth (9), a combination of maternal warmth and reasoning (5), responsive-supportive parenting (1), encouragement of autonomy (1), maternal/paternal/parental positive parenting (8), maternal/paternal/parental mutually responsive orientation (4), parental/maternal parenting quality (8), parental support (1), proactive control (1) and parental monitoring (1). Measures of the negative parenting environment included harsh discipline/parenting (4), parental/maternal negative parenting (6), negative interactions with parents (2), maternal/paternal psychological control (4), parental/maternal power assertive control (3), maternal harsh intrusion (1) and paternal behavioural control (1).

### 3.3.3 Moderating reactivity characteristics

What is apparent from these summary statistics is that studies have operationalised reactivity using an extensive variety of measures. Most significant interactions used genetic measures (38) or report measures of temperament (32) to operationalise reactivity. Genetics measures and temperament observations were combined in 3 significant interactions, whilst another 3 only used observations of temperament. Endophenotypic measurements were the measure of reactivity in 10 significant interactions.

Measures of genetics employed in studies showing significant interactions used SNP's or haplotypes (37 which includes the 3 interactions that also used observed temperament). The genetic components examined focused on several systems that may indicate reactivity including dopamine (DRD4 exon 3 variable repeat=3; DRD2 rs1800497=2; DAT1 SLC6A3 haplotypes = 3; DAT1 (rs27072 + rs40184 = 1); COMT val158met=6), serotonin (5-HTTLPR=6; 5-HTTLPR + STin2 10-12- repeat = 3; MAOA (rs6323=2, T941G=2)), oxytocin (OXT rs2770378=2, OXT rs4813625=1), BDNF (val66met=2), OPRM (rs1799971=3) and GABA (GABRA-2 rs279827 = 1). Similar genes where used in the 4 interactions that used a polygenic score consisting of 5-HTTLPR rs25531, Taq1A rs1800497 in ANKK1/DRD2, DRD4 and val158met rs4680 in COMT.

Measures of temperament employed in studies showing significant interactions operationalised the construct in several ways. Difficult temperament composites were used most often with a total of 14 significant interactions. Individual temperament components were also used. These individual components included behavioural withdrawal (6 interactions), negative emotionality (4), temperamental flexibility (2), impulsivity (1), effortful control (1), uninhibited temperament (1) and unadaptable temperament (1). Moreover, 5 interactions operationalised reactivity using a measure of sensory processing sensitivity. This assesses sensitivity to the environment with items such as *my child finds it unpleasant to have a lot going on at once* (Pluess et al. 2018; Pluess & Boniwell 2015; Slagt et al. 2018). Though not strictly a measure of temperament, the measure correlated with negative emotionality (r=0.43) and will be considered within the suite of report-measures of temperament. The study combining observations of temperament and genetic measures (3 interactions) used anger proneness as its measure of temperament. Temperament survey measures were primarily completed by parents (26), though 6 were completed by teachers.

Endophenotypic measures of reactivity employed in studies showing significant interactions included Respiratory Sinus Arrhythmia (3; RSA; parasympathetic nervous system), cortisol (2; adrenocortical), testosterone (3) and skin conductance level (2; sympathetic nervous system) and were confined to individual studies.

In sum, studies applied a diversity of measures of reactivity, across and within genetic, self-report of temperament, and endophenotypic assessments.

**Table 3.1**: Evidence for the theoretical models from the reviewed studies. Column test indicates the number of interactions tested within the study using a similar child characteristic, sig refers to the number of significant interactions, B is the broad behavioural construct (PS=pro-social, I=Internalizing, E=Externalizing, IE=internalizing and externalizing), Theo indicates the theoretical model supported (DA=diathesis stress, DS=differential susceptibility, VS=vantage sensitivity, CE1=one-way contrastive effects, CE2=two-way contrastive effects) with claimed theory in brackets, a = independence of measures (0 = one information source, 1 = two sources, 2 = three sources), b = was result adjusted for multiple comparisons (1=yes, 0=no), c = was a regions of significance test performed (1=yes, 0=no), d = was a sensitivity analysis performed for how variables were created (2=extensive, 1=some, 0=none, *=only single interaction tested), e = was data plotted with, or separate to, summary graphs (1=yes, 0=no), f =was the effect consistent across different ages, outcomes, child characteristics or parenting environments in the study (1=yes, 0=no) and g = was sensitivity analysis performed for the method used to account for missing data (1=yes, 0=no, *=complete sample available). Codes for information types include O=observation, S=survey, G=genotype, P=parent, M=mother, F=father, C=child, T=teacher SL=saliva, ECG=electrocardiogram, EL=electrodermal testing and RSA=respiratory sinus arrhythmia. Codes of outcomes include Int=intercept, Ch=change, S=slope, G=grade, K=kindergarten, [L]=longitudinal rounded down for age lower limit and up for age upper limit and [CMP]=complete information.

| | | Child characteristics | | | | Parenting | | Behavioural outcome | | | | | Quality of analysis | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study | Age, n, %male | Test | Sig | Type | Scale | Type | Scale | Type | B | Scale | Theo | P-value | a | b | c | d | e | f | g |
| Belsky et al. (2015) | 2–15[L]; 1695; 52% boys | 52 | 3 | G | OPRM1 rs1799971 A/A (reactive) vs G/G & G/A (non-reactive) | O (M) | Sensitivity | S, P (M) | PS | Social competence (Int) | DS | 0.02 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Belsky et al. (2015) | | 52 | 3 | G | OPRM1 rs1799971 A/A (reactive) vs G/G & G/A (non-reactive) | O (M) | Sensitivity | S, T | PS | Social competence (Int) | DA | 0.02 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Belsky et al. (2015) | | 52 | 3 | G | OPRM1 rs1799971 A/A (reactive) vs G/G & G/A (non-reactive) | O (M) | Sensitivity | S, T | IE | Internalizing & Externalizing (Int) | DA | 0.03 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Cao et al. 2016 (translated) | 12–15[L]; 1328; 51.2% boys | 4 | 2 | G | MAOA rs6323 T/T (non-reactive) vs. G/G (reactive) (girls) | S, P (M) | Encouragement of autonomy | S, C | I | Depression | DS | <0.01 | 2 | 1 | 1 | * | 0 | 0 | 0 |
| Cao et al. 2016 (Trans.) | | 4 | 2 | G | MAOA rs6323 T/T (non-reactive) vs. G/G (reactive) (girls) | S, P (M) | Warmth | S, C | I | Depression | DS | 0.01 | 2 | 1 | 1 | * | 0 | 0 | 0 |
| Chen et al. 2018 | 11.9; 445; 50% boys | 4 | 3 | SL | Testosterone (boys) (high) | S, C&P | Harsh discipline | S, P | E | Pro-active aggression | DA | <0.01 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Chen et al. 2018 | | 4 | 3 | SL | Testosterone (boys) (high) | S, C&P | Harsh discipline | S, P | E | Reactive aggression | VS [DS] | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Chen et al. 2018 | | 4 | 3 | SL | Testosterone (girls) (high) | S, C&P | Harsh discipline | S, P | E | Pro-active aggression | DS | <0.01 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Chen, Yu, et al. (2015) | 13.6; 780; 49% boys | 1 | 1 | G | BDNF val66met val/val or val/met (non-reactive) vs met/met (reactive) | S, C (M) | Warmth and reasoning | S, C | I | Anxiety | DS | <0.05 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| Cho et al. (2016) | 13.46; 309; 46% boys | 1 | 1 | G | DRD4 long (1 allele >= 7 repeat; reactive) vs short (both alleles <= 6 repeat; non-reactive) | S, P | Responsive-supportive parenting | S, T | E | Behaviour regulation | DS | <0.05 | 2 | 0 | 1 | 0 | 0 | 1 | 0 |

Table 3.1 continued

| | | | | Child characteristics | | Parenting | | Behavioural outcome | | | | | Quality of analysis | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study | Age, n, %male | Test | Sig | Type | Scale | Type | Scale | Type | B | Scale | Theo | P-value | a | b | c | d | e | f | g |
| Davies et al. (2015) | 2.2; 201; 56% boys | 4 | 1 | G | DAT1 rs27072 and rs40184 C/C alleles (reactive) vs C/T and T/T (non-reactive) (composite 0-2 via C/C = 1, C/T or T/T = 0) | S&O, R&P (M) | Responsiveness | S&O, R | E | Anti-social behaviour (Ch) | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Davies et al. (2015) | | 2 | 1 | O | Uninhibited temperament (high) | S&O, R&P (M) | Responsiveness | S&O, R | E | Anti-social behaviour (Ch) | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Gallitto (2015) | 6.50; 2631; 49% boys | 4 | 2 | S, P | Unadaptable temperament (high) | S, P | Positive parenting | S, P | I | Internalizing | DA | <0.05 | 0 | 0 | 1 | * | 0 | 0 | 0 |
| Gallitto (2015) | | 4 | 2 | S, P | Difficult temperament (high) | S, P | Positive parenting | S, P | E | Externalizing | DA | <0.001 | 0 | 0 | 1 | * | 0 | 0 | 0 |
| Janssens et al. (2016) | 13.76; 985; 51% boys | 240 | 1 | G | DRD4 long (1 allele >= 7 repeat; reactive) vs short (both alleles <= 6 repeat; non-reactive) | S, P | Proactive control | S, P | E | Aggression | DS | <0.001 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Kim & Kochanska (2012) | 2.1; 102; 50% boys | 4 | 2 | O | Negative emotionality (high) | O (M) | Mutually responsive orientation | O | E | Self-regulated compliance | DS | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | |
| Kim & Kochanska (2012) | | 4 | 2 | O | Negative emotionality (high) | O (M) | Mutually responsive orientation | O | E | Effortful control | DS | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| King et al. (2016) | 1.5; 63; 48% boys | 1 | 1 | G | DRD4 exon 3 long (1 allele >= 7 repeat; reactive) vs short (both alleles <= 6 repeat; non-reactive) | O (M) | Sensitivity | S, P (M) | E | Externalizing | DA | 0.023 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| Kochanska et al. (2011) | 5.6; 101; 50% boys | 2 | 2 | G | 5-HTTLPR s/s or s/l alleles (reactive) vs l/l (non-reactive) | O (M) | Responsiveness | O | PS | Moral internalization | DS | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| Kochanska et al. (2011) | | 2 | 2 | G | 5-HTTLPR s/s or s/l alleles (reactive) vs l/l (non-reactive) | O (M) | Responsiveness | S, P (M&F) | PS | Social competence | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| Kochanska, Boldt, et al. (2015) | 10; 102; 50% boys | 6 | 3 | G&O | Biobehavioural risk (5-HTTLPR s/s or s/l and high proneness to anger (reactive) vs l/l and low anger proneness (non-reactive) | O (M&F) | Mutually responsive orientation | S, P (M&F) | E | Cooperation with parental monitoring | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Kochanska, Boldt, et al. (2015) | | 6 | 3 | G&O | Biobehavioural risk (5-HTTLPR s/s or s/l and high proneness to anger (reactive) vs l/l and low anger proneness (non-reactive) | O (M&F) | Power-assertive parenting | S, C | PS | Internalization of adult values | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Kochanska, Boldt, et al. (2015) | | 6 | 3 | G&O | Biobehavioural risk (5-HTTLPR s/s or s/l and high proneness to anger (reactive) vs l/l and low anger proneness (non-reactive) | O (M&F) | Power-assertive parenting | S, P (M&F) | E | Callous–Unemotional behaviour | DA | <0.025 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Kochanska, Brock et al. (2015) | 8–10$^{L}$; 102; 50% boys | 8 | 2 | EL | Skin conductance level (low) | O (M) | Power-assertive control | S, P (M) | E | Externalizing | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Kochanska, Brock et al. (2015) | | 8 | 2 | EL | Skin conductance level (low) | O (F) | Mutually responsive orientation | S, P (F) | E | Externalizing | DS | <0.025 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Li, Sulik et al. (2016) | 2.5–4.5$^{L}$; 145; 54% boys | 3 | 3 | G | SLC6A3 haplotype 1: Intron8 A/A or A/G and Intron 13 A/G or G/G (reactive) vs Intron8 G/G and Intron 13 A/A (non-reactive) | O (M) | Parenting quality | O | E | Effortful control | DA | <0.05 | 2 | 0 | 1 | 1 | 0 | 1 | 0 |
| Li, Sulik et al. (2016) | | 3 | 3 | G | SLC6A3 haplotype 2: Intron8 A/A or A/G and 3'UTR VNTR- 10/09 or 10/10 (non-reactive) vs Intron8 G/G and 3'UTR VNTR 09/09 (reactive) | O (M) | Parenting quality | O | E | Effortful control | DA | <0.05 | 2 | 0 | 1 | 1 | 0 | 1 | 0 |

## Table 3.1 continued

| | | | | Child characteristics | | Parenting | | Behavioural outcome | | | | | Quality of analysis | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study | Age, n, %male | Test | Sig | Type | Scale | Type | Scale | Type | B | Scale | Theo | P-value | a | b | c | d | e | f | g |
| Li, Sulik et al. (2016) | 2.5–4.5[L]; 145; 54% boys | 3 | 3 | G | SLC6A3 haplotype 3: Intron 13 G/G and 3'UTR VNTR- 10/10 (non-reactive) vs Intron 13 A/A or A/G and 3'UTR VNTR-10/09 or 09/09 (reactive) | O (M) | Parenting quality | O | E | Effortful control | DA | <0.05 | 2 | 0 | 1 | 1 | 0 | 1 | 0 |
| Olofsdoter et al. (2018) | 17.3; 1359; 40.6% boys | 4 | 3 | G | OXT rs2770378 G/G=0 (non-reactive) vs G/A=1,A/A=2 (reactive) | S, C | Negative parenting | S, C | I | Social anxiety | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Olofsdoter et al. (2018) | | 4 | 3 | G | OXT rs4813625 G/G=0 (non-reactive) vs G/C=1, C/C=2 (reactive) | S, C | Positive parenting | S, C | I | Social anxiety | DS | <0.001 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Olofsdoter et al. (2018) | | 4 | 3 | G | OXT rs2770378 G/G=0 (non-reactive) vs G/A=1,A/A=2 (reactive) | S, C | Positive parenting | S, C | I | Social anxiety | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Rabinowitz et al. (2016) | 12.99; 775; 71% boys | 8 | 2 | S, P (M) | Temperamental flexibility (low) | S, C (F) | Positive parenting | S, P (F) | E | Externalizing | DA | <0.005 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Rabinowitz et al. (2016) | | 8 | 2 | S, P (M) | Temperamental flexibility (low) | S, C (F) | Positive parenting | S, P (F) | I | Internalizing | DA | <0.005 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | 5–11[L]; 1306; 52% boys | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | IE | Internalizing & Externalizing (Int) | DS | 0.005 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | IE | Internalizing & Externalizing (G1) | DS | 0.012 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | IE | Internalizing & Externalizing (G4) | DA | 0.004 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | IE | Internalizing & Externalizing (G5) | DS | 0.002 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, P (M) | PS | Social competence (Int) | DA | 0.019 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, P (M) | PS | Social competence (K) | DA | 0.008 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, P (M) | PS | Social competence (G1) | DA | 0.013 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, P (M) | PS | Social competence (G3) | DA | 0.017 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | PS | Social competence (Int) | DS | <0.001 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | PS | Social competence (G1) | DS | 0.003 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | PS | Social competence (G2) | DS | 0.002 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | PS | Social competence (G4) | DA | 0.005 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Roisman et al. (2012) | | 35 | 13 | S, P (M) | Difficult temperament (high) | O (M) | Sensitivity | S, T | PS | Social competence (G5) | DS | 0.003 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. (2016b) | 10; 120; 46% boys | 12 | 3 | S, P (M&F) | Negative emotionality (high) | O (M&F) | Harsh parenting | S, P (M&F) | PS | Prosocial behaviour | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. (2016b) | | 12 | 3 | S, P (M&F) | Impulsivity (high) | O (M&F) | Parental responsiveness | S, P (M&F) | E | Externalizing | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. (2016b) | | 12 | 3 | S, P (M&F) | Effortful control (low) | O (M&F) | Parental responsiveness | S, P (M&F) | E | Externalizing | VS | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. 2018 | 4.78; 264; 52.9% boys | 32 | 5 | S, P (M) | Sensory processing sensitivity (high) | S, P (M) | Negative parenting (I) | S, T | E | Externalizing (S) | VS | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |

## Table 3.1 continued

| | | | | Child characteristics | | Parenting | | Behavioural outcome | | | | | Quality of analysis | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study | Age, n, %male | Test | Sig | Type | Scale | Type | Scale | Type | B | Scale | Theo | P-value | a | b | c | d | e | f | g |
| Slagt et al. (2018) | 4.78; 264; 52.9% boys | 32 | 5 | S, P (M) | Sensory processing sensitivity (high) | S, P (M) | Negative parenting (S) | S, T | E | Externalizing (Int) | VS | <0.01 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. (2018) | | 32 | 5 | S, P (M) | Sensory processing sensitivity (high) | S, P (M) | Negative parenting (S) | S, T | E | Externalizing (S) | CE2 [DS] | <0.01 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. (2018) | | 32 | 5 | S, P (M) | Sensory processing sensitivity (high) | S, P (M) | Positive parenting (S) | S, T | E | Externalizing (Int) | VS | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Slagt et al. (2018) | | 32 | 5 | S, P (M) | Sensory processing sensitivity (high) | S, P (M) | Positive parenting (S) | S, T | E | Externalizing (S) | CE2 [DS] | <0.01 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Stocker et al. (2017) | 12–14[L]; 323; 45% boys | 4 | 4 | GP | Polygenic sensitivity | O (M&F) | Parenting quality | S, C | I | Depression (Ch) | DS | 0.03 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Stocker et al. (2017) | | 4 | 4 | GP | Polygenic sensitivity | O (M&F) | Parenting quality | S, C | I | Anxiety (Ch) | DS | 0.02 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Stocker et al. (2017) | | 4 | 4 | GP | Polygenic sensitivity | O (M&F) | Parenting quality | S, C | E | Hostility (Ch) | CE2 [DS] | 0.001 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Stocker et al. (2017) | | 4 | 4 | GP | Polygenic sensitivity | O (M&F) | Parenting quality | S, C | IE | Internalizing & Externalizing (Ch) | CE2 [DS] | 0.0015 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Stoltz et al. (2017) | 5–12[L]; 129; 48–51%[CMP] boys | 2 | 1 | S, P (M) | Negative affectivity (high) | O | Parenting quality | S, T | E | Externalizing | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 1 |
| Sulik et al. (2012) | 4.5; 138; 52% boys | 28 | 4 | G | 5-HTTLPR SLC6A4 l/l (reactive) vs s/s and s/l (non-reactive) | O (M) | Sensitivity-warmth | S, P | E | Non-compliance (Int) | DS | <0.01 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Sulik et al. (2012) | | 28 | 4 | G | 5-HTTLPR l/l alleles and STin2 10- and 12-repeat alleles halpotype (combination of halpotype groups) L10-L12 (reactive) vs S10 (non-reactive) vs S12 (non-reactive) | O (M) | Sensitivity-warmth | S, P | E | Non-compliance (Int) | DS | L10-L12 vs S10 = <0.01 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Sulik et al. (2012) | | 28 | 4 | G | 5-HTTLPR l/l alleles and STin2 10- and 12-repeat alleles halpotype (combination of halpotype groups) L10-L12 (reactive) vs S10 (non-reactive) vs S12 (non-reactive) | O (M) | Sensitivity-warmth | S, P | E | Non-compliance (Int) | DS | L10-L12 vs S12 = <0.05 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Sulik et al. (2012) | | 28 | 4 | G | 5-HTTLPR s/s alleles and STin2 10-repeat allele halpotype (combination of halpotype groups) S10 (reactive) vs S12 (non-reactive) vs L10-L12 (non-reactive) | O (M) | Sensitivity-warmth | S, P | E | Non-compliance (S) | DS | S10 vs L10-L12 <0.05 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| Sulik et al. (2015) | 4.5; 145; 54% | 8 | 5 | G | COMT val158met met/met (reactive) vs met/val and val/val (non-reactive) (boys) | O (M) | Sensitivity-warmth | S, P | I | Internalizing | DA (boys) | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Sulik et al. (2015) | | 8 | 5 | G | COMT val158met val/met or val/val (reactive) vs met/met (non-reactive) (girls) | O (M) | Sensitivity-warmth | S, P | I | Internalizing | DA (girls) | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Sulik et al. (2015) | | 8 | 5 | G | COMT val158met met/met (reactive) vs met/val & val/val (non-reactive) (boys) | O (M) | Sensitivity-warmth | S, P (M) | E | Inhibitory control | VS (boys) | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| Sulik et al. (2015) | | 8 | 5 | G | COMT val158met val/met or val/val (reactive) vs met/met (non-reactive) (girls) | O (M) | Sensitivity-warmth | S, P (M) | E | Inhibitory control | VS (girls) | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| Sulik et al. (2015) | | 8 | 5 | G | COMT val158met val/met or val/val (reactive) vs met/met (non-reactive) (girls) | O (M) | Sensitivity-warmth | S, P (M) | E | Effortful control | VS (girls) | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | * |
| Trucco et al. (2015) | 12–17[L]; 504; 71% boys | 1 | 1 | G | GABRA-2 rs279827 G/G (reactive) vs A/A and A/G (non-reactive) | S, C | Parental monitoring | S, C | E | Externalizing | DS | <0.05 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| Van der Graaff et al. (2016) | 17.04; 379; 56% boys | 16 | 3 | ECG | RSA baseline (lower) (girls) | S, C (M&F) | Parental negative interactions | S, C | E | Externalizing | DS | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |

Table 3.1 continued

| Study | Age, n, %male | Test | Sig | Type | Scale | Type | Scale | Type | B | Scale | Theo | P-value | a | b | c | d | e | f | g |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Child characteristics | | | | | Parenting | | Behavioural outcome | | | | | Quality of analysis | | | | | | |
| Van der Graaff et al. (2016) | 17.04; 379; 56% boys | 16 | 3 | ECG | RSA baseline (higher and lower) (boys) | S, C (M&F) | Parental negative interactions | S, C | PS | Empathic concern | CE1 [DA (high) VS (low)] | <0.01 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Van der Graaff et al. (2016) | | 16 | 3 | ECG | RSA baseline (higher) (girls) | S, C (M&F) | Parental support | S, C | PS | Empathic concern | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Wagner et al. (2017) | 0.5–8[L]; 1234; 43% boys | 18 | 2 | SL | Baseline cortisol (higher) | O (M) | Harsh intrusion | S, P (M) | PS | Empathetic-prosocial behaviour | DA | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Wagner et al. (2017) | 1 | 18 | 2 | SL | Cortisol reactivity (higher) | O (M) | Sensitivity | S, P (M) | E | Conduct problems | VS | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zarra-Nezhad et al. (2014) | 6; 314; 52% boys | 18 | 6 | S, T | Behavioural withdrawal (high) | S, P (M) | Affection | S, T | E | Externalizing | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zarra-Nezhad et al. (2014) | | 18 | 6 | S, T | Behavioural withdrawal (high) | S, P (M) | Psychological control | S, T | I | Internalizing | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zarra-Nezhad et al. (2014) | | 18 | 6 | S, T | Behavioural withdrawal (high) | S, P (M) | Psychological control | S, T | E | Externalizing | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zarra-Nezhad et al. (2014) | | 18 | 6 | S, T | Behavioural withdrawal (high) | S, P (M) | Psychological control | S, T | PS | Pro-social skills | DA | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zarra-Nezhad et al. (2014) | | 18 | 6 | S, T | Behavioural withdrawal (high and low) | S, P (F) | Behavioural control | S, T | I | Internalizing | CE1 [DA (high) VS (low)] | <0.05 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zarra-Nezhad et al. (2014) | | 18 | 6 | S, T | Behavioural withdrawal (high and low) | S, P (F) | Psychological control | S, T | I | Internalizing | CE1 [DA (high) VS (low)] | <0.001 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zhang et al. (2015) | 11–12[L]; 1026; 50% boys | 6 | 2 | G | DRD2 rs1800497 A1/A1 and A1/A2 (reactive) vs A2/A2 (non-reactive) | S, P (M) | Negative parenting | S, C | I | Depression | DS | 0.034 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Zhang et al. (2015) | | 6 | 2 | G | DRD2 rs1800497 A1/A1 and A1/A2 (reactive) vs A2/A2 (non-reactive) | S, P (M) | Negative parenting | S, C | I | Depression | DS | 0.049 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Zhang, Cao et al. 2016 | 12.32; 1399; 53% boys | 12 | 3 | G | MAOA T941G T (reactive) vs G (non-reactive) allele (boys) | S, P (M) | Warmth and reasoning | S, T | E | Reactive aggression | DS | <0.05 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zhang, Cao et al. 2016 | | 12 | 3 | G | MAOA T941G T/T (reactive) vs G/G (non-reactive) (girls) (T/G excluded) | S, P (M) | Warmth and reasoning | S, T | E | Reactive aggression | DS | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zhang, Cao et al. 2016 | | 12 | 3 | G | COMT val158met val/val (non-reactive) vs val/met and met/met (reactive) | S, P (M) | Warmth and reasoning | S, T | E | Reactive aggression | DS | <0.01 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Zhang, Li et al. 2016 | 13.60; 780; 49% boys | 2 | 1 | G | BDNF val66met: val/val or val/met (reactive) vs met/met (non-reactive) | S, C (M) | Warmth and reasoning | S, C | I | Depression | DS | 0.03 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |

### 3.4 Evidence for the each *individual* × *environment* theory

In total 542 interactions were evaluated, however, only 86 evidenced significant effects. Thus, 456 interactions did not provide evidence to support one of the *individual* × *environment* theories. Two studies, however, contribute substantially to non-findings. Janssens et al. (2016) estimated 240 interactions and found only one result, while Belsky et al. (2015) performed a deliberate exploratory study involving 52 interactions assessing parenting, child behaviour, and candidate genes and found three interactions. Excluding these studies reduces the number of significant interactions to 82 and the number of non-significant interactions to a more reasonable 168 (456-49-239). Regardless, finding insufficient evidence to support one of the theoretical models appears to be the norm.

Of the 86 significant interactions, 37 (45%) supported diathesis stress, 33 (37%) supported differential susceptibility, and 9 (13%) supported vantage sensitivity. The generalised findings demonstrate a leniency towards diathesis stress and differential susceptibility hypotheses. Additionally, 3 (3%) interactions evidenced one-way contrastive effects and 4 (5%) an alternative hypothesis of two-way contrastive effects. Figure 3.2 illustrates all models for clarity

In the one-way contrastive model, the effect of the environment is conditionally reversed, yet reactive children fair for worse. Specifically, the moderated effect of the environment is consistent with diathesis stress in *negative environments* for the reactive group, whilst it is consistent with vantage sensitivity for the non-reactive group in the *negative environments* which become conditionally *positive*. On the other hand, in the two-way contrastive model the effect of the environment is entirely reversed depending on the child's level of reactivity. What is positive for the so-labelled *reactive* child is negative for the *non-reactive* child, whilst what is negative for the *reactive* child is positive for the *non-reactive* child (Belsky et al. 2007).

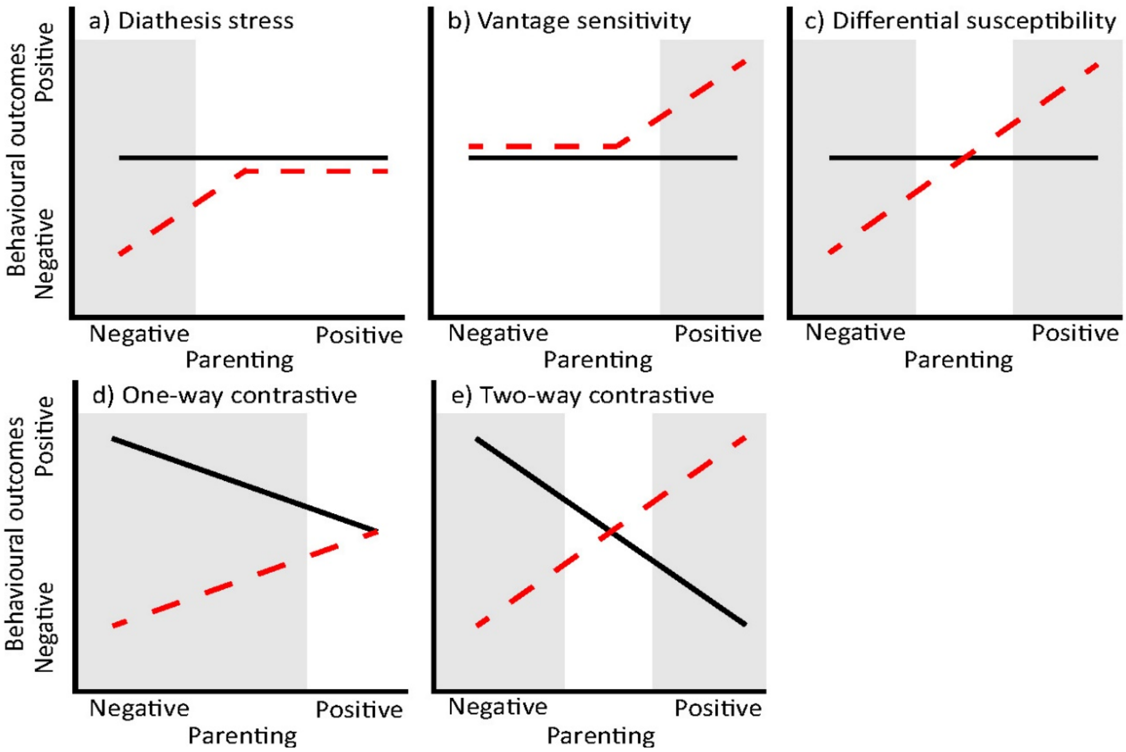



**Figure 3.2:** Theoretical models for which empirical evidence was found; a) diathesis stress, b) vantage sensitivity, c) differential susceptibility, d) one-way contrastive effects, and e) two-way contrastive effects. Reactive children are the dashed, red line. Regions within the grey rectangle are where child reactivity has a statistically significant effect on behaviour given the parenting value.

### 3.4.1 Evidence for diathesis stress

Several diathesis stress effects using either temperament, endophenotypic or genetic characteristics as the moderator and examples are provided here. Roisman et al. (2012) found children reported as difficult temperaments consistently had lower parent rated social competence when maternal sensitivity was low compared to non-difficult counterparts. When maternal sensitivity was high, children with difficult and easy temperaments were equivalent regarding social competence. Likewise, Gallitto (2015) found children with difficult temperaments had more externalising behaviours when there was a lack of positive parenting compared to children with easy temperaments, and children of easy or difficult temperament given the parenting environment was positive. Further, Zarra-Nezhad et al. (2014) demonstrated children high in behavioural withdrawal had more internalising behaviours when paternal and maternal psychological control was high compared to children low on behavioural withdrawal or children exposed to low levels of paternal and maternal psychological control.

Turning to the endophenotype, diathesis stress effects were observed where boys high in testosterone exposed to harsh parental discipline had higher proactive aggression (Chen et al. 2008), girls with a higher baseline RSA exposed to low parental support had reduced empathic concern (Van der Graff et al. 2016) and children with a high baseline cortisol exposed to harsh maternal intrusion had fewer empathic-prosocial behaviours (Wagner et al. 2017). For each of these results, children not exposed to the parenting or who had opposite levels of the moderating characteristic performed equivalently better.

Genotypic examples of diathesis stress include children with the DRD4 seven repeat allele having more externalising behaviours in context of low maternal sensitivity (King et al. 2016), children with the 5-HTTLPR s/s or s/l allele having less social competence in context of low maternal responsiveness (Kochanska et al. 2011) and children with the OXT A/A allele having more social anxiety in the context of negative parenting or an absence of positive parenting (Olofsdoter et al. 2018). As before, children not exposed to negative parenting or with different alleles for the moderating genotype performed equivalently better.

There are also examples of diathesis stress using a measure of genotypic and temperamental sensitivity. Specifically, children with the 5-HTTLPR s/s or s/l allele who were also prone to anger had lower internalisation of adult values and more callous-unemotional behaviour in context of more power-assertive parenting, as well as less cooperation with parental monitoring when parental mutually responsive orientation was low. This is in comparison to children who did not experience such parenting or had the l/l allele.

### 3.4.2 Evidence for differential susceptibility

Evidence for differential susceptibility was found using temperament, endophenotypic and genotypic markers of sensitivity, some examples are presented here. In terms of temperament, Roisman et al. (2012) found children high in difficult temperament consistently had fewer teacher rated internalising and externalising behaviours and more teacher rated social competence in higher maternal sensitivity environments compared to children with easy temperaments, however, the children with difficult temperaments had more internalising and externalising behaviours and less social competence when maternal sensitivity was low. Likewise, Kim and Kochanska (2012) found children observed to have high negative emotionality had either the lowest self-regulated compliance and effortful control when maternal mutually responsive orientation was low, compared to low emotionality children, or the highest self-regulated compliance and effortful control when maternal mutually responsive orientation was high.

In terms of the endophenotype and differential susceptibility, children with a low skin conductance level had either the lowest externalising behaviours if paternal mutually responsive orientation was high or the most externalising behaviours if paternal usually responsive orientation was low

(Kochanska, Brock et al. 2015). Children with a higher skin conductance level were less affected by paternal mutually responsive orientation. In similar terms, girls with a higher level of testosterone had the lowest level of proactive aggression when harsh parental discipline was minimal but the highest level of proactive aggression when harsh parental discipline was high, girls with a low level of testosterone were less affected by harsh parental discipline (Chen et al. 2018). Girls with a low baseline RSA had the most externalising behaviours if parental negative interactions were high but the least number of externalising behaviours if parental negative interactions were minimal, girls with a high baseline RSA were less affected by parental negative interactions (Van der Graaff et al. 2016).

Genotypic evidence for differential susceptibility was also found. Belsky et al. (2015) found children with the OPRM1 rs1799971 A/A allele had the highest social competence when early maternal sensitivity was high but the lowest social competence when early maternal sensitivity was low. Children with G/A or G/G allele were less affected by maternal sensitivity (Belsky et al. 2015). Trucco et al. (2015) found that children with the GABRA-2 rs279827 G/G allele had the most externalising behaviours when parental monitoring was minimal but the least externalising behaviours when parental monitoring was high, children with the G/A or G/G allele were less affected by parental monitoring. Olofsdoter et al. (2018) observed that children with the OXT rs4813625 C/C allele had the least social anxiety if positive parenting was high but the most social anxiety if positive parenting was low, children with the G/C or G/G allele were less affected by positive parenting.

Interestingly, the polygenic sensitivity score implemented by Stocker et al. (2017) also found evidence for differential susceptibility. Specifically, children with high polygenic sensitivity had the least depression and anxiety when parenting quality was high but the most depression and anxiety when parenting quality was low. The effect of parenting quality on depression and anxiety declined in line with a reduction in the polygenic sensitivity score and confirmed differential susceptibility.

### 3.4.3 Evidence for vantage sensitivity

Evidence for the vantage sensitivity model was found when sensitivity measures were temperament, endophenotype and genotype. Children with low behavioural withdrawal, compared to children with high behavioural withdrawal, had fewer internalising behaviours when paternal behavioural and psychological control was low (Zarra-Nezhad et al. 2014). When behavioural and psychological control was high these children had internalising behaviours equivalent to children with high behavioural withdrawal. Likewise, children with low effortful control had fewer externalising behaviours when parental responsiveness was high compared to the equivalent externalising behaviours found in children with low effortful control exposed to low parental responsiveness or children with high effortful control exposed to low or higher parental responsiveness (Slagt et al. 2016b). Finally, children with higher sensory processing sensitivity had the lowest amount of externalising behaviours given decreases in negative parenting or increases in positive parenting (Slagt et al. 2018). Likewise, children with higher sensory processing sensitivity had a decrease in externalising behaviour over time if negative parenting was lower (Slagt et al. 2018). As per the vantage sensitivity model, these results are in comparison to children with low sensory processing sensitivity exposed to low negative or high positive parenting or children with high or low sensory processing sensitivity supposed to high negative or low positive parenting who have equivalent externalising behaviours (Slagt et al. 2018).

Keeping in mind that vantage sensitivity occurs when reactive children perform significantly better than non-reactive children in positive environments, boys high in testosterone had the lowest reactive aggression when harsh parental discipline was low (Chen et al. 2018), boys with a low RSA baseline had the most empathic concern when parental negative interactions were low (Van der Graaff et al. 2016) and children with higher cortisol reactivity had the fewest conduct problems when maternal sensitivity was high (Wagner et al. 2017). Thus, children without these sensitivity characteristics, as measured, or parental environments had equivalently more reactive aggression and conduct problems and equivalently less empathic concern.

In terms of genotype, Sulik et al. (2014) found girls with the COMT val158met val/met or val/val allele had the most inhibitory control and effortful control when maternal sensitivity and warmth was high. Girls with the met/met allele or who were exposed to lower maternal sensitivity and warmth had equivalently less inhibitory control and effortful control. On the other hand, boys with the met/met allele were found to have the most inhibitory control when maternal sensitivity and warmth was high. Boys with the met/val or val/val allele or who were exposed to lower maternal sensitivity and warmth had equivalently less inhibitory control.

### 3.4.4 Evidence for one-way contrastive effects

Two studies report a one-way contrastive effect (Zarra-Nezhad et al. 2014; Van der Graaff et al. 2016). Zarra-Nezhad and colleagues (2014) found that children with high behavioural withdrawal had a decrease in internalizing behaviours if psychological and behavioural control decreased, but children with low behavioural withdrawal had an increase internalizing behaviours if psychological and behavioural control decreased. However, the low and high withdrawal groups always had fewer or more internalizing behaviours, respectively, at high levels of behavioural and psychological control and these internalizing behaviours increased or decreased, respectively, as behavioural and psychological control decreased until the two groups had equivalent internalizing behaviours. Van der Graaff and colleagues (2016) reported that children with high RSA at baseline had an increase in empathic concern if parental negative interactions decreased, but children with low RSA at baseline had a decrease in empathic concern if parental negative interactions decreased. However, the high and low bassline RSA groups always had lower or higher empathic concern, respectively, at high levels of parental negative interactions and empathic concern increased or decreased, respectively, as parental negative interactions decreased until the two groups had equivalent empathic concern. In other words, children with high behavioural withdrawal and high baseline RSA had diathesis stress responses, whilst children with low behavioural withdrawal and low baseline RSA had vantage sensitivity responses.

### 3.4.5 Evidence for two-way contrastive effects

Two interactions using temperament as the measure of reactivity and two interactions employing polygenic sensitivity measures report two-way contrastive effects. Specifically, children with high sensory processing sensitivity had a decrease in externalising behaviours if negative parenting decreased or positive parenting increased, but the children with high sensory processing sensitivity had an increase in externalising behaviours if negative parenting increased or positive parenting decreased. On the other hand, children with low sensory processing sensitivity had an increase in externalising behaviours if negative parenting decreased or positive parenting increased, but the children with low sensory processing sensitivity had a decrease in externalising behaviours if negative parenting increased or positive parenting decreased. Likewise, children who scored high in the polygenic sensitivity score had the lowest hostility and internalising and externalising behaviours if parenting quality was good, but the most hostility and internalising and externalising behaviours if parenting quality was bad. Children who scored low on the polygenic sensitivity score, however, had the opposite relationship with parenting quality.

### 3.4.6 Summary

Overall, these results suggest that evidence has been found for each of the theories using a range of reactivity measurement types, behavioural outcomes and parenting components. However, there is variability in the quantity of support for each theory. A further exploration for why this variation in evidence occurs is provided below by examining the components that make up these interactive models. Systematic analysis seeks to determine if particular forms of measurement or analysis predispose to outcomes supporting a particular theoretical prediction.

## 3.5 Method used to assess reactivity

The first question we might ask is does the measurement type of reactivity make a difference. In Table 3.2, the evidence supporting the theoretical models is broken down by how the study operationalised reactivity to the environment. For example, 19 interactions of differential susceptibility used genotypes operationalised with SNP's or haplotype.

A broad overview of Table 3.2 suggests that finding consistent trends in differential support for the theories by reactivity measurement is complex. The only cells in Table 3.2 with potentially enough information to make comparisons by reactivity measure are genotype SNP and survey reported temperament. The other rows lack a sufficient sample size.

**Table 3.2**: Interactions providing evidence for the theories by assessment method of reactivity. NS refers to non-significant interactions. Number in parentheses refer to interactions excluding Belsky et al. (2015) and Janssens et al. (2017). Differential susceptibility (DS), diathesis stress (DA), vantage sensitivity (VS), one-way contrastive (CE1), two-way contrastive (CE2) and observation (OBS).

| | | DS | DA | VS | CE1 | CE2 | NS |
|---|---|---|---|---|---|---|---|
| Genotype | SNP or haplotype | 19 (17) | 12 (10) | 3 | | | 335 (47) |
| | Polygenic | 2 | | | | 2 | 0 |
| | And temperament (OBS) | | 3 | | | | 3 |
| Temperament | Observation | 2 | 1 | | | | 3 |
| | Survey | 7 | 17 | 4 | 2 | 2 | 79 |
| Endophenotype | Cortisol | 1 | | 1 | | | 16 |
| | Testosterone | 1 | 1 | 1 | | | 1 |
| | SCL | 1 | 1 | | | | 6 |
| | RSA | 1 | 1 | | 1 | | 13 |

Regarding genotype SNP, there was a slight leniency to more evidence for differential susceptibility. Specifically, 19 interactions supported differential susceptibility, 12 supported diathesis stress, and 3 supported vantage sensitivity. However, this difference does not seem systematic enough considering that at minimum 47 interactions found insufficient evidence for a theoretical model and error could be influencing the relative frequencies of support for a theoretical model.

Temperament measured in surveys showed an opposite leniency towards evidence for diathesis stress over differential susceptibility. Seventeen interactions supported diathesis stress, 7 supported differential susceptibility, 4 supported vantage sensitivity, 2 supported one-way contrastive effects, and 2 supported two-way contrastive effects. Again, however, 79 interactions found insufficient evidence for a theoretical model suggesting there could be noise distributed between the evidence for theoretical models.

There were too few interactions to make generalisations about most of the other methods of assessment. Specifically, they showed a mix of evidence for differential susceptibility, diathesis stress, vantage sensitivity, and one and two-way contrastive models. Further, they suffered a similar preponderance for insufficient evidence to support one of the theoretical models.

The polygenic score, interestingly, always found support for one of the theoretical models and was a 50-50 mix of differential susceptibility and two-way contrastive effects. The small number of interactions make generalisations difficult, but the polygenic approach may offer promise for finding consistent effects in the future.

### 3.6 Method used to assess reactivity and parenting

Table 3.3 breaks down the evidence supporting the theoretical models by both the type of measurement for reactivity and the type of measurement of the parenting environment. For example, 6 interactions that used genotype SNP and observations of parenting found support for differential susceptibility, whilst 13 interactions that use genotype SNP and surveys of parenting found support for differential susceptibility.

This Table (3.3) has the same problem of Table 3.2 as the cells are sparsely populated except for genotype SNP and temperament surveys.

Looking at the results for the genotype SNP, when parenting was measured by survey only 2 interactions of diathesis stress were observed compared to 9 diathesis stress interactions when parenting was observed. Further, 13 interactions of differential susceptibility were observed when parenting was measured by survey, whilst only 6 interactions of differential susceptibility occurred when parenting was observed. This suggests that parenting measured by surveys leans towards differential susceptibility when using genotype SNP. Keeping in mind, however, that most findings provided insufficient evidence for any theoretical model and thus there could be error within these comparative estimates.

If we look at the results the surveys of temperament, we find a trend opposite to that of genotype for how parenting was measured. Surveys with report measures of temperament combined with surveys of parenting did not find evidence for differential susceptibility but found 8 interactions of evidence for diathesis stress and 3 interactions of evidence for vantage sensitivity. Combining a survey of temperament with an observation of parenting, however, we see there is 7 interactions of differential susceptibility, 9 interactions of diathesis stress and one interaction of vantage sensitivity. This suggests observations of parenting increases likelihood of finding differential susceptibility when temperament is surveyed. Of course, this should be seen in the light of most interactions showing insufficient evidence for either theoretical model and thus error could be within these comparisons.

**Table 3.3**: Interactions providing evidence for the theories by assessment method of reactivity and parenting. NS refers to non-significant interactions. Number in parentheses refer to interactions excluding Belsky et al. (2015) and Janssens et al. (2017). Differential susceptibility (DS), diathesis stress (DA), vantage sensitivity (VS), one-way contrastive (CE1), two-way contrastive (CE2), observation (OBS) and survey (SUR).

| | | DS | DA | VS | CE1 | CE2 | NS | Parent |
|---|---|---|---|---|---|---|---|---|
| Genotype | SNP or haplotype | 6 (5) | 9 (7) | 3 | | | 76 (27) | OBS |
| | SNP or haplotype | 13 (12) | 2 | | | | 256 (17) | SUR |
| | SNP or haplotype | | 1 | | | | 3 | SUR&OBS |
| | Polygenic | 2 | | | | 2 | 0 | OBS |
| | And temperament (OBS) | | 3 | | | | 3 | OBS |
| Temperament | Observation | 2 | | | | | 2 | OBS |
| | Observation | | | 1 | | | 1 | SUR&OBS |
| | Survey | 7 | 9 | 1 | | | 32 | OBS |
| | Survey | | 8 | 3 | 2 | 2 | 47 | SUR |
| Other | Cortisol | 1 | | 1 | | | 16 | OBS |
| | Testosterone | 1 | 1 | 1 | | | 1 | SUR |
| | SCL | 1 | 1 | | | | 6 | OBS |
| | RSA | 1 | 1 | | 1 | | 13 | SUR |

Combining the insights from the genotype SNP and surveys of temperament contradictory statements could be made in terms of which provides the most accurate way to assess the theoretical models. For example, it could be said that surveys of parenting lack sufficient information to observe differential susceptibility when temperament is measured by survey. However, when using a genotype SNP the measurement of parenting by a survey improves the odds of finding differential susceptibility. These contradictions illustrate that making comparative statements based on what measures studies have included in their models can be quite complex.

### 3.7 Method used to assess reactivity and behaviour

Table 3.4 disaggregates the reactivity measurements by behavioural measurement. For example, an observation of behaviour and genotype SNP for reactivity found 1 incidence of differential susceptibility, whilst using a survey for behaviour and genotype SNP found 18 interactions of differential susceptibility.

An examination of Table 3.4 illustrates that only 8 interactions of evidence for one of the theories used observations or a combination of observations and surveys to assess behaviour. Combining genotype SNP with an observation of behaviour indicated diathesis stress in 4 interactions, whilst differential susceptibility was only observed once. Observations of temperament combined with observations of behaviour indicated support for differential susceptibility twice, and support for diathesis stress once. Therefore, support for the predominant theory for both genotype SNP and temperament observations is reversed. It is likely, therefore, that there are too few observations to make meaningful comparisons and this applies to the other measurements used to operationalise reactivity.

**Table 3.4**: Interactions providing evidence for the theories by assessment method of reactivity and behaviour. NS refers to non-significant interactions. Number in parentheses refer to interactions excluding Belsky et al. (2015) and Janssens et al. (2017). Differential susceptibility (DS), diathesis stress (DA), vantage sensitivity (VS), one-way contrastive (CE1), two-way contrastive (CE2), observation (OBS) and survey (SUR).

| | | DS | DA | VS | CE1 | CE2 | NS | Behaviour |
|---|---|---|---|---|---|---|---|---|
| Genotype | SNP or haplotype | 1 | 3 | | | | 0 | OBS |
| | SNP or haplotype | 18 (16) | 8 (6) | 3 | | | 332 (44) | SUR |
| | SNP or haplotype | | 1 | | | | 3 | SUR&OBS |
| | Polygenic | 2 | | | | 2 | 0 | SUR |
| | And temperament (OBS) | | 3 | | | | 3 | SUR |
| Temp | Observation | 2 | | | | | 2 | OBS |
| | Observation | | 1 | | | | 1 | SUR&OBS |
| | Survey | 7 | 17 | 4 | 2 | 2 | 79 | SUR |
| Other | Cortisol | 1 | | 1 | | | 16 | SUR |
| | Testosterone | 1 | 1 | 1 | | | 1 | SUR |
| | SCL | 1 | 1 | | | | 6 | SUR |
| | RSA | 1 | 1 | 1 | 1 | | 13 | SUR |

In summary, inconsistency in measurement and a tendency to use a single measure of reactivity, rather than combined measures, limits conclusion on systematic association of measurement and substantive findings.

### 3.8 Quality of evidence

Table 3.5 contains a disaggregation of support for the theories based on the quality of the evidence graded using the Newcastle-Ottawa scale (Wells et al. 2015),

**Table 3.5**: Average quality rating for the interactions providing support for the theories. Differential susceptibility (DS), diathesis stress (DA), vantage sensitivity (VS) one-way contrastive (CE1) and two-way contrastive (CE2). Some interactions were excluded from the criteria averages as they had complete data (none missing) or only used single scale items which do not permit sensitivity analysis. Raw data is presented in Table 3.1.

| Theory | Info. sources | Adjusted | ROS | Sensitivity analysis | Plot data | Consistent | Missing | Interactions |
|---|---|---|---|---|---|---|---|---|
| DS | 1.76 | 0.38 | 0.97 | 0.23 | 0 | 0.09 | 0 | 33 |
| DA | 1.51 | 0.27 | 1.00 | 0.08 | 0 | 0.08 | 0.03 | 37 |
| VS | 1.50 | 0 | 1.00 | 0 | 0 | 0 | 0 | 9 |
| CE1 | 1.00 | 0 | 1.00 | 0 | 0 | 0 | 0 | 3 |
| CE2 | 1.50 | 0 | 1.00 | 0.50 | 0 | 0 | 0 | 4 |

Examining each quality criteria separately, we start with the independence of measures. The interactions that supported differential susceptibility, diathesis stress, vantage sensitivity and two-way contrastive effects had an elevated level of independence between measurements. Ranging from 1.5 to 1.76, this indicates the information used in these interactions may not have been heavily biased by having the same informants for each measurement. Nearly all measurements were independent. On the other hand, support for one-way contrastive effects came from a medium level of independence with an average score of one. This indicates on average only two of the measurements of the child characteristic, parenting environment and behaviour were obtained by reports from the same individual. Finally, only two significant interactions, which supported diathesis stress, were based on reports from the same person (a score of 0) whilst all other interactions scored one or higher. This illustrates the independence of the measurements was high.

The criteria of whether studies adjusted for multiple comparisons raises concerns for the quality of statistical rigour of the interactions. For vantage sensitivity and contrastive effects, no studies adjusted the level of statistical significance for the number of statistical comparisons undertaken. Further, only 27% of interactions that supported diathesis stress adjusted for multiple comparisons. However, 38% of interactions supporting differential susceptibility adjusted for multiple comparisons. Overall, however, this shows that statistical rigour is lacking given that most interactions supporting one of the theoretical models did not adjust for multiple comparisons.

All interactions, except for one supporting differential susceptibility, used regions of significance tests. The interaction that did not use regions of significance testing used a latent class trajectory approach to their statistical model. Ascertaining regions of significance is likely complex with this method and therefore it was included for such a reason. This high number is primarily a result of excluding studies that did not analyse regions of significance from this review.

Very few of the interactions were put through sensitivity analysis regarding how the variables used were created. Specifically, the highest individually observed score of 1 indicated that some sensitivity analysis of how variables were created was performed. Further, the average score for two-way contrastive effects was 0.5, differential susceptibility was 0.23, diathesis stress was 0.08 and one-way contrastive effects and vantage sensitivity were 0. This indicates many of the results were not thoroughly evaluated for how the variables were created. It could be important to evaluate the sensitivity of findings to how variables are created when there are multiple decisions in terms of combining multiple scales of similar measurement or similar measures reported by different people. However, the results illustrate this robustness check has not been performed in a systematic, widespread manner.

No studies included in this review provided a plot of the raw data or partial residuals to evaluate their predicted trend lines. This is a shortcoming as plotting the raw data is a relevant way to demonstrate the validity of findings trends. By including this figure other researchers can evaluate how believable the predicted trends are given the distribution of the raw data. For example, are there outliers which appear to be guiding a significant interaction such that their removal might make it disappear? Plotting the raw data is an area that can be improved for research regarding support for these theoretical models.

The results in Table 3.5 demonstrate that there were very few consistent effects regarding support for one of the theoretical models within each study. Consistent effects were defined as support for the same theoretical model being observed using similar measures of behaviour, similar measures of parenting or similar measures of the moderating characteristic. Looking at the evidence, no interactions that provided support for vantage sensitivity or contrastive effects were consistent within a study. On the other hand, 9% of interactions supporting differential susceptibility and 8% of interactions supporting diathesis stress were consistent. However, these consistent effects were only observed in studies that examined a single interaction. Finding more consistent effects will be a relevant consideration for future studies. Or, the inconsistencies illustrate that support for the theoretical model can be complex and context-dependent.

The final quality criteria assessed whether studies with missing data evaluated the sensitivity of their results to the approach they used to handle missing data. No interactions that supported diathesis stress, vantage sensitivity, or contrastive effects evaluated the sensitivity of the result to how missing data was handled. Furthermore, only 3% of results for diathesis stress evaluated sensitivity to missing data methods. Given that missing data can be more problematic with interactions between variables (Bartlett et al. 2015), research in this area could improve by evaluating how different approaches to handling missing data influence the results.

To summarise the results of this section, the analytical quality of the interactions supporting the theoretical models was low and no apparent quality trends explained support for the theories. Studies performed well in terms of independence of measurement and performing regions of significance tests. However, there was only adequate performance in terms of adjusting for multiple statistical comparisons. Further, there was a distinct lack of sensitivity analysis for how variables were created, an absence of plotting raw data to evaluate the results, inconsistent effects and a very low evaluation rate of missing data sensitivity.

### 3.9 Discussion

This review examined how phenotypic, endophenotypic, and genetic characteristics of children moderate the effect of the parenting environment on internalizing and externalizing behaviours and social and emotional competence in children. The review specifically evaluated whether these moderating interactions were consistent with the diathesis stress, differential susceptibility or vantage sensitivity hypotheses. The discussion below focuses on several key themes.

### 3.10 Support for the interactions

Interestingly, most of the interactions evaluated did not support a theoretical model. Excluding two studies that were notably exploratory, nearly half of all interactions investigated were non-significant. This high number of non-significant findings indicates that there may be substantial measurement error or a low signal within the systems investigated such that making inferences on parameters is difficult.

### 3.11 Support for the theories

This review demonstrated that evidence for differential susceptibility, diathesis stress, and vantage sensitivity could all be found within the literature of child characteristics moderating the effect of parenting on behavioural outcomes in children. Moderating characteristics included temperament measures, endophenotypic measures, and genetic variability.

The support for the theoretical models varied substantially, however, between and within different measurement types of reactivity, parenting and behaviour. This made it impractical to be able to ascertain which theory should be expected given a type of reactivity measurement, a certain parenting environment or a type of behavioural outcome. Indeed, most studies varied in how they operationalised each measurement component and this makes comparisons between studies incredibly difficult. Below this is explored in more detail.

## 3.12 Support for the theories: the relationship with measurement

### 3.12.1 Reactivity:

The results of the review suggest that it did not matter how reactivity was operationalised as support for nearly all the theoretical models was found for each measurement type. This may mean individual studies do not generalise to other contexts. It may be that there are many *individual* × *environment* interaction effects and each child is individually reactive to different environments for different reasons that have different effects on developmental outcomes (Belsky et al. 2007). Consequently, if the theoretical model supported can change with reactivity measurements then the global predictive value of understanding differences in reactivity may be of lower consequence.

What may be valuable, therefore, is to examine these reactivity measurements between studies whilst holding constant key parts of the equation. Specifically, a broad range of parenting and behavioural measurements were used in each study and this makes it hard to develop a comparative base to understand the leniency is towards one theory or another based on reactivity measurement.

The polygenic approaches to reactivity appeared to have consistent results supporting either the differential susceptibility or contrastive model. Though only one study used this approach, four of four interactions assessed demonstrated an effect. Future research using these polygenic approaches may be a promising avenue to identify differential susceptibility or even contrastive reactivity.

### 3.12.2 Parenting:

The results in the review demonstrate that parenting was assessed via survey or observation in similar frequency. It appears then that how parenting is measured is not a substantial explanatory component of support for the theoretical models. Generally, observations of parenting have been considered higher quality as they reduce respondent bias. However, surveys of parenting have been validated and can sometimes better capture parenting over a broader period of time. In any case, support for the theories seems to be independent of how parenting was measured and further exploration of this topic holding reactivity and behavioural components constant would be required to properly evaluate such effects. One study did compare survey and observation but they did not estimate the proper statistical criteria to evaluate the theories so we can't make a comparison other than whether it was statistically significant or not.

### 3.12.3 Outcome:

Support for the theoretical models was found for externalising behaviours, internalising behaviours and prosocial behaviours. Moreover, the majority of behavioural measurement was based on surveys. This leads us to conclude that the behaviour measured and type of measurement does not dictate which theoretical model will be supported. To date no authors have suggested that the theoretical models should support only certain types of behaviours, and this largely confirms the absence of a theoretical push. The absence of substantial observational work on behaviour suggests this may be an avenue for future research to compare surveys and observations of behaviour as it has not been evaluated thoroughly.

### 3.12.4 Measurement quality:

A positive finding for most studies and the evidence collected is that there was an elevated level of independence between measurements. Very few studies used the same respondent to measure child characteristics, child behaviour, and the parenting environments. This also implies that variation in support for the theories cannot be explained, based on the summary statistics used in this review, by a lack of measurement independence.

However, the measurement information evaluated by the review could be refined. For example, an exploration of the measurement quality (including reliability and validity) could be undertaken to further assess how measurement properties impact support for the theoretical models.

## 3.13 Statistical analysis and quality

In general, this review demonstrated that the statistical criteria applied to evaluate the theories could be improved. Specifically, few studies adjusted for the number of parameters estimated. This is a basic component to any multivariate study involving the calculations of many frequentist interactions. Without accounting for the number of parameters being estimated the researcher leaves open the possibility that the results occurred by chance. Indeed, given the high number of non-significant findings this could have been a problem. On a similar note, there was little sensitivity analysis regarding how variables were created. When variables are being used that can take multiple forms, for example combining internalising and externalising behaviours compared are looking at them separately, it is useful to perform a sensitivity analysis where these decisions that were made innocently are evaluated with scrutiny. This can demonstrate that it does not matter how the parenting, behavioural or reactivity measurements are created as the same result was found consistently. It may be the case that many authors did this and did not report the results due to space constraints. It may be important, however, that the sensitivity analysis is undertaken and demonstrated to the audience reading the paper to ensure robustness of the results. Moreover, many studies used a method to handle missing data but rarely explored what happens if they used only the complete cases or used an alternate model for missing data. Imputing or modelling missing data with interactions can be particularly complex and time consuming. However, demonstrating that an analysis is robust to missing data methodology will improve the quality of the research in this area. Additionally, no studies plotted the raw data or partial residuals associated with their interactions. Plotting the raw data or partial residuals in part allows the reader to evaluate the strength of support for an observed relationship (Fox & Weisberg et al. 2017). If there are only a few outliers surrounding a key part of the analysis it would indicate that the results are potentially weak. Conversely, raw data that supports the modelled interaction demonstrates a heuristic of strength that the results may be trusted and more fully understood by the reader. Consistent results were, therefore, hard to come by.

### 3.14 Implications for future research

The results of this review highlight two main implications for future research investigating interaction effects of neurological reactivity and the early care environment on child behavioural development. First, to improve replicability and comparisons across studies the measurement of the environment and reactivity in context of other results needs additional consideration. Second, as estimated effects are sensitive to analytical decisions made by researchers there is a need for sensitivity analysis of estimated effects.

Regarding improved measurement, an ideal situation would involve identifying multi-domain reactivity and evaluating how that moderates the effect of parenting measures on behavioural outcomes. The choice of environments and outcomes would be consistent and adequately implemented to reduce measurement error. Nearly every study review used a different implementation of parenting and behaviour and that made it difficult to estimate how reactivity moderated associations between these measures.

The improvement in measurement would also address the observed sensitivity of results between and within studies that used different behavioural, parenting and reactivity measures. Specifically, systematic experimentation with changes in one or two of the measurements would be valuable. By holding some of the measured variables constant, such as reactivity, and making incremental changes, such as different measurements of parenting e.g., emotional climate compared to parental psychological and behavioural control, the more consistent trends for support of the theoretical models can be established.

Though the field already aims to approximate this ideal with some studies using multiple behavioural and parenting measures and a general coherence between the constructs attempted to be measured and investigated (e.g., parental warmth and sensitivity), it is clearly apparent that more coordinated approaches are needed. The lack of consistency between studies suggests finding ways to align and reduce variation between study design and implementation will be valuable to have evidence that converges toward reliable and replicable evidence for the theoretical models.

Finally, the review highlighted that the majority of interaction effects could not be reliably estimated based on statistical significance. Studies need to account for the possibility that the level of uncertainty on the estimated interaction effects may be large and cautiously report results where multiple interactions were estimated but few were statistically reliable. Additionally, although not evaluated in this review, studies should calculate the power of their sample to detect interaction effects in comparison to an optimum design (McClelland & Judd 1993). Findings from studies with high power would be more informative then those of studies with small sample sizes, non-optimum designs, and high measurement error. Further, as there are often multiple analytical decisions made in terms of variable creation and model inclusion, relevant sensitivity analysis for these decision points can be undertaken to reduce the likelihood of following spurious associations.

### 3.15 Application of results for current thesis

As the review demonstrated, measures of child neurological reactivity were associated with moderating the parenting environment consistent with diathesis stress, differential susceptibility, vantage sensitivity, and one-way and two-way contrastive effects. Support for these moderations, however, were far from consistent and there was large variability in the measurements used by each study. The inconsistency in results and variability in measurement implies a need for additional research into the hypotheses for the moderating effects. This additional research could focus on keeping some measurements in the model constant, whilst systematically varying other measures to identify consistencies in evidence. Further, sensitivity analysis of decision points in the analytical process can illustrate how sensitive the results are to different researcher choices.

The empirical component of this thesis, therefore, uses data from two large-scale social science surveys and observations to examine consistencies in support for diathesis stress, differential susceptibility, vantage sensitivity, and one-way and two-way contrastive effects by making systematic changes to key methodological components used in the analysis combined with rigorous sensitivity analysis.

Chapter 4 does this by examining how parent reported temperament moderates the effect of parent reported parenting on parent and teacher reported behavioural development.

Chapter 5, using two data sets, examines how parent reported infant temperament moderates the effect of parent reported and observed parenting on parent reported behavioural development.

The main change between chapters include assessing temperament at different ages, using parent report and observed measures of parenting, and using multiple measures of behaviour and temperament.

## Chapter 4: Testing individual difference models in a longitudinal sample of Australian children: Parenting, child temperament, and behavioural outcomes

### 4.1 Introduction

Individual difference theoretical models of child development have emerged as a focus of significant interest in developmental science both because they have a strong basis in biological theory of inter- and intra-species diversification and because they offer promise for tailoring intervention strategies. Moving beyond uniform environment or child effect models, *child* × *environment* accounts examine how children may be predisposed to different developmental trajectories dependent on their social environments due to observable characteristics such as genotype or temperament (Belsky & Pluess 2009; Moffitt et al. 2005).

Three theoretical models encapsulate the dominant thinking in *child* × *environment* interactions. Risk based, diathesis stress models imply the characteristics of some children predispose them to poor outcomes in developmentally harsh social environments (Monroe & Simons 1991; Zuckerman 1999). Cumulative advantage, vantage sensitivity models posit that characteristics of some children predispose them to optimal outcomes in developmentally rich environments (Pluess & Belsky 2013; Sweitzer et al. 2012). Finally, for better or worse, differential susceptibility and biological sensitivity to context models posit that characteristics of children may make them susceptible to poor outcomes in developmentally harsh environments, yet the same characteristics make them susceptible to optimal outcomes in developmentally rich environments (Belsky et al. 2009; Ellis et al. 2011).

One major field of inquiry has focussed on child temperament as a moderator of the association between the parenting environment and child behavioural development (Kiff et al. 2011). A growing literature in this field has documented empirical support for interactive models whether diathesis stress (Gallitto 2015; Roisman et al. 2012), differential susceptibility (Kim & Kochanska 2012; Roisman et al. 2012) or vantage sensitivity (Slagt et al. 2016b; Zarra-Nezhad et al. 2014) models. However, within this literature there are notable variations in support-evidence for the theoretical models. As documented in the systematic review presented in chapter 3, of 542 interactions that tested for moderating effects of temperament, 33 supported diathesis stress, 33 showed evidence of differential susceptibility, 9 observed evidence for vantage sensitivity, 3 supported an alternative one-way contrastive model, 4 supported another alternative two-way contrastive model and 456 found statistically insufficient evidence of moderation.

The current study undertook analyses of a large, longitudinal sample of Australian children to examine the moderating effects of child temperament on the relationship between parenting and behavioural outcomes. The study applied a detailed analytic plan to provide substantive testing of interactive theories with an express focus on undertaking a critique of emergent findings in the context of both measurement and analytic strategy. The study tests individual difference theory in a population outside the United Kingdom and United States of America, where most of the analyses have been conducted, and advances knowledge through critical attention to analytical pathways and measurement.

### 4.2 Measurement: Parenting, child temperament and child behaviour

A considerable number of studies have documented the association of parenting environments and child behavioural development, with more recent studies attending to the moderating role of child characteristics, particularly temperament. Figure 4.1 presents the basic moderation model. As indicated in Figure 4.1, a range of measurements are employed to assess the key components of the model. The conceptual basis of these model components are briefly outlined and discussed below.

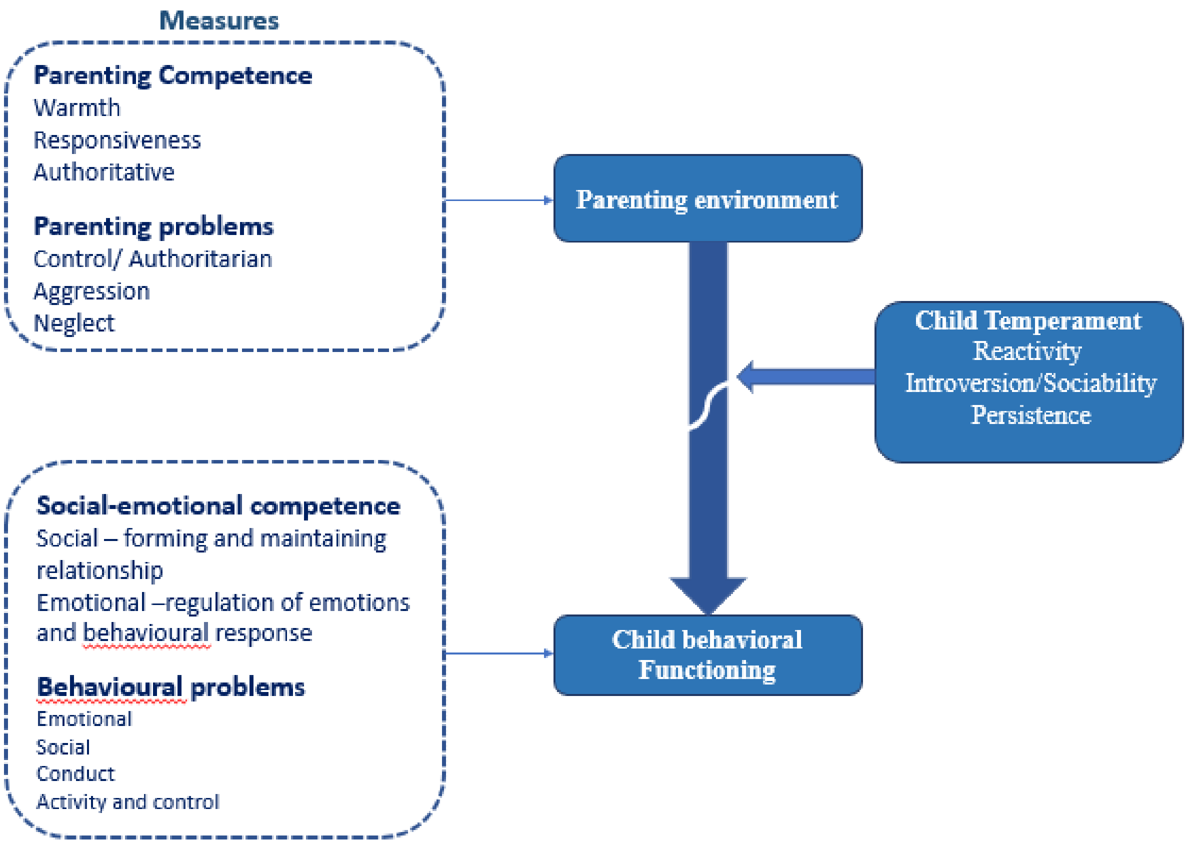



**Figure 4.1:** Moderating role of temperament on the effect of parenting on child behavioural functioning.

*Parent Care environment:* broadly refers to all actions by parents related to raising children and is one of the most strongly implicated factors in child behavioural development (Pinquart 2016a, 2017). Measurement of parenting includes both parent competency in supporting their child's social and emotional functioning and learning, including parental behavioural and psychological control, emotional quality of parent-child relationships, and parent responsiveness to children's emotions and needs (Kiff et al. 2011; Maccoby 2000; Pinquart 2016a, 2017; van den Boom 1994). Regarding parent emotional competency, it is typically conceptualised in terms of warmth, closeness, and acceptance compared to negativity and rejection (Clark & Ladd 2000; Maccoby 1992; Spera 2005). The warmth component measures affection, positive emotions, involvement, and admiration of parents with their children (MacDonald 1992). Parenting environments high in warmth and acceptance are associated with fewer externalizing and internalizing behaviours and greater social and emotional competence in children (Chen et al. 2000; Davidov & Grusec 2006; Rothbaum & Weisz 1994; Stormshak et al. 2010). Conversely, negativity, rejection, and a lack of emotional support from the parents for their children captures the opposite side of parenting emotional quality (Rohner 2004) and is associated with more internalizing and externalizing behaviours and reduced social and emotional competence (Deater-Deckard et al. 2009; Knafo & Plomin 2006; Miner & Clarke-Stewart 2008; Viding et al. 2009). Measurement of parenting is typically made through surveys answered by parents, children, teachers and other care-givers or through codified researcher observations of the parent. Further, studies often create psychometric indices of parenting that combine similar measurements, as single items may not capture the breadth of parenting experienced by a child.

*Child behavioural response to environment:* Social and emotional competence and behavioural difficulties, both externalising and internalising, are constructs that collectively measure behavioural functioning. Social competence are behaviours that facilitate positive social interactions and maintain peer relationships (e.g., cooperation and sharing; Rose-Krasnor 1997), whilst emotional competence involves using identification and regulatory behaviours to manage emotions and emotions arising through social interactions with others (Denham et al. 2003; Saarni 1990). Externalising and internalising behaviours, on the other hand, are considered as maladjustment that contrast with positive or adaptive social-emotional competence development (Ladd & Profilet 1996). Measures of externalising behaviours observe or seek reports on overt negative behaviours directed towards other people and objects and includes aggression, defiance and antisocial conduct (Collett et al. 2003). Measures of internalising behaviours, on the other hand are observed or reported emotional responses symptomatic of depression, anxiety and rumination. These are typically less overt and therefore may be harder to observe (Myers et al. 2002).

*Temperament:* has been defined and measured as *constitutionally based individual differences in reactivity and self-regulation in the domains of affect, activity and attention* (Rothbart & Bates 2006, pg. 100). Constitutional in this definition refers to the biological basis of temperament, in the form of genetic expression, which is subsequently shaped by interactions with a child's environment over time. Reactivity encapsulates how responsive children are to changes in the internal and external environment. Reactivity can be both specific (e.g., emotion in response to negative events) and general (e.g., fear), and is measured by the response threshold, length, duration, and intensity of a child's physical, affective, and attentional reactions to stimulus (Rothbart & Derryberry, 1981). Self-regulation, on the other hand, refers to the processes used to control and regulate a child's reactivity. Temperament measures in the form of surveys and observations thus aim to assess the predispositions of children to respond and react to the environment (Thomas & Chess 1977).

Children's temperaments have been classified as difficult (high withdrawal, intense emotions and negativity and low adaptability and rhythmicity), easy (opposite of difficult), slow-to-warm-up (low activity, approach and adaptability and high negativity), or intermediate (any other combination) (Carey & McDevitt 1978). Sub-components of temperament are often examined separately and include negative emotionality, effortful control, impulsivity, and inhibition. Children characterised with a difficult temperament and higher negative emotionality, lower effortful control, higher impulsivity, and higher inhibition have been associated with the development of more externalizing and internalizing behaviours and lower social and emotional competence (Capsi et al. 1995; Frick & Morris 2004; Muris & Ollendick 2005; Rothbart 2007; Sanson et al. 2004).

### 4.3 Theory testing: Temperament as a moderator of parenting environment

Although parenting and temperament may influence behavioural development, empirical evidence shows that the effects of parenting on behavioural development can be moderated by the temperament of the child (e.g., Kiff et al. 2011). By capturing regulatory differences in reactivity and self-regulation to experience, temperament has been able to index children predisposed to the effects of the quality of the parenting environment.

The shape of this moderating effect, however, has drawn interest and evaluation within the child development literature. Risk based models have defined the moderating role of temperament in terms of a diathesis to stress. In this model, children with difficult temperaments, high reactivity, and low self-regulatory capability are predisposed to develop behavioural difficulties when the environment is poor due to their heightened sensitivity. On the other hand, the vantage sensitivity model posits the interaction between the environment and temperament is one of cumulative advantage. Due to differences in reactivity, in the context of a positive environment reactive children may be predisposed to exhibit better behavioural development as they could be more receptive to positive cues and reward. Finally, differential susceptibility thinking supposes that children exhibiting more reactive and difficult temperament may be predisposed to develop behavioural problems in poor environments, yet have better behavioural development in rich environments. The reactivity would thus operate in a manner of for 'better or worse' (Belsky et al. 2007).

Research, nonetheless, has found that difficult, reactive and low self-regulatory temperaments have moderated the effect of parenting on the behavioural development of children consistent with both the diathesis stress, differential susceptibility and vantage sensitivity accounts. For respective examples; Gallitto (2015) found children with difficult temperaments had more externalising behaviours when there was a lack of positive parenting compared to children with easy temperaments; Roisman et al. (2012) found children high in difficult temperament consistently had fewer teacher rated internalising and externalising behaviours and more teacher rated social competence in higher maternal sensitivity environments compared to children with easy temperaments, however, the children with difficult temperaments had more internalising and externalising behaviours and less social competence when maternal sensitivity was low; and, children with low behavioural withdrawal, compared to children with high behavioural withdrawal, had fewer internalising behaviours when paternal behavioural and psychological control was low (Zarra-Nezhad et al. 2014). The systematic review presented in chapter 3 further highlights this complexity. Specifically, this review identified that evidence for the theoretical models was inconsistent and most of the time (84%) there was insufficient evidence of moderation to make statistically robust conclusions.

Given research investigating the moderating role of temperament has had such inconsistent findings, it is possible that additional and consistent evidence can be observed by performing extended analyses with large, longitudinal surveys containing data from multiple respondents. The use of large surveys may help overcome the inability to make statistically robust conclusions, expand the representativeness of the research sample, and can contribute additional information regarding how temperament moderates the effect of the parenting environment. Additionally, extended investigations with multiple sensitivity analyses may help find consistencies in moderating effects.

### 4.4 Current research

This study aimed to examine how child temperament moderates the effects of parenting emotional quality on behavioural development. The study used a large, longitudinal sample of Australian children with behavioural measurements from multiple respondents and a detailed and predetermined analytical plan to try and establish rigorous consistencies in evidence. Specifically, the study tested whether, and how, mother reported temperament components (difficult temperament, reactivity to negative experiences, non-persistence and introversion) moderate the effects of mother reported parenting emotional quality (warmth and harsh parenting) on mother and teacher reported externalizing and internalizing behaviours and social and emotional competence in children at ages 4, 6 and 8 and across time. A first exploratory hypothesis was that if temperament moderated the effect of parenting on behavioural development it could reasonably be expected to take the shape of either a diathesis-stress, differential susceptibility or vantage sensitivity model. A second direct hypothesis, however, was that if temperament moderated the effect of parenting on an aspect of behaviour, the interaction would be consistent at all ages and across time, consistent across reporters of behaviour, parenting environments and similar behaviour measures and generally consistent across temperament measures. This hypothesis implies a focus on the biological underpinnings of temperament and asks the evidence for moderating effects to be strong and consistent to provide useful research implications.

### 4.5 Methods

### 4.6 Sample

The Longitudinal Study of Australian children (LSAC) is a large longitudinal survey following two cohorts of Australian children recruited in across 2003-2004 to evaluate the influence of Australia's social and cultural environment on childhood development (LSAC 2016). The first cohort, the birth cohort, recruited infants younger than 1 year of age and was used in the current study. These infants were born between March 2003 and February 2004. The initial 5107 primary parents and their infants were selected using stratified random sampling from Australian Medicare records to be representative of all children in Australia. However, the Medicare record may not include all Australian residents and thus LSAC approximates a representative sample of children from urban and rural areas of all states and territories in Australia at baseline. A new wave of data was collected every two years, with waves 2 (age 4-5), 3 (age 6-7) and 5 (age 8-9) constituting the frame for key variables used in this study.

From the entire sample (5107), the analytical sample was restricted to surveys of the primary caregiver who filled out the survey in wave 1 to maintain consistency of the reporter of information (Murray et al. 2007; Kersten et al. 2016). Nearly all primary caregivers at age 0–1 were the child's biological mother and the sample was thus restricted to mother reported information (excludes 74 biological fathers and 14 others). Children were also excluded if at any time point (ages 0-1 to 10; wave 6 information included for these purposes) the parent indicated the child had a neurological or medical disorder including attention deficit disorder (ADD) or attention deficit hyperactive disorder (ADHD) (n=150), anxiety or depression (n=179), autism spectrum disorder (n=155), epilepsy (n=33), chronic fatigue (n=7), diabetes (n=12), blackouts, fits or loss of consciousness (n=58), mental illness for over 6 months (n=13), a nervous or emotional condition (that requires treatment) (n=46), any mental illness for which help or supervision is required long-term (n=47), long term effects as a result of a head injury, stroke or other brain damage (n=12) and chronic or recurring pain (n=109). Additionally, children were excluded if they had ever taken (n=91), or were taking (n=84), any medication for ADD or ADHD. Exclusion decisions were based on prior exclusions documented in the literature (e.g., Roisman et al. 2012) with the underlying rationale relating to the biasing effect of such disorders on parenting, temperament and behaviour (D'Souza & Karmiloff-Smith 2017; Kim et al. 2016). Following exclusions, the remaining sample comprised 3408.

Only children with complete information were included in the analysis and implications are discussed thereafter. Of the base sample of 3408 children, 1289 had information for all variables across the period of interest (4, 6 and 8 years). A comparison of children with missing and complete data revealed they had similar distributions for the key variables of interest in the moderation (maximum difference parent report = 0.3, teacher report = 0.39 of a pooled standard deviation, Appendix A). The first section of the results contains full information of this sensitivity comparison. Differences in results due to missing data may be minimal – due to similar distributions and a comparatively large sample size.

### 4.6.1 Procedure

Data from the primary parent were collected during home visits using face-to-face interviews and during which questionnaires were provided for postal return. Teacher data were collected with mail out questionnaires. Ages of children for the complete case sample in years were Mean (M) = 4.84 Standard Deviation (SD) = 0.23 at age 4 (wave 3), M = 6.85 SD = 0.29 at age 6 (wave 4), M = 8.93 SD = 0.31 at age 8 (wave 5).

## 4.7 Measures

### 4.7.1 Moderator: Temperament (ages 4 and 6)

Mothers completed a reduced version of the Short Temperament Scale for Children (STSC) at ages 4 and 6 (Sanson et al. 1994). Three STSC subscales were measured; reactivity to negative experiences (hereafter reactivity; *when shopping together, if I do not buy what this child wants (e.g. sweets, clothing), he/she cries and yells*), persistence (*when a toy or game is difficult, this child quickly turns to another activity*) and introversion/sociability (*when in a park or visiting, this child will go up to strange children and join in their play*). Each subscale is the average of four items regarding how often a child's behaviour matches a statement; *almost never=1* to *almost always=6*. Persistence was reverse coded so that it indicated non-persistence. Higher scores indicate higher reactivity, higher non-persistence and higher introversion. Difficult temperament is the overarching scale and is the average of reactivity, non-persistence and introversion. Higher scores indicate a more difficult temperament. Reliability coefficients ranged from α=0.67 to 0.81.

### 4.7.2 Environment: Parenting warmth and harsh parenting (ages 4, 6 and 8)

Parental warmth was measured with six questions from the Child Rearing Questionnaire (Paterson & Sanson 1999). Specifically, mothers were surveyed about their interactions with study child over the last six months in terms of question such as *how often did you express affection by hugging, kissing and holding this child?* Parents responded *never/almost never=1* to *all the time=5*. Reliability coefficients ranged from α=0.88 to 0.89.

Harsh parenting was measured using five items based on the National Longitudinal Study of Children and Youth (2005) implemented by Statistics Canada. The mother was surveyed for questions such as *how often are you angry when you punish this child?* In response, the parent answered *never/almost never=1* to *all the time=5*. Reliability coefficients ranged from α=0.65 to 0.68.

### 4.7.3 Outcome: Externalizing and internalizing behaviours and social-emotional competence (ages 4, 6 and 8)

Externalizing and internalizing behaviours and social-emotional competence were measured using the parent and teacher form of the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997; Kersten et al. 2016). The SDQ consists of five subscales including emotional difficulties (*often unhappy, depressed or tearful*), peer problems (*picked on or bullied by other children*), conduct problems (*often fights with other children or bullies them*), hyperactivity (*constantly fidgeting or squirming*) and pro-social (*considerate of other people's feelings*). Each subscale ranges from 0 to 10 and is the sum of five questions with three possible options: (reverse coded where relevant) '0 = not true', '1 = somewhat true', and '2 = certainly true'. Questions in the SDQ are consistent across time, though at age 4 two items were different to coincide with toddler appropriate wording. Following standard procedures the subscales were aggregated into internalizing (emotional difficulties and peer problems; 0–20), externalizing (conduct problems and hyperactivity subscale; 0–20) and total difficulties (internalizing and externalizing; 0–40) scores. Higher scores for the emotional difficulties, peer problems, conduct problems and hyperactivity subscale indicate more behavioural difficulties, whilst higher scores on the pro-social scale indicate more social and emotional competence. Reliability coefficients for externalizing type scales ranged from α=0.53 to 0.87, internalizing type scales α=0.48 to 0.78, pro-social scales α=0.66 to 0.88 and total externalizing and internalizing scales α=0.74 to 0.86.

### 4.7.4 Covariates (ages 0-1, 2, 4, 6 and 8)

A range of covariates, with documented evidence of association with behavioural development (e.g., see Bayer et al. 2011; Haltigan et al. 2017; Kuckertz et al. 2017; Masten et al. 2005) were included in the analysis. These included child gender (613 male, 676 female) and age. Children were coded Indigenous if at any point (ages 0-1 to 10) the mother identified them as Aboriginal or Torres Strait Islander (Indigenous n=21). Maternal education at child age 4 was used throughout the analysis and coded to three categories: 1. year 11 and certificate or trade qualification (n=249), 2. year 12 and certificate or trade qualification or diploma (n=487) and 3. bachelor degree or higher (n=553). Maternal depression was assessed with the Kessler K6 screening scale (Kessler et al. 2003). Six items asked how the parent felt the last four weeks, e.g., [*how often*] *did you feel worthless*, ranging from *all of the time*=1 to *none of the time*=5. The scale was reverse coded so that higher scores indicate more psychological distress. Reliability coefficients at ages 4, 6 and 8 were $\alpha$=0.81, 0.82 and 0.82, respectively. Socioeconomic status was assessed with the Socio-Economic Indexes for Areas (SEIFA) advantage/disadvantage metric at 2006 for ages 4 and 6, and 2011 for age 8. This ranks areas in Australia according to relative socio-economic advantage and disadvantage based on information from the five-yearly national census (Australian Bureau of Statistics 2018b). The Home Learning Environment (HLE) consisted of the average of seven items that assessed the frequency of the child and parent participating in activities, e.g., *read to the child from a book*, rated from *not in the past week*=0 to *6-7 days*=3 (reliability coefficients at age 4, 6 and 8 were $\alpha$=0.70, 0.66 and 0.63, respectively; Melhuish et al. 2008). Stressful Life Events (SLE) was the sum of 16 *yes*=1, *no*=0 questions regarding if the parent had experienced a range of stressful life events in the past year e.g., *suffered a serious illness, injury or assault* or *separation due to relationship or marital difficulties*. A range of dummy variables were also created if at any time in the available data (ages 0-1 to 10, waves 1 to 6) conditions were observed. An any present (1) or all absent (0) index of physical disabilities was created based on whether mothers reported the child has limited use of arms or fingers, has difficulty gripping things, has limited use of legs or feet, has other physical condition or has other disfigurement (n1=37). Likewise, an any present (1) or all absent (0) index of sight, speech and hearing problems was created based on whether mothers reported the child has sight problems, has hearing problems, has speech problems or has ongoing problems with eyes or seeing properly or hearing problems (n1=269). A dummy variable was also created for children with other non-specific medical condition or disability or other condition (1 = present, 0 = absent; n1=155). Finally, children with either difficulty learning or understanding things or developmental delay (only age 0–1) were coded 1 = developmental delay present (n1=31) or 0 = developmental delay absent.

### 4.7.5 Variable transformations

All time-varying independent, predictor variables—excluding behavioural outcomes and SEIFA—were standardised to Percentage of Maximum Possible (POMP). This means each variable ranges from 0 to 100, standard coefficients are easier to interpret (coefficient × 100 = maximum effect), interactions can be evaluated correctly and variables can be standardised over time (Cohen et al. 1999).

## 4.8 Analytical plan

A multiverse approach to the analysis was undertaken (Steegen et al. 2016). The multiverse analytical plan recognises that variables can be operationalised in multiple ways and therefore to avoid capitalising on chance, and to promote transparency, a large number of possibilities should be explored. Specifically, by rotating through the numerous ways variables might be analysed the sensitivity of the results to research choices can be visualised. If consistent trends are identified confidence in the results is raised (Dick et al. 2015). The approach to the multiverse is detailed below and was important for the current study because there are 8 outcome measures from 2 reporters modelled at 4 time specifications in 2 environments moderated by 4 reactivity measures, equalling 512 interactions.

First, using the sample with complete information, the behavioural outcomes in all scale forms (e.g., conduct problems, hyperactivity and externalizing) rated by the mother and teachers were predicted by interactions between parenting (warmth and harsh parenting) and temperament (difficult, reactivity, non-persistence and introversion) at ages 4 and 6 separately using general linear models, controlling for covariates. A simplified general linear model (Aiken et al. 1991) takes the form;

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i \qquad \text{Equation 4.1}$$

where *y* is the behavioural outcome, *i* indexes the child, *x* is the parenting measure, *z* is the measure of temperament, $\beta$ indicates the regression coefficients, and $\varepsilon$ is the residual errors that are normally distributed and independent with mean 0 and standard deviation $\sigma_\varepsilon$, specifically, $\varepsilon_i \sim$ Normal(0, $\sigma_\varepsilon^2$).The interaction coefficient is $\beta_3$.

Second, using the sample with complete information, the same interactions were examined across ages 4, 6 and 8 using hybrid random effect linear models, controlling for covariates. The hybrid model (Equation 4.2) is a cross between a random effects model and a fixed effects model and includes an average and change term (difference between observed at age and average across age) for covariates that vary with time (e.g., parenting) (Allison 2009; Schunck 2013). The average term captures between-person variation whilst the change term captures the within-person variation (within person estimates being equivalent to those economists typically estimate in a *fixed effect model*). The random effects component assigned a random intercept to each individual to capture differences in starting points between individuals. The between person effects are equivalent to the intercepts-as-outcomes model advocated by Roisman et al. (2012). Interactions in the hybrid model (Equation 4.3) were between temperament at age 4 and parenting at age 4, 6, and 8. The interaction covariate was appropriately generated before estimating effects (as recommended in Schunck 2013). The hybrid model was chosen because it meets the analytical requirements to estimate the between-person effect (average over time) and within-person effect (effect of changes across time), model relevant time invariant and time varying covariates, and model growth to describe individual behavioural trajectories if desired.

$$y_{it} = \beta_0 + \beta_1 \left(x_{it} - \bar{x}_i\right) + \beta_2 c_i + \beta_3 \bar{x}_i + \mu_i + \epsilon_{it} \qquad \text{Equation 4.2}$$

In the hybrid random effect model equation (4.2) above (Schunck 2013), $y$ is the behavioural outcome for individual $i$ at time $t$. $\beta$ indicates regression coefficients, $x$ is the score on a time varying covariate (e.g., parenting), $\overline{x}$ is the average score for a time varying covariate (e.g., parenting), $c$ is a time invariant covariate (e.g., child gender), $\mu$ is the random intercept with standard deviation $\sigma_\mu$, $\mu_i \sim$ Normal(0, $\sigma_\mu^2$), and $\varepsilon$ is the residual error ($\varepsilon_{it} \sim$ Normal(0, $\sigma_\varepsilon^2$)). $\beta_1$ is a within-person effect, whilst $\beta_3$ is a between-person effect.

The equation (4.2) below illustrates a reduced hybrid model with interactions as estimated in this study. Specifically, there is an interaction between time invariant temperament (age 4) and time varying parenting (age 4, 6, and 8);

$$y_{it} = \beta_0 + \beta_1 \left(x_{it} - \bar{x}_i\right) + \beta_2 \bar{x}_i + \beta_3 z_i + \beta_4 \bar{x}_i z_i + \beta_5 \left(x_{it} - \bar{x}_i\right) z_i + \mu_i + \epsilon_{it} \qquad \text{Equation 4.3}$$

where $x$ is the parenting score, $z$ is the temperament score, $\beta_4$ is the between-person interaction, and $\beta_5$ is the within-person interaction.

Interactions were evaluated for evidence of differential susceptibility, diathesis stress and vantage sensitivity using several criteria proposed by Roisman et al. (2012). First, the level of statistical significance was based on Benjamini and Hochberg (1995) procedure (critical value = 0.05) using only the interaction coefficients (512) given the number of coefficients calculated in the multiverse analysis. Second, the Regions of Significance (ROS; Dearing and Hamilton 2006) for temperament given parenting, and parenting given temperament were calculated. Temperament must be a significant predictor at low and high, but not medium, values of parenting for differential susceptibility to occur, whilst diathesis stress and vantage sensitivity generally occur when temperament is significant at low, high or all values of parenting. Parenting should be significant at high or all values of temperament for differential susceptibility, diathesis stress or vantage sensitivity to occur. When parenting is significant at low and high values of temperament it can suggest an alternative model of contrastive effects. These contrastive effects either operate in two-way (cross-over) or one-way (no cross-over) fashion that depend on the temperament coefficient given parenting. Low and high refers to the mean minus 2 Standard Deviations (SD) or plus 2SD, respectively. Third, the cross-over point where the lines of high and low temperament or parenting cross was calculated. The lines must cross-over within the range of observed parenting values for differential susceptibility, whilst cross-over points within the range of observed temperament suggest contrastive effects. Fourth, the proportion of children with parenting or temperament scores above the cross-over points was calculated (proportion affected; PA). Generally, 0.16 of children above (positive parenting) or below (negative parenting) the cross-over point on parenting provides confidence that differential susceptibility is observed, and 0.16 or more children above (reactive) or below (non-reactive) the cross-over point on temperament provides confidence in two-way contrastive effects. Finally, if differential susceptibility was observed, quadratic parenting and quadratic parenting by temperament interactions were included to rule out non-linear functions of parenting.

For each analysis, diagnostic plots of the residuals were examined to evaluate the normality of the residuals and validate assumptions of the statistical models (Pinheiro & Bates 2000, pg. 174). Analysis was undertaken using R (version 3.5.0; R Core Team 2018), with hybrid models estimated using the lme4 (version 1.1.17; Bates et al. 2015) and nlme (version 3.1.137; Pinheiro et al. 2018; to extract p-values) packages, ROS calculated using the interplot package (version 0.1.5; Solt & Hu 2016), and conditional interaction plots created using the visreg (version 2.5.0; Breheny & Burchett 2017) and ggplot2 (version 3.0.0; Wickham 2016) packages.

### 4.9 Results

Descriptive statistics for a main sample of variables used in the analysis are presented in Table 4.1. In terms of key correlations; **parenting;** parental warmth and harsh parenting were significantly ($p$<0.05) negatively correlated at ages 4, 6 and 8 ($r$=-0.35 to -0.4). **Temperament–parenting**; difficult temperament at age 4 and 6 had significant correlations with harsh parenting ($r$=0.18 to 0.35) and parental warmth ($r$=-0.1 to -0.21) that were weaker when not measured in the same time period. **Behaviour–parenting**; at age 4, 6 and 8, parent rated externalizing and internalizing behaviours were significantly negatively correlated with parental warmth ($r$=-0.12 to -0.16, $r$=-0.07 to -0.11, respectively) and positively correlated with harsh parenting ($r$=0.45 to 0.49, $r$=0.19 to 0.25, respectively), whilst parent rated pro-social behaviours were significantly positively correlated with parental warmth (*r*=0.25 to 0.26) and negatively correlated with harsh parenting (*r*=-0.25 to -0.36). **Teacher rated behaviour–parenting**; when significant, similar but weaker correlations were observed between parental warmth and harsh parenting and teacher rated externalizing behaviours (*r*=-0.06, *r*=0.13 to 0.26, respectively) and pro-social behaviours (*r*=0.06 to 0.12, *r*=-0.12 to -0.23, respectively). However, teacher rated internalizing behaviours had no significant correlations with parental warmth or harsh parenting. **Temperament–behaviour**; difficult temperament at age 4 and 6 was significantly correlated with more parent rated externalizing behaviours (*r*=0.2 to 0.38) and internalizing behaviours (*r*=0.19 to 0.33) and fewer pro-social behaviours (*r*=-0.19 to -0.32). **Temperament–teacher rated behaviour;** similar but weaker significant correlations were observed between difficult temperament at age 4 and 6 and teacher rated pro-social behaviour (*r*=-0.06 to -0.11), whilst very inconsistent and smaller correlations were observed with teacher rated externalizing (*r*=0.06 to 0.09) and internalizing behaviours (*r*=0.08 to 0.09). **Time reliability**; significant correlations increasing over time at ages 4, 6 and 8 were observed between teacher and parent reports of externalizing behaviour (*r*=0.25, 0.44, 0.5, respectively), internalizing behaviour (*r*=0.22, 0.29, 0.39, respectively) and pro-social behaviour (*r*=0.22, 0.25, 0.29, respectively). **Scale reliability**; regarding scale reliability, parental warmth (α=0.88 to 0.89) had better reliability than harsh parenting (α=0.65 to 0.68). Teacher ratings had higher reliability than parents, respectively, for externalizing (α=0.85 to 0.87 vs. 0.75 to 0.77), internalizing (α=0.72 to 0.78 vs. 0.6 to 0.67) and pro-social behaviours (α=0.81 vs. 0.6 to 0.67). **Missing data;** finally, looking at a comparison between children with partial data excluded from the study and children with complete information, the excluded children had lower parental warmth (max difference between excluded and included children = -0.03 pooled SD), higher harsh parenting (max = 0.04 pooled SD), more difficult temperaments (max = 0.07 pooled SD), more parent and teacher rated externalizing (max = 0.17 pooled SD) and internalizing (max = 0.14 pooled SD) behaviours and fewer parent and teacher rated pro-social behaviours (max = -0.18 pooled SD).

**Table 4.1**: Correlations, means, standard deviations, reliability coefficients and effect sizes of difference (Hedge 1981) to excluded sample for key variables, for different ages, used in the analysis. Summary information, except correlations, for all variables used in the analysis is provided in Appendix A. Underlined numbers are significant at *p*<0.05. T = teacher report, all others are parent report.

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Parental warmth 4 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 Parental warmth 6 | 0.62 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 Parental warmth 8 | 0.57 | 0.66 | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 Harsh parenting 4 | -0.35 | -0.31 | -0.27 | | | | | | | | | | | | | | | | | | | | | | | |
| 5 Harsh parenting 6 | -0.31 | -0.4 | -0.31 | 0.6 | | | | | | | | | | | | | | | | | | | | | | |
| 6 Harsh parenting 8 | -0.28 | -0.35 | -0.39 | 0.55 | 0.65 | | | | | | | | | | | | | | | | | | | | | |
| 7 Difficult at 4 | -0.18 | -0.11 | -0.1 | 0.27 | 0.21 | 0.18 | | | | | | | | | | | | | | | | | | | | |
| 8 Difficult 6 | -0.17 | -0.21 | -0.16 | 0.28 | 0.35 | 0.28 | 0.63 | | | | | | | | | | | | | | | | | | | |
| 9 Externalizing 4 | -0.12 | -0.12 | -0.08 | 0.45 | 0.42 | 0.34 | 0.38 | 0.29 | | | | | | | | | | | | | | | | | | |
| 10 Externalizing 6 | -0.08 | -0.16 | -0.08 | 0.34 | 0.49 | 0.39 | 0.25 | 0.36 | 0.63 | | | | | | | | | | | | | | | | | |
| 11 Externalizing 8 | -0.09 | -0.12 | -0.16 | 0.32 | 0.41 | 0.47 | 0.2 | 0.31 | 0.54 | 0.71 | | | | | | | | | | | | | | | | |
| 12 Internalizing 4 | -0.09 | -0.05 | -0.05 | 0.19 | 0.15 | 0.12 | 0.32 | 0.29 | 0.25 | 0.17 | 0.15 | | | | | | | | | | | | | | | |
| 13 Internalizing 6 | -0.09 | -0.11 | -0.08 | 0.2 | 0.2 | 0.17 | 0.24 | 0.33 | 0.24 | 0.29 | 0.23 | 0.56 | | | | | | | | | | | | | | |
| 14 Internalizing 8 | -0.04 | -0.06 | -0.07 | 0.17 | 0.18 | 0.25 | 0.19 | 0.22 | 0.21 | 0.24 | 0.32 | 0.42 | 0.57 | | | | | | | | | | | | | |
| 15 Pro-Social 4 | 0.26 | 0.21 | 0.18 | -0.25 | -0.26 | -0.21 | -0.32 | -0.28 | -0.35 | -0.27 | -0.22 | -0.13 | -0.08 | -0.04 | | | | | | | | | | | | |
| 16 Pro-Social 6 | 0.22 | 0.25 | 0.19 | -0.24 | -0.34 | -0.25 | -0.19 | -0.31 | -0.26 | -0.32 | -0.25 | -0.15 | -0.11 | -0.07 | 0.47 | | | | | | | | | | | |
| 17 Pro-Social 8 | 0.21 | 0.21 | 0.25 | -0.24 | -0.27 | -0.36 | -0.21 | -0.28 | -0.23 | -0.25 | -0.33 | -0.17 | -0.15 | -0.17 | 0.45 | 0.53 | | | | | | | | | | |
| 18 T Externalizing 4 | -0.03 | -0.04 | -0.01 | 0.13 | 0.15 | 0.16 | 0.04 | 0.03 | 0.25 | 0.3 | 0.28 | 0.04 | 0.1 | 0.1 | -0.16 | -0.11 | -0.13 | | | | | | | | | |
| 19 T Externalizing 6 | -0.02 | -0.06 | -0.05 | 0.13 | 0.21 | 0.19 | 0.02 | 0.06 | 0.27 | 0.44 | 0.45 | 0.03 | 0.07 | 0.09 | -0.14 | -0.21 | -0.23 | 0.45 | | | | | | | | |
| 20 T Externalizing 8 | 0 | -0.05 | -0.06 | 0.16 | 0.23 | 0.26 | 0.04 | 0.09 | 0.3 | 0.43 | 0.5 | 0.03 | 0.09 | 0.16 | -0.15 | -0.19 | -0.26 | 0.42 | 0.63 | | | | | | | |
| 21 T Internalizing 4 | -0.05 | -0.03 | -0.02 | -0.02 | -0.05 | -0.02 | 0.09 | 0.08 | 0.01 | 0.05 | 0.06 | 0.22 | 0.16 | 0.13 | -0.05 | -0.05 | -0.07 | 0.33 | 0.09 | 0.06 | | | | | | |
| 22 T Internalizing 6 | -0.01 | -0.01 | -0.02 | 0.01 | 0 | 0.05 | 0 | 0.04 | 0.04 | 0.11 | 0.11 | 0.13 | 0.29 | 0.21 | 0 | -0.03 | -0.04 | 0.16 | 0.33 | 0.16 | 0.2 | | | | | |
| 23 T Internalizing 8 | 0.03 | 0 | 0.02 | 0.01 | 0.04 | 0.04 | 0.03 | 0.02 | 0.08 | 0.15 | 0.16 | 0.1 | 0.24 | 0.39 | 0 | -0.05 | -0.06 | 0.22 | 0.25 | 0.35 | 0.2 | 0.3 | | | | |
| 24 T Pro-Social 4 | 0.06 | 0.05 | 0.04 | -0.12 | -0.12 | -0.15 | -0.09 | -0.07 | -0.17 | -0.19 | -0.17 | -0.09 | -0.04 | -0.04 | 0.22 | 0.2 | 0.19 | -0.57 | -0.29 | -0.27 | -0.31 | -0.08 | -0.12 | | | |
| 25 T Pro-Social 6 | 0.07 | 0.12 | 0.1 | -0.15 | -0.19 | -0.2 | -0.06 | -0.11 | -0.19 | -0.29 | -0.3 | -0.07 | -0.07 | -0.09 | 0.17 | 0.25 | 0.3 | -0.31 | -0.59 | -0.4 | -0.1 | -0.24 | -0.14 | 0.28 | | |
| 26 T Pro-Social 8 | 0.02 | 0.09 | 0.07 | -0.16 | -0.19 | -0.23 | -0.07 | -0.11 | -0.21 | -0.29 | -0.31 | -0.06 | -0.07 | -0.13 | 0.13 | 0.22 | 0.29 | -0.28 | -0.43 | -0.61 | -0.09 | -0.11 | -0.26 | 0.28 | 0.39 | |
| Mean | 87.53 | 88.61 | 86.02 | 22.98 | 22.88 | 23.35 | 38.09 | 33.47 | 4.92 | 4.37 | 4.08 | 4.41 | 4.58 | 4.66 | 7.85 | 8.58 | 8.69 | 2.8 | 3.09 | 3.01 | 4.02 | 4.01 | 4.19 | 7.6 | 8 | 8.01 |
| Standard deviation | 11.5 | 12.12 | 13.18 | 12.04 | 12.36 | 12.62 | 11.72 | 11.81 | 3.06 | 2.99 | 3.07 | 2.12 | 2.31 | 2.46 | 1.67 | 1.57 | 1.53 | 3.33 | 3.42 | 3.58 | 2.43 | 2.43 | 2.78 | 2.11 | 2.07 | 2.08 |
| Reliability (α) | 0.88 | 0.89 | 0.89 | 0.65 | 0.67 | 0.68 | 0.71 | 0.71 | 0.75 | 0.76 | 0.77 | 0.6 | 0.65 | 0.67 | 0.66 | 0.67 | 0.67 | 0.86 | 0.85 | 0.87 | 0.72 | 0.73 | 0.78 | 0.81 | 0.81 | 0.81 |
| Effect size *Hedges' G* | -0.01 | -0.03 | -0.01 | 0.04 | 0.02 | 0.03 | 0.07 | 0.04 | 0.09 | 0.11 | 0.12 | 0.14 | 0.08 | 0.11 | -0.04 | -0.1 | -0.1 | 0.17 | 0.14 | 0.14 | 0.12 | 0.13 | 0.05 | -0.14 | -0.18 | -0.08 |
| | Warmth | | | Harsh parenting | | | Difficult | | P Externalizing | | | P Internalizing | | | P Pro-social | | | T Externalizing | | | T Internalizing | | | T Pro-social | | |

### 4.9.1 Theory testing and measurement: Interaction models

The *first exploratory hypothesis* of this study was that if temperament moderated the effect of parenting on behavioural development it could reasonably be expected to take the shape of either a diathesis-stress, differential susceptibility, or vantage sensitivity model.

After adjustment for multiple comparison only thirteen of the 512 interactions remained as robust evidence for the moderating effect of temperament. These significant interactions and associated criteria for assessing the theoretical models are presented in Table 4.2.

These interactions partially support the first hypotheses and are also presented in figures and illustrate eight cases of two-way contrastive effects, two cases of one-way contrastive effects and three cases of diathesis stress. One-way contrastive effects mean the effects of parenting are the opposite for reactive and non-reactive children and these children are significantly different along the whole, or at one end of, the parenting distribution. On the other hand, two-way contrastive effects mean the effects of parenting are opposite for reactive and non-reactive children and these children are significantly different at both ends of the parenting distribution, but not-significantly different at some point where there predicted behaviours become equivalent and cross-over. Diathesis stress effects are significant differences between reactive and non-reactive children at one end of the, or for the entire, range of parenting and the effect of parenting is stronger for reactive children and non-significant or weaker (in the same direction) for non-reactive children.

Table 4.2 confirms this interpretation of the interactions. The interactions consistent with two-way contrastive effects have lower and upper bounds of the regions of significance for both temperament on parenting and parenting on temperament within two standard deviations of the mean of parenting or temperament, respectively (parenting warmth, between term M=87.4, SD=10.6, change term M>0.01, SD=6.3; harsh parenting, between term M=23.1, SD=10.6, change term M>0.01, SD=10.6; temperament available in Appendix A). Further, they had PA (proportion affected; see Section 4.9) indexes for parenting as the *x-axis* variable ranging from 0.45 to 0.78 and PA indexes for temperament as the *x-axis* variable ranging from 0.19 to 0.53. On the other hand, the one-way contrastive effects had only the lower or upper bounds of the regions of significance for temperament on parenting which were within two standard deviations of the mean of parenting, whilst lower and upper bounds of the regions of significance for parenting on temperament were within two standard deviations of the mean. The PA indices for parenting as the *x-axis* variable were 0.19 and 1 and PA indexes for temperament as the *x-axis* variable were 0.26 and 0.35, respectively. Regarding the diathesis stress interactions, they had regions of significance within two standard deviations of the mean in only the upper bound for temperament on parenting, whilst parenting was only significant in context of higher temperament scores. The PA indices for parenting as the *x-axis* variable ranged from <0.01 to 0.16 and PA indexes for temperament as the *x-axis* variable ranged from 0.67 to 1, respectively. Furthermore, non-linear parenting effects and non-linear *parenting* × *temperament* interactions were estimated. None of these interactions were statistically significant after adjusting for multiple comparisons and did not change interpretation of any linear interaction effects.

These interactions are presented in terms of outcome domains below. Notable trends, observable below and in Table 4.2, are that the two-way contrastive effects were restricted to teacher reports of behaviour, whilst diathesis stress interactions were only observed for parent reports of behaviour.

### 4.9.2 Predictors of externalizing behaviour

Changes in harsh parenting, the more reliable, within person term, predicted significantly different parent rated conduct problems for reactive and non-reactive children consistent with diathesis stress (Figure 4.2a). Additionally, harsh parenting and parental warmth at age 4 interacted with difficult temperament to predict parent and teacher rated conduct problems, respectively. The harsh parenting interaction was consistent with diathesis stress (Figure 4.2b), whilst the parental warmth interaction was consistent with two-way contrastive effects (Figure 4.2c).

### 4.9.3 Predictors of internalizing behaviour

Changes in parental warmth, the more reliable, within person term, interacted with introversion to predict parent rated overall internalizing behaviours consistent with one-way contrastive effects (Figure 4.3a), but in a direction opposite to that presupposed. That is, children with high introversion had an unexpected increase in internalizing behaviours when parental warmth increased.

Average parental warmth interacted with difficult temperament to predict teacher rated emotional difficulties consistent with one-way contrastive effects (Figure 4.3b), and average parental warmth also interacted with difficult temperament and reactivity to predict teacher rated peer problems (Figure 4.4a,b) and overall internalizing behaviour (Figure 4.4c,d) consistent with two-way contrastive effects.

Average harsh parenting interacted with reactivity to predict parent rated overall internalizing behaviour consistent with diathesis stress (Figure 4.5a) and teacher rated peer problems (Figure 4.5b) and overall internalizing behaviour (Figure 4.5c) consistent with two-way contrastive effects.

### 4.9.4 Predictors of internalizing and externalizing behaviour

Average parental warmth interacted with difficult temperament to predict overall externalizing and internalizing behaviour consistent with a two-way constative effect (Figure 4.6).

### 4.9.5 Predictors of pro-social behaviour

No interactions between parenting and temperament were observed to significantly predict pro-social behaviour after adjusting for multiple comparisons.

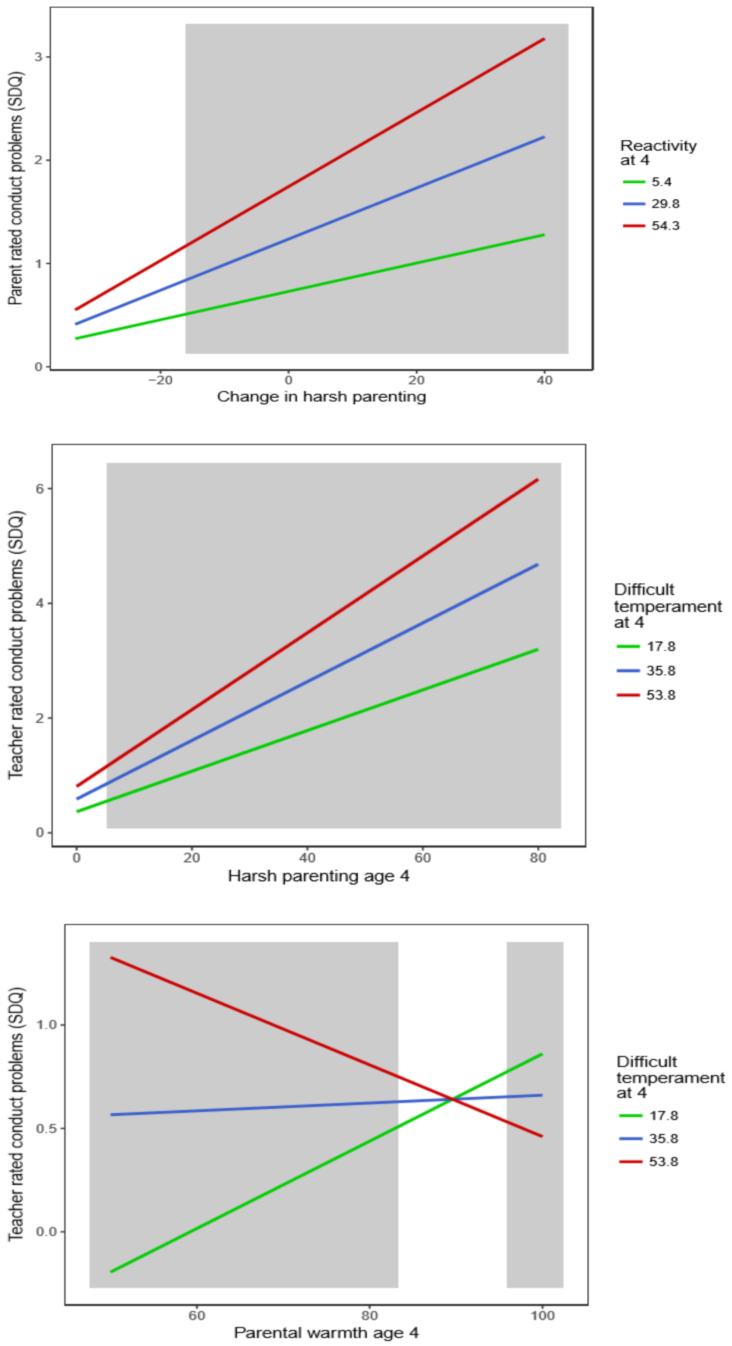



**Figure 4.2:** Conditional interaction plot with regions of significance (grey polygons) where harsh and warm parenting effects on conduct problems significantly differ due to temperament (red +1.5 SD, blue = average and green -1.5 SD). Model a) is consistent with diathesis stress, b) is consistent with diathesis stress, and c) is consistent with two-way contrastive effects. The cross-over point in a) is not identical to Table 4.2 because the plotted interaction uses age 6.

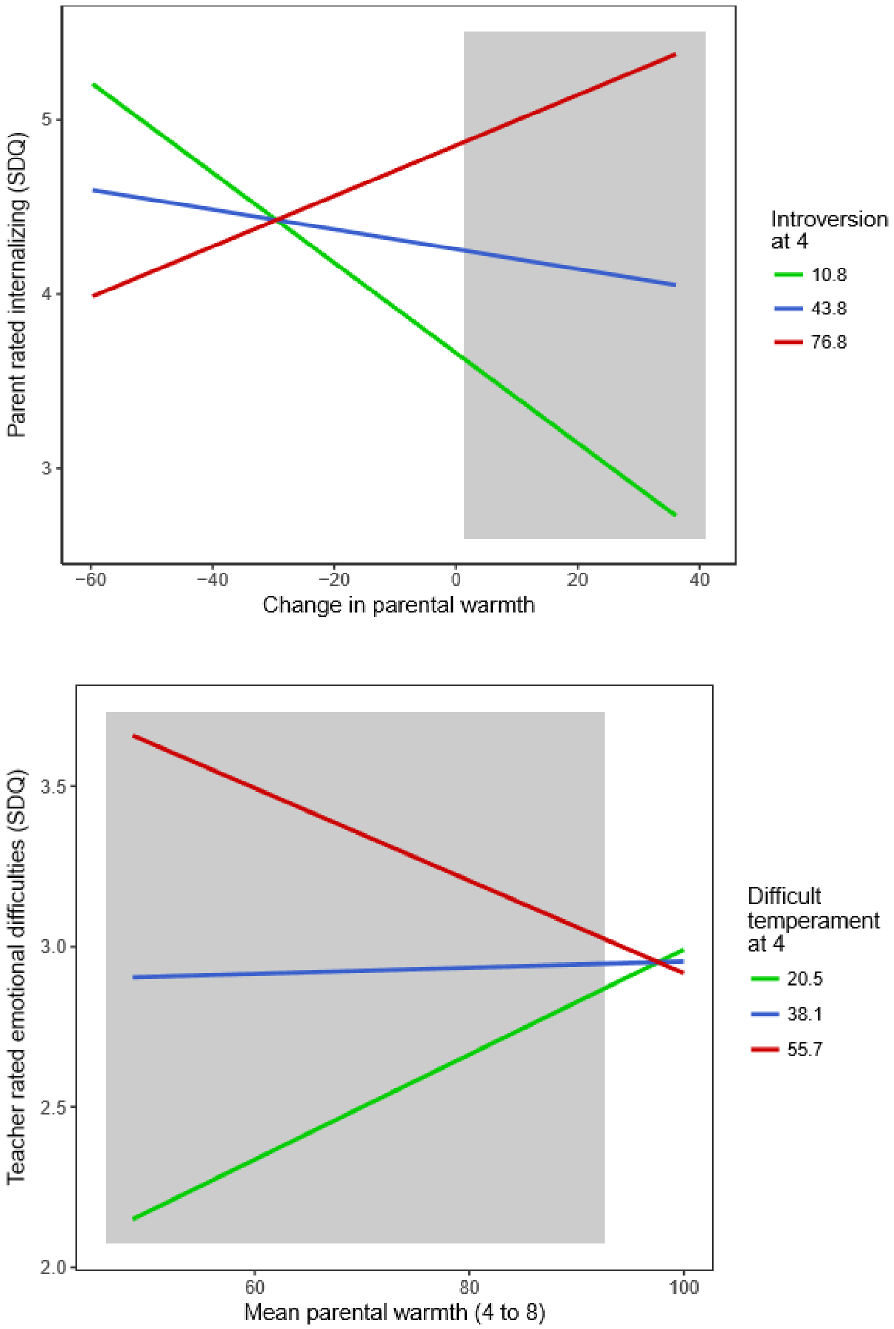



**Figure 4.3:** Conditional interaction plot with regions of significance (grey polygons) where parental warmth effects on teacher rated emotional difficulties and parent rated internalizing behaviours significantly differ due to temperament (red +1.5 SD, blue = average and green -1.5 SD). Model a) and b) are consistent with one-way contrastive effects. The cross-over point for a) is not identical to Table 4.2 because the plotted interaction uses age 6.

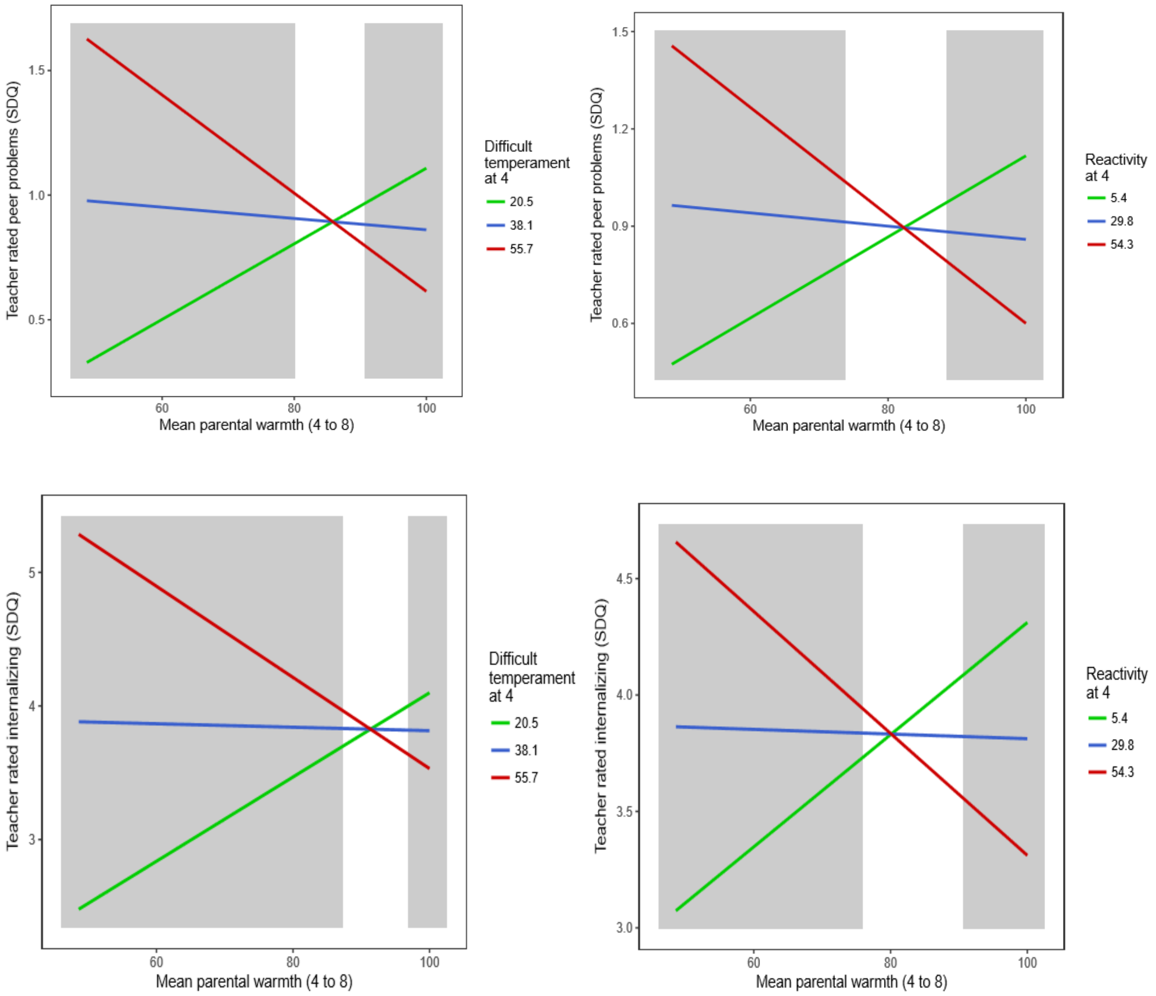



**Figure 4.4:** Conditional interaction plot with regions of significance (grey polygons) where parental warmth effects on teacher rated peer problems and internalizing behaviour significantly differ due to temperament (red +1.5 SD, blue = average and green -1.5 SD). Model a), b), c) and d) are consistent with two-way contrastive effects.

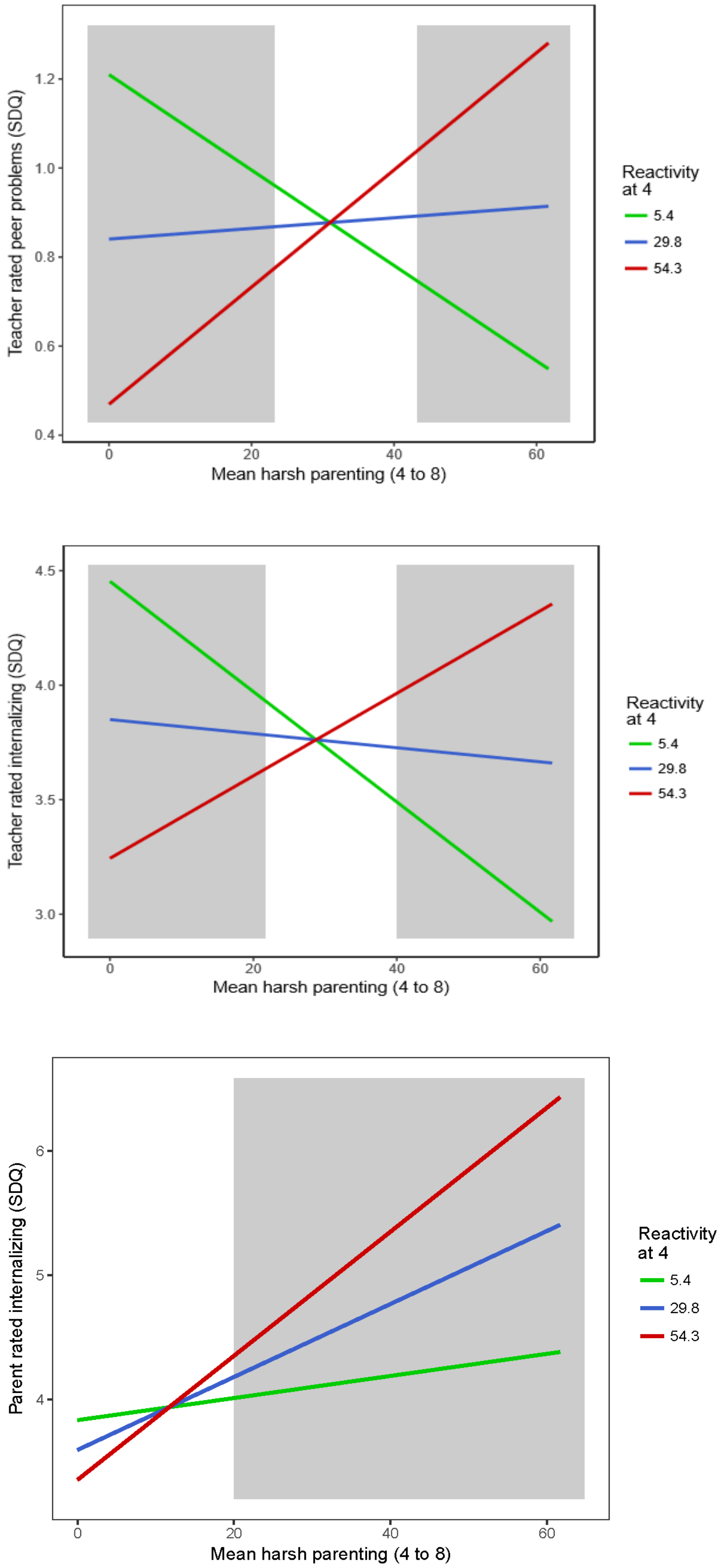



**Figure 4.5:** Conditional interaction plot with regions of significance (grey polygons) where harsh parenting effects on teacher rated internalizing and peer problem behaviour and parent rated internalizing behaviour significantly differ due to temperament (red +1.5 SD, blue = average and green -1.5 SD). Model a) and b) are consistent with two-way contrastive effects, whilst c) is consistent with diathesis stress.

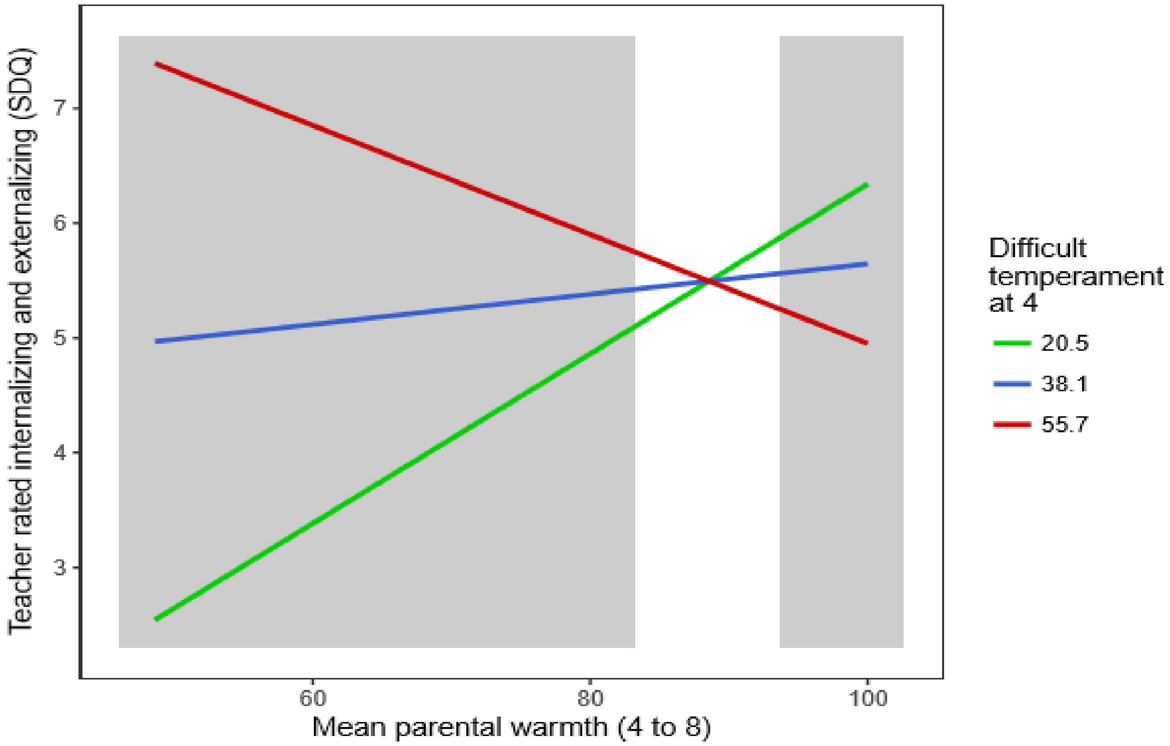



**Figure 4.6:** Conditional interaction plot with regions of significance (grey polygons) where parental warmth effects on parent rated internalizing and externalizing behaviour significantly differ due to temperament (red +1.5 SD, blue = average and green -1.5 SD). Model is consistent with two-way contrastive effects.

**Table 4.2**: Criteria to evaluate theoretical models. Columns 2–5 indicate regression coefficients for the intercept, parenting environment, temperament and interaction, respectively, p-value refers to interaction p-value. All interactions are significant after adjusting for multiple comparisons. T = teacher report, P = parent report, M = mean over time, C = change over time, 4 = measured at age 4, CO = crossover point, PA = proportion affected. Outcome codes; EmotD = emotional difficulties, INT = internalizing, PeerP = peer problems, TSDQ = internalizing and externalizing, CondP = conduct problems. Parenting codes; Warmth = parental warmth, Harsh P = harsh parenting. Temperament codes; Diff = difficult, React = reactivity, Intro = introversion. Model codes; CE1 = one-way contrastive effects, CE2 = two-way contrastive effects, DA = diathesis stress. * = $p<0.05$, ** <0.01, ***<0.001. < and > indicate significance occurred outside observed values. All ROS bounds within 2SD of mean, see results text and Appendix A.

| | | | | | | ROS temperament on parent | | | | ROS parent on temperament | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome | Intercept | M Warmth | Diff 4 | Inter | p-value | Lower | Upper | CO | PA > | Lower | Upper | CO | PA > | Model |
| T EmotD | 0.07 | 0.03 | 0.09 | -0.0009 | 0.000286 | 92.7 | > | 97.6 | 0.19 | 31.2 | 46.9 | 39.2 | 0.26 | CE1 |
| T INT | -1.35 | 0.07 | 0.17 | -0.0019 | 0.000004 | 87.4 | 96.9 | 91.4 | 0.45 | 31.2 | 41.7 | 37.4 | 0.29 | CE2 |
| T PeerP | -1.42 | 0.04 | 0.09 | -0.001 | 0.000023 | 80.1 | 90.6 | 85.8 | 0.61 | 27.8 | 41.7 | 35.8 | 0.33 | CE2 |
| T TSDQ | -3.57 | 0.14 | 0.31 | -0.0034 | 0.000038 | 83.2 | 93.7 | 88.5 | 0.54 | 36.5 | 48.7 | 41.9 | 0.19 | CE2 |
| Outcome | Intercept | Warmth | Diff | Inter | p-value | Lower | Upper | CO | PA > | Lower | Upper | CO | PA > | Model |
| T CondP 4 | -2.71 | 0.04 | 0.10 | -0.0011 | 0.000957 | 83.3 | 95.8 | 89.6 | 0.51 | 27.8 | 45.2 | 37.6 | 0.53 | CE2 |
| Outcome | Intercept | M Warmth | React 4 | Inter | p-value | Lower | Upper | CO | PA > | Lower | Upper | CO | PA > | Model |
| T PeerP | 0.27 | 0.02 | 0.05 | -0.0006 | 0.000221 | 73.8 | 88.5 | 82.6 | 0.69 | 15.8 | 31.6 | 26.4 | 0.52 | CE2 |
| T INT | 2.23 | 0.03 | 0.09 | -0.0010 | 0.000300 | 75.9 | 90.6 | 84.3 | 0.65 | 15.8 | 36.8 | 28.8 | 0.52 | CE2 |
| Outcome | Intercept | C Warmth | Intro 4 | Inter | p-value | Lower | Upper | CO | PA > | Lower | Upper | CO | PA > | Model |
| P INT | 2.15 | -0.03 | 0.03 | 0.0006 | 0.001134 | < | 1.2 | -56.8 | 1.00 | 40 | 80 | 53.1 | 0.35 | CE1 |
| Outcome | Intercept | Harsh P | Diff | Inter | p-value | Lower | Upper | CO | PA < | Lower | Upper | CO | PA > | Model |
| P CondP 4 | 1.42 | 0.02 | 0.01 | 0.0009 | 0.000870 | < | 5.3 | -14.0 | 0.00 | ALL | ALL | -22.6 | 1.00 | DA |
| Outcome | Intercept | M Harsh P | React 4 | Inter | p-value | Lower | Upper | CO | PA < | Lower | Upper | CO | PA > | Model |
| P INT | 3.58 | 0 | -0.01 | 0.0008 | 0.000305 | < | 20 | 11.7 | 0.16 | 15.8 | > | -5.2 | 0.67 | DA |
| T INT | 5.82 | -0.03 | -0.02 | 0.0009 | 0.000508 | 21.7 | 40 | 28.7 | 0.75 | 21.1 | 47.4 | 33.4 | 0.22 | CE2 |
| T PeerP | 2.38 | -0.01 | -0.02 | 0.0005 | 0.000676 | 23.3 | 43.3 | 31 | 0.78 | 10.5 | 36.8 | 27.4 | 0.3 | CE2 |
| Outcome | Intercept | C Harsh P | React 4 | Inter | p-value | Lower | Upper | CO | PA < | Lower | Upper | CO | PA > | Model |
| P CondP | 1.51 | 0.01 | 0.02 | 0.0005 | 0.000788 | < | -16.1 | -34.7 | 0.00 | 0 | > | -25 | 1 | DA |

### 4.9.6 Theory testing and consistency of findings: Multiverse analysis:

The *second direct hypothesis* of this study was that if temperament moderated the effect of parenting on an aspect of behaviour, the interaction would be consistent at all ages and across time, consistent across reporters of behaviour, parenting environments and similar behaviour measures and generally consistent across temperament measures.

Despite the demonstration in the data that temperament moderates the effects of parenting and is consistent with contrastive effects and diathesis stress, the multiverse analysis does not support the second hypothesis and demonstrates that evidence for interactions is sporadic and inconsistent (Table 4.3). Examination of results by data source, temperament, parenting, age, and outcome show few consistencies.

With regard to data source consistency, average harsh parenting and reactivity at age 4 predicted the internalizing behaviour for both mother and teacher reported behaviour, yet mother reports aligned with one-way contrastive effects and teacher reports indicated two-way contrastive effects.

Looking at temperament measures, consistencies were evident in that difficult temperament at age 4 and reactivity at age 4 moderated the effect of average parental warmth on teacher rated peer problems and internalizing behaviours consistent with two-way contrastive effects.

There were two consistencies observed across parenting environments as reactivity at age 4 moderated the effect of average harsh parenting and average parental warmth on teacher rated peer problems and internalizing behaviours consistent with two-way contrastive effects.

On the other hand, no significant interactions were consistently found at ages 4 and 6 and over time.

Changes in outcome scales evidenced more comparable results where average harsh parenting moderated by reactivity significantly predicted teacher rated peer problems and internalizing behaviours (consistent with two-way contrastive effects) and parent rated internalizing behaviours (consistent with diathesis stress). Further, average parental warmth moderated by difficult temperament significantly predicted teacher rated emotional difficulties (consistent with one-way contrastive effects) and teacher rated peer problems, internalizing behaviours and total internalizing and externalizing behaviours (consistent with two-way contrastive effects), whilst parental warmth moderated by difficult temperament significantly predicted teacher rated conduct problems at age 4 (consistent with two-way contrastive effects). In addition, average parental warmth moderated by reactivity significantly predicted teacher rated peer problems and internalizing behaviours (consistent with two-way contrastive effects).

In sum, while some consistencies were seen among outcomes, parenting environments, and measures of temperament, the results seem substantially influenced by analytical choices in the construction of variables used in the interactions. Further illustrating inconsistencies, relaxing statistical criteria for significance to $p<0.01$ would see 7% of interactions (38) as significant and relaxing it further to $p<0.05$ would see 18% of interactions (94) as significant. However, many gaps and inconsistencies would remain even with such criteria (Table 4.3).

**Table 4.3**: *P*-values for assessed interactions. Bold, underlined interactions significant after adjusting for multiple comparisons. Light grey shading $p<0.05$, dark grey shading $p<0.01$. Codes; P=parent rated, T=teacher rated, CondP=conduct problems, Hyper=hyperactivity, Extern=Externalizing, EmD= emotional difficulties, PeerP=peer problems, Intern=Internalizing, Diff=difficult, React = reactivity, Intro = introversion, Non-pers = non-persistence, M=average (between-person), C=change (within-person).

| | | Harsh parenting | | | | Parental warmth | | | |
|---|---|---|---|---|---|---|---|---|---|
| Outcome | Age | Diff | Intro | React | Non-pers | Diff | Intro | React | Non-pers |
| P CondP | 4 | **0.0009** | 0.0115 | 0.3405 | 0.2771 | 0.6547 | 0.4868 | 0.8762 | 0.4492 |
| P CondP | 6 | 0.0021 | 0.9565 | 0.0981 | 0.0944 | 0.6141 | 0.7386 | 0.5467 | 0.2915 |
| P CondP | M | 0.0034 | 0.5988 | 0.0547 | 0.1196 | 0.2105 | 0.235 | 0.24 | 0.9804 |
| P CondP | C | 0.0089 | 0.4847 | **0.0008** | 0.2529 | 0.1688 | 0.8061 | 0.4399 | 0.0702 |
| T CondP | 4 | 0.1419 | 0.9258 | 0.1003 | 0.4957 | **0.001** | 0.2453 | 0.0059 | 0.0062 |
| T CondP | 6 | 0.5082 | 0.0952 | 0.9873 | 0.0356 | 0.1315 | 0.7709 | 0.0274 | 0.7471 |
| T CondP | M | 0.2207 | 0.2582 | 0.0565 | 0.1501 | 0.002 | 0.2823 | 0.0034 | 0.0258 |
| T CondP | C | 0.0449 | 0.2295 | 0.1343 | 0.2466 | 0.1729 | 0.4218 | 0.4449 | 0.2742 |
| P Hyper | 4 | 0.6747 | 0.4821 | 0.8682 | 0.7892 | 0.7523 | 0.775 | 0.391 | 0.4344 |
| P Hyper | 6 | 0.9437 | 0.1838 | 0.633 | 0.6683 | 0.9446 | 0.4894 | 0.4775 | 0.498 |
| P Hyper | M | 0.8031 | 0.9869 | 0.6573 | 0.8534 | 0.0337 | 0.0581 | 0.0253 | 0.6825 |
| P Hyper | C | 0.1038 | 0.0059 | 0.7007 | 0.6311 | 0.8594 | 0.7423 | 0.8789 | 0.3571 |
| T Hyper | 4 | 0.3638 | 0.4139 | 0.0491 | 0.1559 | 0.1978 | 0.1057 | 0.7749 | 0.684 |
| T Hyper | 6 | 0.6134 | 0.0129 | 0.6792 | 0.2149 | 0.3957 | 0.8487 | 0.8266 | 0.8815 |
| T Hyper | M | 0.7627 | 0.3769 | 0.8178 | 0.8683 | 0.04 | 0.0564 | 0.3075 | 0.2967 |
| T Hyper | C | 0.0297 | 0.0732 | 0.0184 | 0.8739 | 0.2854 | 0.3981 | 0.3369 | 0.8052 |
| P Extern | 4 | 0.0297 | 0.0505 | 0.6561 | 0.4136 | 0.6385 | 0.8312 | 0.5007 | 0.9303 |
| P Extern | 6 | 0.1748 | 0.3302 | 0.2504 | 0.6339 | 0.8581 | 0.7185 | 0.8054 | 0.3108 |
| P Extern | M | 0.2659 | 0.8212 | 0.5911 | 0.5713 | 0.0305 | 0.0502 | 0.0271 | 0.7676 |
| P Extern | C | 0.774 | 0.1067 | 0.1074 | 0.3198 | 0.3655 | 0.9191 | 0.7458 | 0.0905 |
| T Extern | 4 | 0.9208 | 0.5608 | 0.6144 | 0.545 | 0.0156 | 0.105 | 0.1314 | 0.1141 |
| T Extern | 6 | 0.9077 | 0.0126 | 0.7552 | 0.0811 | 0.2253 | 0.7983 | 0.3032 | 0.9865 |
| T Extern | M | 0.7801 | 0.2821 | 0.5394 | 0.4839 | 0.0075 | 0.0772 | 0.0575 | 0.1025 |
| T Extern | C | 0.0124 | 0.0631 | 0.0161 | 0.6876 | 0.1622 | 0.3259 | 0.2938 | 0.5019 |
| P EmD | 4 | 0.3002 | 0.9613 | 0.1134 | 0.3225 | 0.5959 | 0.6076 | 0.5647 | 0.2219 |
| P EmD | 6 | 0.0484 | 0.0672 | 0.0779 | 0.3141 | 0.414 | 0.0755 | 0.8247 | 0.6027 |
| P EmD | M | 0.0026 | 0.291 | 0.0015 | 0.0524 | 0.0582 | 0.1244 | 0.4663 | 0.2286 |
| P EmD | C | 0.5213 | 0.0654 | 0.5334 | 0.0942 | 0.3166 | 0.0142 | 0.8194 | 0.2275 |
| T EmD | 4 | 0.7032 | 0.2852 | 0.3528 | 0.2117 | 0.7256 | 0.3928 | 0.1623 | 0.5964 |
| T EmD | 6 | 0.2461 | 0.697 | 0.3443 | 0.3414 | 0.0567 | 0.0543 | 0.3401 | 0.4087 |
| T EmD | M | 0.0418 | 0.1315 | 0.0116 | 0.9675 | **0.0003** | 0.0119 | 0.0135 | 0.0632 |
| T EmD | C | 0.8519 | 0.9403 | 0.8432 | 0.4782 | 0.8855 | 0.1004 | 0.5266 | 0.0186 |
| P PeerP | 4 | 0.2766 | 0.1936 | 0.0308 | 0.0931 | 0.1864 | 0.605 | 0.1802 | 0.4674 |
| P PeerP | 6 | 0.4631 | 0.3963 | 0.4353 | 0.192 | 0.4314 | 0.1679 | 0.0991 | 0.2037 |
| P PeerP | M | 0.1427 | 0.7086 | 0.0094 | 0.7233 | 0.1258 | 0.8618 | 0.0032 | 0.6601 |
| P PeerP | C | 0.3096 | 0.1601 | 0.9622 | 0.7547 | 0.2033 | 0.0099 | 0.4851 | 0.948 |
| T PeerP | 4 | 0.4173 | 0.2318 | 0.0217 | 0.0447 | 0.0368 | 0.2315 | 0.1139 | 0.161 |
| T PeerP | 6 | 0.6425 | 0.6585 | 0.9452 | 0.2595 | 0.2402 | 0.1608 | 0.5632 | 0.8521 |
| T PeerP | M | 0.0074 | 0.4928 | **0.0007** | 0.1901 | **0.00002** | 0.0128 | **0.0002** | 0.0505 |
| T PeerP | C | 0.4673 | 0.3557 | 0.0658 | 0.4229 | 0.9333 | 0.388 | 0.3869 | 0.085 |
| P Intern | 4 | 0.1803 | 0.4302 | 0.0175 | 0.6698 | 0.2454 | 0.9984 | 0.2249 | 0.2153 |
| P Intern | 6 | 0.0759 | 0.4714 | 0.098 | 0.144 | 0.3058 | 0.7143 | 0.3998 | 0.2653 |
| P Intern | M | 0.0044 | 0.3603 | **0.0003** | 0.283 | 0.0322 | 0.3615 | 0.028 | 0.2933 |
| P Intern | C | 0.8498 | 0.6985 | 0.6935 | 0.1837 | 0.1433 | **0.0011** | 0.7838 | 0.4354 |
| T Intern | 4 | 0.4663 | 0.1724 | 0.0474 | 0.0475 | 0.129 | 0.7847 | 0.0726 | 0.2344 |
| T Intern | 6 | 0.3061 | 0.9873 | 0.5562 | 0.2123 | 0.0564 | 0.0411 | 0.343 | 0.6564 |
| T Intern | M | 0.0057 | 0.1951 | **0.0005** | 0.462 | **0.000004** | 0.0033 | **0.0003** | 0.025 |
| T Intern | C | 0.763 | 0.6266 | 0.234 | 0.9905 | 0.8886 | 0.1223 | 0.3662 | 0.0126 |

**Table 4.3 continued**: *P*-values for assessed interactions. Bold, underlined interactions significant after adjusting for multiple comparisons. Light grey shading $p<0.05$, dark grey shading $p<0.01$. Codes; P=parent rated, T=teacher rated, CondP=conduct problems, Hyper=hyperactivity, Extern=Externalizing, EmD= emotional difficulties, PeerP=peer problems, Intern=Internalizing, Diff=difficult, React = reactivity, Intro = introversion, Non-pers = non-persistence, M=average (between-person), C=change (within-person).

| | | Harsh parenting | | | | Parental warmth | | | |
|---|---|---|---|---|---|---|---|---|---|
| Outcome | Age | Diff | Intro | React | Non-pers | Diff | Intro | React | Non-pers |
| P TSDQ | 4 | 0.0142 | 0.0626 | 0.086 | 0.7391 | 0.2965 | 0.8799 | 0.2317 | 0.4282 |
| P TSDQ | 6 | 0.0418 | 0.8303 | 0.0726 | 0.222 | 0.4575 | 0.9855 | 0.7249 | 0.9991 |
| P TSDQ | M | 0.0132 | 0.4855 | 0.0122 | 0.3049 | 0.0047 | 0.0588 | 0.0047 | 0.4088 |
| P TSDQ | C | 0.9342 | 0.3781 | 0.1732 | 0.8966 | 0.7831 | 0.0375 | 0.693 | 0.0974 |
| T TSDQ | 4 | 0.6528 | 0.2668 | 0.4952 | 0.1479 | 0.0129 | 0.2076 | 0.0484 | 0.0876 |
| T TSDQ | 6 | 0.6523 | 0.0886 | 0.9233 | 0.0636 | 0.0666 | 0.2105 | 0.2275 | 0.8235 |
| T TSDQ | M | 0.1257 | 0.9062 | 0.0343 | 0.398 | <u>**0.00004**</u> | 0.0078 | 0.0021 | 0.0259 |
| T TSDQ | C | 0.0714 | 0.3603 | 0.0229 | 0.7886 | 0.4104 | 0.7746 | 0.888 | 0.0539 |
| P ProSo | 4 | 0.0711 | 0.0284 | 0.9901 | 0.781 | 0.8711 | 0.995 | 0.4027 | 0.7914 |
| P ProSo | 6 | 0.8561 | 0.0237 | 0.223 | 0.0171 | 0.873 | 0.0423 | 0.0258 | 0.8192 |
| P ProSo | M | 0.0257 | 0.004 | 0.5311 | 0.7279 | 0.7412 | 0.1875 | 0.6701 | 0.5929 |
| P ProSo | C | 0.4983 | 0.4662 | 0.7669 | 0.8453 | 0.2693 | 0.948 | 0.1446 | 0.4084 |
| T ProSo | 4 | 0.8877 | 0.4996 | 0.8404 | 0.1542 | 0.3711 | 0.5155 | 0.6073 | 0.1717 |
| T ProSo | 6 | 0.7043 | 0.2477 | 0.1871 | 0.5845 | 0.7638 | 0.5541 | 0.279 | 0.2219 |
| T ProSo | M | 0.9008 | 0.4608 | 0.0889 | 0.1324 | 0.1004 | 0.3819 | 0.1873 | 0.326 |
| T ProSo | C | 0.0035 | 0.0087 | 0.0195 | 0.6968 | 0.5986 | 0.8487 | 0.3122 | 0.9181 |

### 4.10 Discussion

This study investigated how child temperament moderated the effects of parenting emotional quality on behavioural development in a large longitudinal sample of Australian children surveyed at ages 4, 6 and 8. Temperament was indexed by mother reported reactivity, non-persistence, introversion, and a composite score defining difficult temperament. Parenting emotional quality was captured by mother reported parental warmth and harsh parenting. Behavioural outcomes included mother and teacher reports of the strength and difficulties questionnaire. A detailed analytical plan, the multiverse, was used to establish rigorous consistencies in evidence. The first hypothesis of the study was that if temperament moderated the effect of parenting on behavioural development it could take the shape of either a diathesis-stress, differential susceptibility, or vantage sensitivity model. A second hypothesis was that if temperament moderated the effect of parenting on an aspect of behaviour the interaction would be consistent at all ages and across time, consistent across reporters of behaviour, parenting environments, and similar behaviour measures, and generally consistent across temperament measures.

#### 4.10.1 Hypothesis 1: Interaction effects and support for the theoretical models

To test the hypothesis that if temperament moderated the effect of parenting on behavioural development it would take the shape of either a diathesis stress, differential susceptibility, or vantage sensitivity model, this study examined numerous statistical models that included interactions between parenting and temperament predicting behavioural outcomes.

Detailed exploration of these interactions with discipline relevant statistical criteria showed that there was very limited support for the diathesis stress model (3 interactions) and no support for the differential susceptibility or vantage sensitivity models. Moreover, the study found evidence for one-way (2 interactions) and two-way (8 interactions) contrastive models. Thus, the first hypothesis was weakly supported.

Regarding the three diathesis stress interactions, more reactive children exhibited more behaviour problems, compared to less reactive children, in context of developmentally poor parenting. Similar diathesis stress effects have been observed in previous research (Gallitto 2015; Kiff et al. 2011; Morris et al. 2002; Roisman et al. 2012; Slagt et al. 2016b; Stoltz et al. 2017) and may be due to children with a reactive disposition to negative outcomes or more difficult temperament responding to harsh parenting more than non-reactive children. The reactivity may cause a heightened stress response system and subsequent issues with behavioural regulation when parenting is rejecting and hostile (Belsky et al. 1998, 2007; Moore & Depue 2016).

In terms of the ten contrastive models, where beneficial and detrimental parenting for reactive and non-reactive children are at opposite ends of the distribution, these interactions have been illustrated in theoretical accounts of how child characteristics moderate the effect of parenting on behavioural outcomes (Belsky et al. 2007), yet a mechanism remains elusive. The mechanism for the reactive or difficult child may be that negative or positive experiences register more heavily on their experience and thus encourage poor or excellent behavioural regulation (Boyce 2016; Moore & Depue 2016), with similar findings to those above found for reactive children (Kim & Kochanska 2012; Roisman et al. 2012). How the mechanism operates for non-reactive children, however, is much more uncertain. For example, Stocker et al. (2017) used a polygenic index of reactivity and found that in low quality parenting environments (more hostility and less warmth) the proposedly least reactive children had comparably fewer externalizing and externalizing and internalizing behaviours than reactive children, yet in high quality parenting environments (less hostility and more warmth) the least reactive children had comparably more externalizing and externalizing and internalizing behaviours than reactive children. However, using only internalizing type behaviours did not replicate for better or worse patterns for non-reactive children. In any case, they limit the discussion to concluding the result is *difficult to explain and awaits replication* (Stocker et al. 2017). Likewise, Slagt et al. (2018) found that children with low sensory processing sensitivity (non-reactive) had increases in externalizing behaviours if positive (warm, responsive, autonomy granting and reasoned) parenting increased or negative (harsh, inconsistent and punitive) parenting decreased, when positive parenting decreased or negative parenting increased these non-reactive children had declines in externalizing behaviours. For children high in sensory processing sensitivity (reactive) the opposite trends was observed. They find the interaction effects were less influenced by the use of lower cut-points (e.g., 1 SD) for the reactive children and put forth that *future research would have to bear out the robustness of these findings and their potential explanations*. Finally, results from Leerkes et al. (2009) suggested two-way contrastive effects where higher parental sensitivity to distress protected temperamentally reactive children from behavioural dysregulation, but increased behavioural dysregulation in non-reactive infants. They proposed the sensitivity to distress may intrude on non-reactive children and limit self-regulation development or reduce compliance with parental assistance. Yet, again concluding *further research is needed to examine these findings*.

One possibility to explain the contrastive outcomes of non-reactive children is that high parental warmth and non-harsh parenting may indicate parenting that is over-protective of the non-sensitive child (Leerkes et al. 2009; Spokas & Heimberg 2009; Wood et al. 2002). Specifically, a high warmth or low harsh parenting environment that indicates overprotection may restrict stress regulation experiences that would register on less sensitive sensory systems and facilitate the development of coping strategies in non-reactive children (Edwards et al. 2010; Rapee 1997; Rubin et al. 2009; Ungar 2009). This may lead to further self-regulation issues when dealing with stressful interactions with peers or schools (Day et al. 2018). By not being exposed to experiences that allow them to develop stress regulation capability, these overprotected children may internalize or externalize stressful experiences they have not yet adjusted to experiencing at home (Rubin et al. 2009).

On the other hand, this overprotection idea runs counter to one observed interaction in this study. Specifically, children with high introversion (reactive) were associated with more internalizing behaviours if parental warmth increased over time, whilst children with low introversion (non-reactive) decreased in internalizing behaviours over time if parental warmth increased (Figure 4.3a). It could, therefore, be that increases in parental warmth may register as overprotection for already sensitive children such that they develop inadequate stress regulation skills. Whilst increases in parental warmth provide positive emotional support for the non-sensitive children who may be less receptive to parental warmth. As this is the opposite to the earlier findings of higher average warmth being a predictor of fewer or more internalizing behaviours for the reactive and non-reactive children, respectively, it seems difficult to reconcile these contrasting results without further study or the presence of other moderating effects of introversion.

Alternatively, high parental warmth may not fully capture the range of parenting experiences necessary to explain the development of internalizing behaviour. For example, because the typical reactive child might require more parental investment in behavioural management, equivalently high parental warmth for non-reactive children could actually indicate indulgent parenting and limited behavioural management (Hart et al. 2003; Sanson et al. 2004; Super et al. 2008). Indeed, there was a significant negative correlation between difficult temperament and parental warmth and a positive correlation between difficult temperament and harsh parenting. Thus, it could be that the proportion of non-reactive children with high warmth, indulgent parents is higher than that of reactive children due to bidirectional associations between child temperament and parenting (Kiff et al. 2011; Paulussen-Hoogeboom et al. 2007). This expansion to how the parenting environment is indexed could in part describe the cross-over interaction pattern but would need a multidimensional parenting measure to be clarified.

A final thought may be that this study may have disproportionally excluded children with relevant characteristics that could have influenced the estimation of interactions in the statistical model. As only one diathesis stress interaction and a single one-way contrastive interaction were from within person terms, most interactions were average, between-person terms that may be modelling unobserved variation between children. Thus, less rigorous evidence supported the two-way contrastive model.

### 4.10.2 Hypothesis 2: Multiverse analysis and consistency of effects

A multiverse analytical approach, that estimated nearly all possible variable combinations, was used to test the second hypothesis. Specifically, the second hypothesis was that if temperament moderated the effects of parenting on behaviour, then the interaction would be consistent at all ages and across time, consistent across reporters of behaviour, parenting environments, and similar behaviour measures, and generally consistent across temperament measures.

Support for this second hypothesis was largely absent as evidence that temperament moderated the effect of parenting on a behaviour had substantial inconsistencies tied to analytical decisions. There were no systematic pockets of evidence. Instead, overwhelming inconsistencies were observed across ages, measures of temperament, parenting, and behaviour, and mother and teacher reports of behaviour.

Two potential explanations for the relative inconsistencies are evident. First, the interactions may be adequate statistical descriptors and simply reflect high levels of within-population variability. All relevant criteria for determining an interaction were met and the results were occasionally replicated across different outcome types, parenting environments, and temperament scales. Thus, these results may be an example of different sensitivities to different environmental influences for different outcomes due to different reasons (Moore & Depue 2016). Indeed, variation in evidence for the theoretical models has been previously reported. Other studies have found variation across ages (Roisman et al. 2012), gender (Sulik et al. 2014) negative and positive parenting environments (Zhang et al. 2015), outcome domains (Roisman et al. 2012; Slagt et al. 2016b), study informants (Belsky et al. 2015; Hastings et al. 2015; Rabinowitz et al. 2016) and measures of sensitivity (Belsky et al. 2015; Slagt et al. 2018). Subsequently, associated reasons for this variation can be proposed, such as sensitive periods of development, gender based variation, and contextual variation of informants (Rabinowitz & Drabick et al. 2017).

The second explanation for inconsistency may relate to inadequacies in the measurement. Given extensive theoretical and empirical evidence supporting biological variation in reaction to environment (Boyce 2016; Moore & Depue 2016), a closer look at the measures used to index parenting, behaviour, and the sensitivity moderator may identify limitations.

### 4.10.3 Do limits of measurement explain inconsistent evidence?

Testing of Individual difference models require measurement of environment, child reactivity/sensitivity, and child outcome. Limitations in measurement of any one of these components may limit theoretical testing. Below, several limitations in these measurements are outlined.

In this study, parenting was the environmental measure. The sample had very high levels of parental warmth and low levels of harsh parenting. This indicates that variation in these practices may not have been adequately measured or that the parents were highly similar in these practices. Moreover, the reliability of harsh parenting was low in some instances. Further, as the parenting measures were self-report and not validated by observation (Hawes & Dadds 2006) there may have been measurement error resulting from social desirability of answers (Bennet et al. 2006), such that the range of parenting was not adequately captured (Gardner 2000).

The child outcome focused on behaviour. Again the data evidenced low levels of variability. Few children had extremes of behavioural difficulties. The behaviour measure, the SDQ, may not be appropriate in distinguishing children performing much better, or worse, than other children (Belsky & Pluess 2009; Rosiman et al. 2012). A historical examination of the development of the SDQ reveals it was initially used to classify children with and without psychiatric disorders (Goodman 1994, 1997). Instead of a dimensional measure of behavioural difficulties, this makes it a classifier of potential psychiatric disorder. Further, a later study aiming to prove its dimension aspect made incorrect inference. Specifically, Goodman & Goodman (2009) compared the ability of the SDQ to predict psychiatric disorder as a continuous (e.g, 0 to >32) and categorical measure (e.g., 0 vs 1, 0 vs 2, 0 vs 4 etc.). In all models they observed the continuous SDQ fit better overall than categorical SDQ categories. They incorrectly inferred this meant the "*odds of disorder increased at a constant rate across the range"* of the SDQ. Specifically, the tests of model fit used (log-likelihood ratio tests) penalise a model with more coefficients. Therefore, any non-significant difference between two sequential categories could demonstrate the categorical variable as a worse fit and a significant difference between one or a few sequential categories could not make enough difference to sway the overall model test. Thus, the overall model test is not appropriate to evaluate the dimensionality of the SDQ. Complicating their interpretation further, Figure 1 of their paper illustrating the prevalence of disorder (*y-axis*) against SDQ (*x-axis*) clearly shows overlapping 95% confidence intervals of disorder prevalence between many sequential SDQ scores. This implies there is a non-significant difference between these SDQ scores and thus their conclusions are misstated. In context of the present study, this suggests the SDQ may have captured problematic behaviours but may not represent normative and comparative behavioural development of children. In addition to these issues, reliability of behaviour reports was sometimes low, parent and teacher reports of behaviour could be inaccurate (An et al. 2018), and parent reports could correlate with their desired answers for temperament and parenting based questions.

The sensitivity moderator was indexed by parent reports of temperament at age 4 and 6 and may not have adequately captured biological sensitivity for two main reasons. First, the 12 items used were taken from a pool of 72, where originally 17 items measured reactivity, 10 measured persistence, and 9 items measured introversion (Sanson et al. 1994). It may be that the reduced number of items missed key variation between children. For example, Rosiman et al. (2012) found more consistent evidence than this study for difficult temperament moderating the effect of maternal sensitivity on behavioural development, but they used a 50 item indicator of infant temperament. Validation of the key temperament components Roisman et al. (2012) used may offer insight into why temperament was not as predictive in this study. Moreover, the biological aetiology of temperament was not confirmed with indicators of biological sensitivity, such as genetic and endophenotypic variation (Howarth et al. 2016; Nigg 2006; Saudino & Micalizzi 2015; Whittle et al. 2006). Similarly, due to strong links to biology temperament may be better indexed during infancy as temperament and environment interact across time and later measurement may not adequately capture biologically based sensitivity but rather behavioural reactions (Bornstein et al. 2015a; Roisman et al. 2012). Second, and more concerning, the 12 items chosen to index easy and difficult temperament were selected because they had the "*strongest relationship with our behaviour problem measures*" (Prior et al. 1989). These behaviour problems were indexed by the Preschool Behaviour Questionnaire (Behar & Stringfield 1974), which was based on the Rutter (1967) Children's Behaviour Questionnaire which, in slightly different form (Rutter 1970), was also used by Goodman (1994) in the initial development of the SDQ. Thus, by design the 12 items used to measure temperament in the present study have the strongest overlap with the measure of behavioural problems. It therefore seems possible the excluded (up to 60) temperament items might tap regulatory and sensitivity differences between children that may be relevant for individual by environment models of behavioural development (Lengua et al. 1998). Finally, parents may have been biased to indicate socially desirable child temperament, though because temperament requires information on child reactions to experiences over time mothers are often considered acceptable informants (Achenbach & Edelbrock 1978; Sanson et al. 1994).

Overall, considering there were very few consistent trends and the possible issues with measurement of parenting, temperament, and behaviour, it appears that the measurement limitations are a likely explanation for the varied and sporadic findings of this study.

### 4.11 Strengths, limitations and improvements

This study was strengthened by a large sample size, longitudinal design, use of both mother and teacher reports of behaviour, comprehensive analytical investigation, and indictors of both negative and positive aspects of parenting.

Limitations include that in several instances, residual diagnostics indicated that very high and low values of externalizing and internalizing behaviours could be predicted better. Improvements might be made using an alternative modelling strategy, such as using a Poisson or Negative Binomial model to account for the left-skewed distribution of these behaviours (Gardner et al. 1995). Nonetheless, Gaussian linear models, in various forms, are the primary tool that has been used within the child development literature to assess interactions for the hypotheses of interest (e.g., Roisman et al. 2012).

A further limitation is that a comparison of results using complete cases and partial information offset by data imputation was not performed in this study. Although the complete case was quite similar to the full sample on the variables of interest and the influence may likely be small, it may be that magnitude and precision of effects were underestimated (Goldstein et al. 2014; Graham et al. 2007; Grund et al. 2018; Mister & Enders 2017; Schafer & Yucel 2002). Imputing data would be a valuable contribution of future research to more fully explore result sensitivity. It must be kept in mind, however, that performing such imputation in an adequate way can be logistically difficult. Recent progress has made it possible to impute data with non-normal distributions and ordinal categories in multilevel models with substantive model compatible specifications that also allow for interactions to be imputed directly (Bartlett et al. 2015; Enders et al. 2018a; Enders et al. 2018b). Nonetheless, from experience, achieving convergence can be computationally burdensome and in some cases infeasible with large sample sizes, complex models and non-normally distributed, ordinal variables with many values (>10). Moreover, a model must be specified for each outcome variable. In context of the current study, this would equate to creating 32 separate data sets to validate the hybrid model results. Though difficult, however, this would improve confidence in the effects for a larger sample of children.

Finally, the limitations of measurement, as discussed above, present significant level of concern in interpretation of the findings. The study was dependent on secondary data analysis where a large sample size is traded-off for available measures with a limited number of items.

A major improvement to this study would be to apply the same rigorous analytical approach to higher quality measures of the parenting environment, higher quality and more extensive temperament measures combined with other indicators of biological sensitivity and high quality measures of behaviour that have more variation and dimensionality. By systematically varying key components of models and asking for rigorous evidence a detailed understanding of how child sensitivity moderates the effect of parenting on behavioural development may emerge.

### 4.12 Conclusion

This study examined how child temperament moderated the effects of parenting emotional quality on behavioural development in a large longitudinal sample of Australian children surveyed at ages 4, 6 and 8. Evidence was found that child temperament (difficult, reactivity and introversion) moderated the effect of parenting warmth and harsh parenting on the development of teacher and mother reported internalizing behaviours and conduct problems consistent with either two-way contrastive effects, one-way contrastive effects or diathesis stress. However, evidence of interaction effects was inconsistent across child ages, temperament components, parenting measures, teacher and mother behavioural reports and similar behavioural outcomes. The study highlighted the need to evaluate interactive models in detail and suggests future research could improve on this study by using higher quality and more extensive measures of temperament, parenting and behaviour combined with the same rigorous evaluation of statistical models.

## Chapter 5: Testing individual difference models in a longitudinal sample of Australian and US children: Parenting, infant temperament, and behavioural outcomes

### 5.1 Introduction

A long standing principle of developmental psychology is that individuals vary in their response to the environment (Slagt et al. 2016b; Wachs & Gandour 1983). That is, each child's development is the product of the interaction between their individual biology and developmentally salient experiences (Plomin et al. 2006; Rutter 2007). Several theoretical models have been proposed to explain how an individual's characteristics may moderate the effects of their early care environments on their ongoing behavioural development. These models include diathesis stress (Zuckerman 1991), differential susceptibility (Belsky 1997; Ellis et al. 2011), and vantage sensitivity (Pluess & Belsky 2013).

While there is a strong theoretical basis for the principle of individual variability and associated theoretical models (Boyce 2016; Pluess 2015), and empirical evidence in animal models and some experimental studies (Caspi & Moffitt 2006; Pawlak et al. 2008), the literature has so far been largely unable to produce systematic and convincing evidence in population samples to elucidate mechanisms. That is to say, our understanding of exactly how individual differences in reactivity and regulation moderate the effects of early care experiences on behavioural development remains unclear (Rabinowitz & Drabick 2017). Chapter 3 and 4 of this thesis have illustrated this point. Chapter 3 presented a literature review and found that of 542 interactions examining how individual differences moderate the influence of parenting on behavioural development, only 86 (16%) were considered statistically reliable. Further, these 86 significant interactions varied in support for individual by environment theoretical models. This variation was not clearly linked to differences between studies in terms of measurement and analytical design. This inconsistent evidence was elucidated in Chapter 4, where the multiple ways of exploring an interaction between temperament and parenting when predicting behavioural development showed two important insights. First, the large majority of interactions (97.5%) were sporadic and not statistically reliable. And second, the statistically reliable interactions (2.5%) supported varying theoretical models and differed across similar measures of parenting, temperament, behaviour, reporters of behaviour, and child age. This implies, therefore, that there are gaps in the theory-to-evidence knowledge continuum for individual by environment interactions.

As this gap in the theory-to-evidence knowledge continuum exists, it is important that additional research apply a more systematic strategy in testing theoretical models. By systematically exploring evidence generated over multiple studies, trends may be observed that highlight the optimal methods to test these theoretical models within the complexity of influences on human development. While *individual × environment* models are well supported in less complex developmental models, there are significant challenges when studying human development that has multiple 'moving parts' in terms of individual differences, environments, important developmental outcomes, and the methods used to measure these components (Dick et al. 2015; Meany 2010; Plomin & Bergeman 1991; Rutter et al. 1997; Sameroff 2000).

### 5.2 Role of measurement in individual × environment interactions

One possible source of variation in testing individual difference theoretical models of behavioural development is measurement (Ellis et al. 2011; Roisman et al. 2012). In this chapter, again examining *parenting × temperament* interactions as predictors of behaviour, four areas of variation in measurement are considered: coverage, content, information source, and child age at assessment

Regarding coverage, the distributions for some measures of parenting, temperament, or behaviour can be highly skewed and, therefore, not capture the range of population experiences. For example, the parental warmth scale analysed in Chapter 4 had a positively skewed distribution in which parental warmth reported at or below the scale mid-point in only 0.7% of the observations over three measurement occasions. Likewise, the harsh parenting scale was only reported at or below the midpoint for 4.1% of observations. Although some children may be exposed to hostile, low warmth, and invasively restrictive early parenting, if the parenting measure does not sufficiently capture these experiences inferential ability is lowered or entirely removed and efficiency of detecting interactions drastically declines (McClelland & Judd 1993). Using measures with more coverage of possible variation may therefore identify a greater number of individual by environment interactions (Ellis et al. 2011).

With regards to content, short-version scales used to measure temperament, parenting, and behaviour may not capture as much variation between children and parents as their full-item counterparts. More specifically, when measuring temperament the use of shortened scales may compromise the construct of temperament assessed and limit the potential to capture the range of underlying sensitivity to parenting (Levy 1968; Putnam & Rothbart 2006; Smith et al. 2000). For example, the easy–difficult scale from the Short Temperament Scale for Infants (STSI; Prior et al. 1989) consists of three dimensions (approach, irritability, cooperation), whereas the full STSI has two further dimensions (activity-reactivity, rhythmicity) that tap important regulatory and reactivity characteristics likely to distinguish children more sensitive to developmental experiences. By capturing more variation, longer, full-item scales may enhance the identification of individual by environment interactions (McClelland & Judd 1993). However, short-version scales are more likely to be favoured in large population studies as a means to reduce respondent burden and maximise the number of constructs measured (Elgar et al. 2006; Putnam et al. 2014).

Regarding measurement source, variation in the methods used to obtain information on behaviour, temperament, and parenting may explain variation in findings across studies (Achenbach et al. 1987; Patterson & Forgatch 1995; Seifer et al. 1994). For example, there are merits and limitations of measuring parenting behaviour via self-report or observation. Self-reports can be valuable for obtaining information on uncommon parenting behaviours and providing a general parenting overview, though there may be bias if parents choose to report a false ideal image of their parenting, have difficulty accurately recalling their parenting behaviour, or have an overly positive or negative view of their parenting (Bornstein et al. 2015b; Locke & Prinz 2002; Milner & Crouch 1997; Morsbach & Prinz 2006). Additionally, parental characteristics, such as mental health, may bias reported information (Richters 1992). On the other hand, whilst calibrated observations of researchers may accurately identify parenting experiences on a comparative scale, observation methods are typically limited in scope (Gardner 2000; Zaslow et al. 2006). Observation requires sampling an event or time period, raising questions about representativeness. Additionally, parents may modify their behaviour under observation or the observation context may not elicit parenting behaviour that is proposed as developmentally important (e.g., physically aggressive parenting in response to child behaviour; Hill et al. 2008). Therefore, the relative strengths and weaknesses of observation and self-report measures may play an important role in evaluating individual by environment interaction effects (Rabinowitz & Drabick 2017; Slagt et al. 2016b; Uher & McGuffin 2010).

Time of assessment, in the study of human behavioural development, is a substantial and complex source of measurement variation that may influence the identification of individual by environment interactions (Hall & Perona 2012). For instance, temperament while biologically driven and visible from early age is subject to environmental influence (Blandon et al. 2010; Bornstein et al. 2019; Janson & Mathiesen 2008). Thus, the measurement of genetically-driven temperament may be more reliable when measured early in life when effects of early developmental experiences are least likely to obscure the measurement of an underlying biological reactivity (Slagt et al. 2016a). Yet, at the same time, it may be that as temperament changes a child can fluctuate from sensitive to non-sensitive (Belsky & Pluess 2009; Roisman et al. 2012). Thus, earlier temperamental sensitivity may not identify later sensitivity (Roberts & DelVecchio 2000). These examples show that time of assessment maps to key theoretical considerations that may influence the identification of individual by environment interactions.

## 5.3 Aims of this chapter

Given the potential role of measurement in detecting individual by environment interactions, this chapter systematically explores how different measures of parenting, behaviour, and temperament may influence support for theoretical models of individual by environment interactions.

Specifically, the temperament by parenting interactions analysed in chapter 4 are modified to accommodate several new, potentially more promising measures. Data from the Longitudinal Study of Australian Children (LSAC) are used again, but employing measurement of temperament assessed when children were younger and employing an alternative measure of parenting that may cover more content through wider scaling.

In addition, this chapter incorporates a similar aged sample of children from the US Family Life Project (FLP) that has alternative and more extensive measures of temperament, two additional behaviour scales with more coverage, and measures of observed parenting assessed on two occasions. Importantly, an identical behavioural outcome is available for both studies.

By systematically exploring how different measures may influence support for the individual by environment theoretical models, this chapter may highlight key areas where measurement choices can bridge the gap between theory and evidence.

The following research questions, indicated by underlined text in Figure 5.1., are explored;

1) Do more symmetric distributions of parenting and behaviour measures identify more interactions than skewed distributions?
2) Does a longer (more items) measurement of parent reported temperament identify more interactions than a shorter one?
3) Are measures derived from observations of parenting behaviour more revealing of interaction effects compared to self-reported parenting measures?
4) Are more interactions identified when temperament is measured at a younger age compared to when temperament is assessed at a later age?

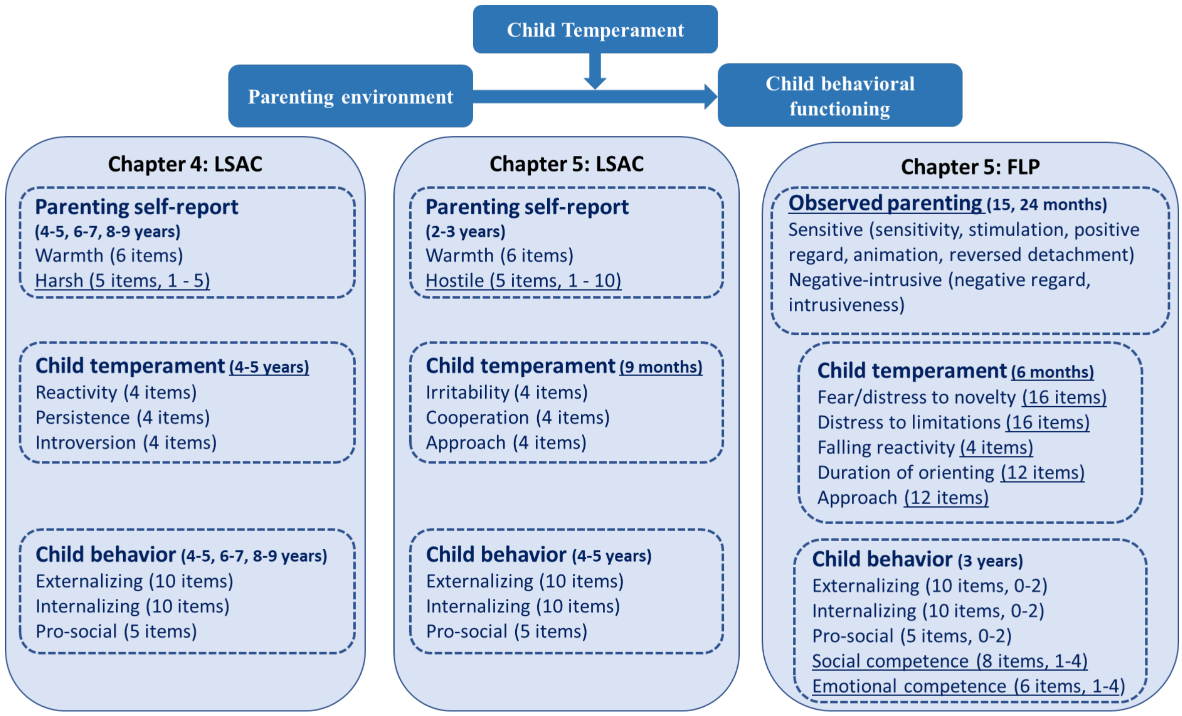



**Figure 5.1.** Outline of measurements used in chapter 4 and this chapter (5) to test *temperament* × *parenting* interactions predicting behavioural development. Systematic changes in chapter 5 are underlined and include 1) two behaviour and one parenting measure(s) with greater variance of response scales, 2) a longer (more items) and more comprehensive measure of temperament, 3) observed parenting measures (all other measures parent report), and 4) temperament measured when the child was younger.

## 5.4 Methods

## 5.5 Data

To address the research questions, we used secondary data from two different studies: The LSAC and the FLP.

### 5.5.1 Longitudinal Study of Australian children

The LSAC is a large longitudinal survey following two cohorts of Australian children recruited in across 2003-2004 to evaluate the influence of Australia's social and cultural environment on childhood development (LSAC 2016). The first cohort, the birth cohort, recruited infants younger than 1 year of age and was used in the current study. These infants were born between March 2003 and February 2004. The initial 5107 primary parents and their infants were selected using stratified random sampling from Australian Medicare records to be representative of all children in Australia. However, the Medicare record may not include all Australian residents and thus LSAC approximates a representative sample of children from urban and rural areas of all states and territories in Australia at baseline. A new wave of data was collected every two years, with waves 1 (age 0-1), 2 (age 2-3) and 3 (age 4-5) constituting the frame for key variables used in this study.

Data from the primary parent were collected for LSAC during home visits using face-to-face interviews and during which questionnaires were provided for postal return. Ages of children for the analysis sample in years were Mean (M) = 0.77 Standard Deviation (SD) = 0.21 at age 0-1 (wave 1), M = 2.90 SD = 0.24 at age 2-3 (wave 2), and M = 4.83 SD = 0.23 at age 4-5 (wave 3).

From the entire LSAC sample (n=5107), the analytical sample was first restricted to surveys of the primary caregiver who filled out the main externalizing and internalizing behavioural outcome survey in wave 3 (n=3609). The same caregiver was used to maintain consistency of the reporter of information and to use covariates observed before parenting was collected (Murray et al. 2007; Kersten et al. 2016). Nearly all primary caregivers at wave 3 were the child's biological mother (99%). Children were excluded if at the first wave (ages 0-1) the parent indicated the child had long term effects as a result of a head injury, stroke or other brain damage (n=1), blackouts (n=2), or chronic or recurring pain (n=0). Exclusion decisions were based on prior exclusions documented in the literature (e.g., Roisman et al. 2012) with the underlying rationale relating to the biasing effect of such disorders on parenting, temperament and behaviour (D'Souza & Karmiloff-Smith 2017; Kim et al. 2016). Additionally, due to convergence issues with missing categorical data when using imputation routines, parents who did not report education in wave 1 were excluded (n=4). Following exclusions, the remaining sample comprised 3602 children. Missing data were imputed using fully conditional specification methods specified in the analytical plan (Section 5.9). A comparison of included and excluded children on analytical variables is provided in the results section. Note that the 1289 children analysed in Chapter 4 are also in this chapter's included sample.

### 5.5.2 Family Life Project

The FLP is a longitudinal study of 1292 children and families living in non-urban, lower income communities in the United States of America (see Willoughby et al. 2013 for sampling procedure). Families and their newborns living in two major geographical areas characterised by being both rural and having high levels of child poverty (including three counties in eastern North Carolina and three counties in central Pennsylvania) were recruited using a stratified random sampling procedure. This stratification yielded an area based representative sample of families with a child born between September 15, 2003, and September 14, 2004. African American families were oversampled in North Carolina, but this was not necessary in Pennsylvania as the majority of the population of sampled communities was non-African American (over 95%).

The current study uses computerised parent-report and observational FLP data collected during home visits when participating children approximated 2, 6, 15, 24, and 35 months of age. Because parts of the FLP sample are restricted-access and data security requirements were infeasible to meet at time of analysis, some demographic information was unavailable for this analysis (e.g., caregiver education and relationship to child). However, a range of more detailed demographic processes influencing behavioural development were available. From information available, ages of children in months at each time point were M = 2.64 SD = 1.26 at 2 months (wave 1), M = 7.72 SD = 1.47 at 6 months (wave 2), M = 15.68 SD = 1.3 at 15 months (wave 3), M = 24.8 SD = 1.87 at 24 months (wave 4), and M = 37 SD = 1.75 at 35 months (wave 5).

In line with existing research conducted using the FLP (e.g., Raver et al. 2013; Mills-Koonce et al. 2016), children were not excluded from analysis on the basis of caregiver gender or possible impairments. Parenting data used in the analyses were from the care giver who reported on child behaviour at age 35 months. Nearly all primary care-givers of the entire sample were the child's biological mother (e.g., >96% at 35 month sample, Vernon-Feagans et al. 2013), so it is likely similar proportions are represented in the analysis sample (but this information was unavailable in the general release). Specifically, the analytical sample was restricted to children with information on the main externalizing and internalizing outcome variable at 35 months of age (n=1093). Missing data were imputed using fully conditional specification methods specified in the analytical plan (Section 5.9). A comparison of included and excluded children on analytical variables is provided in the results section.

For the FLP 15-month home visit, parents and children were provided a standardized set of developmentally appropriate toys and asked to play together as they normally would during free time. At the 24-month home visit, the parent and child were asked to complete a 10 minute puzzle task consisting of three puzzles of increasing difficulty. The parent was instructed that the task was for the child to complete, but they could help as needed. The play and puzzle tasks were video recorded for 10 minutes and later coded for parenting behaviours.

## 5.6 Measures

### 5.6.1 Moderator: Temperament (age 9 months) LSAC

Temperament was measured with a reduced version of the Short Temperament Scale for Infants (STSI; Sanson et al. 1987) when the child was around 9 months old. Three STSI subscales were measured; irritability (4 items; e.g., *this baby continues to cry in spite of several minutes of soothing*); cooperation (4 items; e.g., *this baby stays still during procedures like hair brushing or nail cutting*) and approach (4 items; e.g., *this baby's first reaction (at home) to approach by strangers is acceptance)*. Each subscale is the average of four items regarding how often a child's behaviour matches a statement; *almost never=1* to *almost always=6*. Cooperation and approach were reverse coded so that they indicated non-cooperation and non-approach. Higher scores indicate higher irritability, higher non-cooperation and higher non-approach. Difficult temperament was calculated as the average of irritability, non-cooperation and non-approach (Prior et al. 1989). Higher scores indicate a more difficult temperament. Reliability coefficients ranged from $\alpha=0.57$ to 0.74.

### 5.6.2 Moderator: Temperament (age 6 months) FLP

Temperament was measured with the revised Infant Behavior Questionnaire Revised (IBQ-R; Gartstein & Rothbart, 2003) via computer using Blaise software when the child was approximately 6 months old. Five IBQ-R subscales were measured regarding child temperament over the past week to two weeks; fear/distress to novelty (16 items, 5 required; e.g., *how often during the last week did the baby cry or show distress at a change in parents appearance, (glasses off, shower cap on, etc.)?*), distress to limitations (16 items, 5 required; e.g., *when something the baby was playing with had to be removed, how often did the baby cry or show distress for a time?*), approach (12 items, 4 required; e.g., *when the baby saw a toy they wanted, how often did the baby immediately go after it?*), duration of orienting *(*12 items, 4 required; e.g., *how often during the last week did the baby look at pictures in books and/or magazines for 5 minutes or longer at a time?*), and falling reactivity/recovery from distress (4 of 13 items, 2 required; e.g., *when the baby was upset about something, how often did they stay upset for up to 10 minutes or longer?*). Each subscale is the average of the items answered from *never=1* to *always=7*, and higher scores indicate greater frequency of the temperament construct. Though the IBQ-R was partially administered, these subscales are from higher level factors of extraversion/surgency (approach), negative affectivity (fear/distress to novelty, distress to limitations, (reversed) falling reactivity) and orienting/regulation (duration of orienting). Thus, only the higher level negative affectivity was calculated if two of the subscales were present. Pairwise (given enough required items were answered) reliability coefficients ranged from α=0.53 to 0.89.

### 5.6.3 Environment: Parenting warmth and hostile parenting (age 3 years) LSAC

Parental warmth was measured with six questions from the Child Rearing Questionnaire (Paterson & Sanson 1999) when the child was around 3 years old. Specifically, mothers were surveyed about their interactions with study child over the last six months in terms of question such as *how often did you express affection by hugging, kissing and holding this child?* Parents responded *never/almost never=1* to *all the time=5*. Reliability coefficient for this parental warmth measure was α=0.85.

Hostile parenting was measured using five items from the Early Childhood Longitudinal Study of Children, Birth Cohort (US Department of Education 2001) when the child was around 3 years old. The parent was surveyed for questions such as *I have raised my voice and shouted at this child.* In response, the parent answered *not at all=1* to *all the time=10*. Reliability coefficient for this hostile parenting measure was α=0.85.

### 5.6.4 Environment: Parenting during child play or challenge task (age 15 and 24 months) FLP

The 10 minute videos of parent and child interactions during play (around 15 months) and a challenge task (around 24 months) were coded for seven parenting dimensions by trained observers (Cox & Crnic 2002; NICHD Early Child Care Research Network 1999). Specifically, coders rated the entire interaction period for sensitivity (level of responsiveness and support offered to the child contingent on the child's needs), positive regard (positive feelings and warmth directed toward the child), stimulation (developmentally appropriate language use), animation (level of facial and tonal affect), detachment (degree to which the mother is disengaged), intrusiveness (controlling, parent-agenda driven behaviours) and negative regard (hostile verbal and physical treatment of the child). Over 45% of observations used were double coded (15 months n=976, Kappa with squared weights range κ=0.71 to 0.80; 24 months n=515, Kappa with squared weights range κ=0.82 to 0.89) and coders discussed score differences before assigning a final score. Ratings ranged from *not at all characteristic=1* to *highly characteristic=5* at 15 months and from *not at all characteristic=1* to *highly characteristic=7* at 24 months. The 15 and 24 month measures were analysed separately and as an average. Consequently, the 15–24 month average used a rescaled 24 months measure where 1=1, 2 and 3 = 2, 4 = 3, 5 and 6 = 4, and 7 = 5 (Mills-Koonce et al. 2011). Sensitive and negative-intrusive parenting composites were also calculated (Mills-Koonce et al. 2011). Sensitive parenting is average of, sensitivity, positive regard, stimulation, animation, and reversed-detachment, whilst negative-intrusive parenting is the average of negative regard and intrusiveness. The reliability coefficients for sensitive parenting (15 months, 24 and 15–24 average) ranged from α=0.89 to 0.91, whilst the respective range for negative-intrusive parenting was α=0.57 to 0.75. If parenting was missing at 15 or 24 months, the average consisted of the observed value.

### 5.6.5 Outcome: Externalizing and internalizing behaviours and social-emotional competence (age 35 months FLP; age 5 years LSAC)

Externalizing and internalizing behaviours and social-emotional competence were measured using the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997; Kersten et al. 2016) at approximately 5 years for the LSAC sample and 35 months for the FLP. The SDQ consists of five subscales including emotional difficulties (5 items; e.g., *often unhappy, depressed or tearful*), peer problems (5 items; e.g., *picked on or bullied by other children*), conduct problems (5 items; e.g., *often fights with other children or bullies them*), hyperactivity (5 items; e.g., *constantly fidgeting or squirming*) and pro-social (5 items; e.g., *considerate of other people's feelings*). Each subscale is the sum of five questions scored (reverse coded where relevant) *not true=0*, *somewhat true=1*, or *certainly true=2*. Higher level scales include internalizing (emotional difficulties and peer problems; 0–20), externalizing (conduct problems and hyperactivity; 0–20) and total difficulties (internalizing and externalizing; 0–40). Higher scores indicate more behavioural difficulties, except for the pro-social scale which indicates more social and emotional competence. For LSAC reliability coefficients for externalizing type scales ranged from α=0.67 to 0.77 and for internalizing type scales from α=0.54 to 0.65, whilst the pro-social scale was α=0.69 and the total scale was α=0.78. For FLP reliability coefficients for externalizing type scales ranged from α=0.68 to 0.78 and for internalizing type scales from α=0.5 to 0.7, whilst the pro-social scale was α=0.67 and the total scale was α=0.82.

Additionally, social and emotional competence were measured in the FLP sample using the Head Start Competence Scale (Domitrovich et al. 2000). Social competence is the average of 8 items (6 required) e.g., *resolves problems with fiends/siblings*, whilst emotional competence is the average of 6 items (4 required) e.g., *copes with sadness*. Items from *not at all well=1* to *very well=4*, thus higher scores indicate competence. Pairwise (as not all items were required) reliability coefficients were α=0.86 for social competence and α=0.86 for emotional competence.

### 5.6.6 Covariates (age 9 months LSAC, age 2–6 months FLP)

A range of covariates, with documented evidence of association with behavioural development (e.g., see Bayer et al. 2011; Haltigan et al. 2017; Kuckertz et al. 2017; Masten et al. 2005) were included in the analysis. If a covariate was observed more than once the average of that covariate was used. All time-varying covariates were taken before the first parenting measures used in analysis to avoid the bias of post-treatment variables (Gelman & Hill 2007, pg. 188-190; Montgomery et al. 2018). Covariates and age of collection are detailed in Table 5.1.

**Table 5.1:** Additional covariates used in analytical models.

| | Longitudinal Study of Australian Children | Family Life Project |
|---|---|---|
| Child age (continuous) | Age at 9 month sample. Years; M = 0.77 SD = 0.21. Months; M = 9.24 SD = 2.52. | Age at 6 month sample or age at 2 month sample plus 4 months. Years; M 0.64 SD = 0.12. Months; M = 7.72 SD = 1.48. |
| Parent age years (continuous) | Age at 9 month sample. M = 32.02 SD = 5.16. | Parent age at 2 month sample. If data was only available at later ages (6, 15, or 24 months), the closest age to 2 months (e.g., 6 over 15, 15 over 24) was reduced to approximate age at 2 months. M = 26.37 SD = 6.53. |
| Child gender | Male = 51.1%, Female = 48.9% | Male = 49.6%, Female = 50.4% |
| Child race | Aboriginal or Torres Strait Islander at 9 month sample (3.0%), Other (97%). | Black = 43.7%, Other = 56.3%. More detail unavailable in general release. |
| Caregiver education | At 9 month sample; Year 11 and certificate or trade qualification (25.3%); Year 12 and trade certificate or qualification or diploma (38.7%); Bachelor degree or higher (36.1%) | Unavailable in general release. |
| Socioeconomic status | Financial hardship scale (9 month sample). Six items assessing if parent experienced financial hardship, yes (1) or no (0), in the last 12 months e.g., *not been able to pay gas, electricity or telephone bills on time?* α=0.62. | Income-to-needs ratio (6 month sample; 2004). Household income divided by poverty thresholds defined by the United States Census Bureau. Observed minimum 0 and maximum 16.49 used for POMP calculation. |
| Caregiver depression | Kessler K6 screen scale (9 month sample; Kessler et al. 2003). Six items on how the parent felt the last four weeks, e.g., [*how often*] *did you feel worthless*, ranging from *all of the time* (1) to *none of the time* (5). Reverse coded so higher scores indicate more psychological distress. α=0.82. | Brief Symptom Inventory (Derogatis & Fitzpatrick 2004). Parent depression (6 items; e.g., *feelings of worthlessness*), anxiety (6 items; e.g., *nervousness or shakiness inside*), and somatization (6 items; e.g., *nausea or upset stomach*) symptoms ranging from *not at all* (0) to *extremely* (4). Global severity index calculated as average of 18 items across 2 and 6 month sampling periods. 2 month α=0.86 and 6 month α=0.89. |
| Home learning environment | Unavailable. | Infant Home Observation for Measurement of the Environment (HOME; Bradely et al. 1984). Sum of 11, minimum 4 required, observed (1) or not observed (0) conditions in home of child relating to learning materials e.g., *at least 10 books are present and visible*. 6 month sample, α=0.78. |
| Stressful experiences | Stressful Life Events, at 9 months. Sum of 14 *yes*=1, *no*=0 items regarding if the parent experienced stressful life event in the past year e.g., *suffered a serious illness, injury or assault.* α=0.5. | Life events checklist was used to create a sum of 47 experiences in the last 6 months noted as negative events weighted by the effect they had from no effect (0) to great effect (3); e.g., *major personal illness/injury*. 2 month sample, α=0.73. |
| | Longitudinal Study of Australian Children only | |
| Child impairments | Any present (1; n=17) or all absent (0) index of physical disabilities based on parent reporting child has limited use of arms or fingers, has difficulty gripping things, has limited use of legs or feet, has other physical condition or has other disfigurement. 9 month sample.<br>Any present (1; n=65) or all absent (0) index of sight, speech and hearing problems based on parent reporting child has sight problems, hearing problems, speech problems or ongoing problems with eyes or seeing properly or hearing problems. 9 month sample.<br>Children with either difficulty learning or understanding things or developmental delay were coded developmental delay present (1; n=34) or absent (0). 9 month sample. | |

## 5.7 Missing data imputation

Missing data were imputed using substantive model compatible fully conditional specification (SMC-FCS; Enders et al. 2018). Fully conditional specification imputes missing data for variable *y* using observed values of $y$ and a vector of observed covariates $\boldsymbol{x}$ (Mistler & Enders 2017). Specifically, Bayesian techniques are used to estimate and sample from univariate conditional distributions to impute values for missing $y$ given observed $y$ and $\boldsymbol{x}$. Because of the univariate conditional nature of imputation, once a variable $y$ has been sampled any missing data in $\boldsymbol{x}$ are sampled using information from the imputed $y$. Thus, *variables are imputed one at a time, with the filled-in variable from one step serving as a predictor in all subsequent imputation steps* (Enders et al. 2017).

Substantive model compatible implies that there is a joint model for which the estimated conditional associations for analysed variables equal those used for the conditional imputation (Bartlett et al. 2015). In practice, this involves imputing explanatory variables conditional on other explanatory variables with any polynomial and interaction terms of interest directly modelled, whilst the outcome variable does not impute the explanatory variables (Bartlett et al. 2015; Zhang & Wang 2017). When examining interactions the SMC-FCS approach is preferable because the interaction is not directly imputed (the just-another-variable approach) or passively computed (the impute-then-transform approach; Zhang & Wang 2017). Instead, in a single-step imputations from the sampling process are evaluated against the analytical model of interest and only valid observations are retained (Zhang & Wang 2017). Thus, SMC-FCS can reduce bias in parameter estimates of interaction effects (Bartlett et al. 2015; Zhang & Wang 2017).

To capture the combination of different measures of temperament (difficult, negative affectivity, separate components) and parenting (hostile, warmth, negative/intrusive, sensitive, separate components) missing data models were run for each outcome including as many substantive interactions as feasible. In total, this amounted to 136 datasets that each contained 20 imputed datasets. Imputation was performed using Blimp software (version 1.1; Enders et al. 2018; Keller & Enders 2018).

### 5.8 Analytical approach

A multiverse approach to the analysis was undertaken (Steegen et al. 2016). The multiverse analytical plan recognises that variables can be operationalised in multiple ways and therefore to avoid capitalising on chance, and to promote transparency, a large number of possibilities should be explored. Specifically, by rotating through the numerous ways variables might be analysed the sensitivity of the results to research choices can be visualised. If consistent trends are identified confidence in the results is raised (Dick et al. 2015). The approach to the multiverse is detailed below and was important for the current study because there are 8 to 10 outcome measures modelled by 2 to 9 parenting measures moderated by 4 to 6 reactivity measures in two data sets, equalling 1684 interactions.

Using the imputed sample, the parent rated behavioural outcomes in all scale forms (e.g., conduct problems, hyperactivity and externalizing) were predicted by interactions between parenting and temperament such that all unique and plausible combinations were evaluated without doubling up modelled information e.g., not including temperament subscales with overall temperament scales. These interactions were evaluated using a general linear model and coefficients combined using Rubin's rules for multiply imputed data (Barnard & Rubin 1999; Rubin 1987) with the simplified model form (Aiken et al. 1991) being;

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i$$ Equation 5.1

where *y* is the behavioural outcome, *i* indexes the child, *x* is the parenting measure, *z* is the measure of temperament, $\beta$ indicates the regression coefficients and, $\varepsilon$ is the residual errors that are normally distributed and independent with mean 0 and standard deviation σ, specifically, $\varepsilon_i \sim$ Normal(0, $\sigma^2$). $\beta_3$ is the interaction coefficient.

Interactions were evaluated for evidence of differential susceptibility, diathesis stress and vantage sensitivity using several criteria proposed by Roisman et al. (2012). First, the level of statistical significance was based on Benjamini and Hochberg (1995) procedure (critical value = 0.05) using only the interaction coefficients (1620 FLP, 64 LSAC) given the number of coefficients calculated in the multiverse analysis. Second, the Regions of Significance (ROS: Dearing and Hamilton 2006) for temperament given parenting, and parenting given temperament were calculated. Temperament must be a significant predictor at low and high, but not medium, values of parenting for differential susceptibility to occur, whilst diathesis stress and vantage sensitivity generally occur when temperament is significant at low, high or all values of parenting. Parenting should be significant at high or all values of temperament for differential susceptibility, diathesis stress or vantage sensitivity to occur. When parenting is significant at low and high values of temperament it can suggest an alternative model of contrastive effects. These contrastive effects either operate in two-way (cross-over) or one-way (no cross-over) fashion that depend on the temperament coefficient given parenting. Low and high refers to the mean minus 2 Standard Deviations (SD) or plus 2SD, respectively. Third, the cross-over point where the lines of high and low temperament or parenting cross was calculated. The lines must cross-over within the range of observed parenting values for differential susceptibility, whilst cross-over points within the range of observed temperament suggest contrastive effects. Fourth, the proportion of children with parenting or temperament scores above the cross-over points was calculated (proportion affected; PA). Generally, 0.16 of children above (positive parenting) or below (negative parenting) the cross-over point on parenting provides confidence that differential susceptibility is observed, and 0.16 or more children above (reactive) or below (non-reactive) the cross-over point on temperament provides confidence in two-way contrastive effects. Finally, if differential susceptibility was observed, quadratic parenting and quadratic parenting by temperament interactions were included to rule out non-linear functions of parenting.

For each analysis, diagnostic plots of the residuals were examined to evaluate the normality of the residuals and validate assumptions of the statistical models (Pinheiro & Bates 2000, pg. 174). All continuous covariates were rescaled to Percentage of Maximum Possible (POMP) so that they ranged from 0 to 100 (Cohen et al. 1999). General linear models were estimated using R (version 3.5.0; R Core Team 2018), with regression coefficients pooled using the MICE package (version 3.3.0; van Buuren & Groothuis-Oudshoorn 2011), ROS calculated using the interplot package (version 0.2.1; Solt & Hu 2018), and conditional interaction plots created using the visreg (version 2.5.0; Breheny & Burchett 2017) and ggplot2 (version 3.0.0; Wickham 2016) packages.

## 5.9 Results

### 5.9.1 Descriptive statistics

Descriptive statistics (means and correlations) for the focus parenting, temperament and child behavioural outcomes, using complete case data, are presented in Table 5.2 (LSAC) and Table 5.3 (FLP). Those for all other variables included in the analyses are provided in Appendix B. In terms of key and expected significant ($p<0.05$) associations; **parenting;** hostile/negative-intrusive parenting were inversely associated with warmth/sensitive parenting, respectively. **Temperament–parenting**; hostile parenting was associated with more difficult temperament, whilst parental warmth was associated with less difficult temperament. Likewise, negative-intrusive parenting was associated with more negative affect (higher distress to limitations and fear, lower falling reactivity), whilst sensitive parenting with respectively lower negative affect. **Behaviour–parenting**; positive parenting (warmth/sensitive) was associated with fewer externalizing and internalizing behaviours and more pro-social behaviour social and emotional competence, whilst these associations reversed for negative parenting (harsh/negative-intrusive). **Temperament–behaviour**; difficult temperament was associated with more externalizing and internalizing behaviour and fewer pro-social behaviours. Similarly, negative affect (higher distress to limitations and fear, lower falling reactivity) was associated with more externalizing and internalizing behaviour and lower pro-social behaviour and social and emotional competence. On the other hand, approach was associated with fewer internalizing behaviours and more pro-social behaviour, as well as more social and emotional competence in the FLP sample. A longer duration of orienting was associated with more social and emotional competence. **Scale reliability**; regarding scale reliability, parental warmth (α=0.846) and hostile parenting (α=0.846) were similar, whilst sensitive parenting (α=0.913) exceeded negative-intrusive parenting (α=0.702). Likewise, social (α=0.864) and emotional competence (α=0.86) had higher internal consistency than internalizing (α=0.651 to α=0.698), externalizing (α=0.772 to α=0.783), and pro-social behaviour (α=0.673 to α=0.685). Temperament measures in FLP had higher reliability (α=0.77 to α=0.888), except falling reactivity (α=0.526), compared to temperament in LSAC (α=0.57 to α=0.73). **Missing data;** finally, comparing included and excluded children, showed included children had higher parental warmth (0.10 pooled SD) and parenting sensitivity (0.40 pooled SD), slightly higher harsh parenting (0.02 pooled SD), lower negative-intrusive parenting (-0.16 pooled SD), less difficult temperament (-0.07 pooled SD), substantially lower fear (-1.04 pooled SD), fewer externalizing (-0.15 pooled SD) and internalizing (-0.31 pooled SD) behaviours, more pro-social behaviours (0.07 pooled SD) and, though compared to only a small number of children, lower social (-0.45 pooled SD) and emotional competence (-0.11 pooled SD). Thus, the analytical samples are generally weighted toward children with fewer behavioural difficulties and developmentally richer parenting.

**Table 5.2**: Correlations, means, standard deviations, reliability coefficients and effect sizes of difference (Hedge 1981) to excluded sample (positive numbers indicate included sample was higher for measure) for key variables used in the analysis of LSAC data. Summary information, except correlations, for all variables used in the analysis is provided in Appendix B. Underlined numbers are significant at $p<0.05$. All measures are parent report. Additional numbers for *temperament × parenting* cells are relative efficiency, compared to an optimal sampling design, and the equivalent sample size if the research design was optimal (McClelland & Judd 1993).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 Hostile parenting | | | | | | | | | |
| 2 Parental warmth | -0.14 | | | | | | | | |
| 3 Irritability | 0.14<br>0.022/61 | -0.12<br>0.021/65 | | | | | | | |
| 4 Un-cooperative | 0.11<br>0.025/69 | -0.18<br>0.024/76 | 0.3 | | | | | | |
| 5 Non-approach | 0.1<br>0.02/54 | -0.14<br>0.018/56 | 0.27 | 0.33 | | | | | |
| 6 Difficult temperament | 0.15<br>0.016/43 | -0.2<br>0.018/56 | 0.71 | 0.76 | 0.73 | | | | |
| 7 Externalizing | 0.3 | -0.11 | 0.13 | 0.13 | 0.01 | 0.12 | | | |
| 8 Internalizing | 0.14 | -0.09 | 0.1 | 0.08 | 0.12 | 0.13 | 0.34 | | |
| 9 Pro-social | -0.16 | 0.22 | -0.11 | -0.11 | -0.11 | -0.15 | -0.4 | -0.24 | |
| Mean | 23.3 | 90.4 | 29.8 | 36.6 | 25 | 30.4 | 5.4 | 4.8 | 7.7 |
| Standard Deviation | 14.39 | 10.39 | 16.09 | 17.51 | 16.53 | 12.22 | 3.28 | 2.42 | 1.76 |
| n | 2930 | 3448 | 3231 | 3235 | 3200 | 3145 | 3602 | 3602 | 3584 |
| Reliability (α) | 0.846 | 0.846 | 0.57 | 0.655 | 0.73 | 0.744 | 0.772 | 0.651 | 0.685 |
| Effect size Hedges' G | 0.02 | 0.10 | -0.04 | -0.03 | -0.11 | -0.07 | -0.15 | -0.31 | 0.07 |

**Table 5.3**: Correlations, means, standard deviations, reliability coefficients and effect sizes of difference (Hedge 1981) to excluded sample for key variables used in the analysis of FLP data. Summary information, except correlations, for all variables used in the analysis is provided in Appendix B. Underlined numbers are significant at $p<0.05$. Obs. measures are observed, all others are parent report. No children with externalizing, internalizing or pro-social behaviour measures were excluded. Additional numbers for *temperament × parenting* cells are relative efficiency, compared to an optimal sampling design, and the equivalent sample size if the research design was optimal (McClelland & Judd 1993).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Obs. Negative-Intrusiveness (15 – 24 month average) | | | | | | | | | | | | |
| 2 Obs. Sensitivity (15 – 24 month average) | -0.5 | | | | | | | | | | | |
| 3 Distress to limitations | 0.11<br>0.031/31 | -0.21<br>0.047/47 | | | | | | | | | | |
| 4 Fear | 0.24<br>0.043/41 | -0.34<br>0.057/55 | 0.35 | | | | | | | | | |
| 5 Falling reactivity | -0.19<br>0.053/52 | 0.28<br>0.081/80 | -0.28 | -0.3 | | | | | | | | |
| 6 Approach | -0.03<br>0.013/12 | 0.08<br>0.038/38 | 0.09 | -0.07 | 0.31 | | | | | | | |
| 7 Orienting | 0.09<br>0.036/35 | -0.06<br>0.035/34 | 0.01 | 0.13 | 0.07 | 0.35 | | | | | | |
| 8 Externalizing | 0.31 | -0.35 | 0.24 | 0.17 | -0.18 | -0.06 | -0.01 | | | | | |
| 9 Internalizing | 0.25 | -0.34 | 0.16 | 0.29 | -0.22 | -0.16 | 0.03 | 0.49 | | | | |
| 10 Pro-social | -0.22 | 0.24 | -0.17 | -0.18 | 0.18 | 0.18 | 0.06 | -0.37 | -0.27 | | | |
| 11 Emotional competence | -0.21 | 0.25 | -0.17 | -0.19 | 0.21 | 0.21 | 0.1 | -0.43 | -0.32 | 0.56 | | |
| 12 Social competence | -0.18 | 0.21 | -0.2 | -0.12 | 0.17 | 0.12 | 0.11 | -0.49 | -0.25 | 0.47 | 0.78 | |
| Mean | 33.5 | 46.3 | 40.9 | 30 | 70.3 | 71.4 | 44.4 | 7.8 | 3.9 | 6.4 | 2.5 | 2.2 |
| Standard Deviation | 16.87 | 18.36 | 13.23 | 16.54 | 15.27 | 14.99 | 16.05 | 3.73 | 2.94 | 1.97 | 0.66 | 0.59 |
| n | 1048 | 1048 | 1014 | 977 | 1003 | 1011 | 1001 | 1093 | 1093 | 1093 | 1092 | 1093 |
| Reliability (α) | 0.702 | 0.913 | 0.77 | 0.888 | 0.526 | 0.846 | 0.816 | 0.783 | 0.698 | 0.673 | 0.86 | 0.864 |
| Effect size *Hedges' G* | -0.16 | 0.40 | -0.02 | -1.04 | 0.11 | 0.11 | -0.02 | | | | -0.11 | -0.45 |

### 5.9.2 Key research questions

### 5.9.3 Interaction effects present in LSAC and FLP

After adjustment for multiple comparison **two** of the 64 interactions in the LSAC sample indicated robust evidence for the moderating effect of temperament (Table 5.4). On the other hand, **none** of the 1620 interactions in the FLP sample demonstrated robust evidence, even when the multiple comparison test was made less conservative by splitting the sample three-ways based on age parenting was assessed (Tables 5.5, 5.6, and 5.7).

### 5.9.4 Research Question 1: Do more symmetric distributions of parenting and behaviour identify more interactions than skewed distributions?

To assess this question, comparison is made across studies as the observed parenting measures of the FLP have more variation and are less skewed than the parent reported measures of LSAC. Additionally, FLP parenting measures have higher relative efficacy in terms of how close the variance matches an optimal design to test for interactions (McClelland & Judd 1993). However, even though the LSAC hostile parenting measure had more answer capability ranging from 1 to 10, hostile parenting remained highly skewed. Despite this, the larger sample size in LSAC overall gave it a similar efficacy in absolute terms (sample equivalent to optimal design) to the FLP. Thus, comparing FLP and LSAC parenting partially allows the test of whether wider distributions detect more interactions, but it is confounded with the FLP using observed parenting measures. Additionally, the relative efficacy of the two samples was quite low (max 0.081), indicating interaction effects will likely be hard to detect. On the other hand, the social and emotional competence measures in FLP had more variation and less skew then externalizing, internalizing, and pro-social behaviour. This allows a comparison of whether the distribution of behaviour matters for detecting interactions within the FLP.

As no interactions were observed in the FLP sample, however, there is insufficient evidence if wider distributions of parenting or behaviour identify a higher number of interactions. Likewise, the two interactions found in the LSAC sample do not convincingly (3% success rate) indicate skewed distributions of behaviour and parenting identify more interactions.

### 5.9.5 Research question 2: Does a longer measurement of parent reported temperament identify more interactions than a shorter one?

A longer measure of temperament used in the FLP sample did not identify interactions, whilst the shorter measure of temperament used in LSAC had a 3% success rate. This provides insufficient evidence that shorter or longer measures of temperament identify substantially more interactions.

### 5.9.6 Research question 3: Are observations of parenting more revealing of interaction effects compared to self-reported parenting?

The FLP sample used observations of parenting and did not find interaction effects, whilst the self-reported parenting in LSAC identified two interactions for a 3% success rate. These results do not support either observed or self-reported parenting as greatly improving the identification of interactions.

### 5.9.7 Research question 4: Is temperament measured at a younger age associated with more interactions compared to temperament assessed later?

The analysis of LSAC in chapter 4 used temperament measured at 4–5 years and identified interactions with a success rate of 2.5%, whilst the LSAC analyses here used infant temperament assessed at approximately 9 months and had a 3% success rate. This evidence indicates little advantage for either younger or older age of temperament assessment.

**Table 5.4**: *P*-values for assessed interactions in the LSAC sample. Bold, underlined interactions significant after adjusting for multiple comparisons. Light grey shading $p<0.05$, dark grey shading $p<0.01$. Codes; CondP=conduct problems, Hyper=hyperactivity, Extern=Externalizing, EmD= emotional difficulties, PeerP=peer problems, Intern=Internalizing, Total = total externalizing and internalizing, ProSo = pro-social, Diff=difficult, Irrit = irritability, N-Appr = non-approach, UnCoop = uncooperative.

| | Hostile parenting | | | | Parental warmth | | | |
|---|---|---|---|---|---|---|---|---|
| Outcome | Diff | Irrit | N-Appr | UnCoop | Diff | Irrit | N-Appr | UnCoop |
| CondP | 0.208 | 0.962 | 0.46 | 0.166 | 0.217 | 0.704 | 0.127 | 0.606 |
| Hyper | 0.441 | 0.747 | 0.082 | 0.534 | 0.226 | 0.995 | 0.062 | 0.492 |
| Extern | 0.181 | 0.758 | 0.113 | 0.209 | 0.112 | 0.922 | 0.041 | 0.5 |
| EmD | 0.015 | **<u>0.0004</u>** | 0.053 | 0.739 | 0.913 | 0.944 | 0.634 | 0.624 |
| PeerP | 0.685 | 0.0024 | 0.233 | 0.742 | 0.422 | 0.387 | 0.981 | 0.515 |
| Intern | 0.029 | **<u>0.00003</u>** | 0.384 | 0.935 | 0.565 | 0.663 | 0.585 | 0.918 |
| Total | 0.581 | 0.011 | 0.648 | 0.341 | 0.463 | 0.804 | 0.211 | 0.47 |
| ProSo | 0.617 | 0.978 | 0.467 | 0.172 | 0.17 | 0.011 | 0.841 | 0.36 |

**Table 5.5**: 15 – 24 month's average parenting p-values for assessed interactions in the FLP sample. No interactions significant after adjusting for multiple comparisons. Light grey shading *p*<0.05, dark grey shading *p*<0.01.

| 15 – 24 months | | Parenting | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Temperament** | **Behaviour** | Sensitivity | Positive regard | Stimulation | Animation | Detachment | Sensitive parenting | Negative regard | Intrusiveness | Negative-intrusive |
| Approach | Conduct problems | 0.868 | 0.167 | **0.019** | 0.218 | 0.085 | 0.074 | 0.12 | 0.37 | 0.184 |
| | Externalizing | 0.862 | 0.209 | **0.028** | 0.522 | 0.148 | 0.135 | 0.309 | 0.746 | 0.521 |
| | Hyperactivity | 0.804 | 0.536 | 0.135 | 0.919 | 0.474 | 0.391 | 0.954 | 0.735 | 0.92 |
| | Emotional difficulties | 0.311 | 0.291 | 0.757 | 0.568 | 0.533 | 0.511 | 0.334 | 0.731 | 0.459 |
| | Peer problems | 0.275 | 0.144 | 0.944 | 0.59 | 0.691 | 0.418 | 0.17 | **0.038** | **0.047** |
| | Internalizing | 0.194 | 0.165 | 0.971 | 0.561 | 0.559 | 0.304 | 0.148 | 0.13 | 0.093 |
| | Total SDQ | 0.63 | 0.934 | 0.144 | 0.917 | 0.525 | 0.7 | 0.151 | 0.348 | 0.165 |
| | Pro-social | 0.675 | 0.555 | 0.5 | 0.74 | 0.276 | 0.642 | 0.887 | 0.603 | 0.857 |
| | Social competence | 0.135 | 0.385 | 0.097 | 0.182 | 0.203 | 0.16 | 0.294 | 0.401 | 0.227 |
| | Emotional competence | 0.45 | 0.53 | 0.2 | 0.349 | 0.32 | 0.231 | 0.679 | 0.767 | 0.657 |
| Distress to limitations | Conduct problems | 0.744 | 0.973 | 0.874 | 0.828 | 0.469 | 0.998 | 0.232 | 0.087 | 0.097 |
| | Externalizing | 0.433 | 0.847 | 0.7 | 0.55 | 0.83 | 0.763 | 0.228 | 0.146 | 0.121 |
| | Hyperactivity | 0.376 | 0.766 | 0.697 | 0.419 | 0.843 | 0.702 | 0.421 | 0.481 | 0.436 |
| | Emotional difficulties | 0.206 | 0.114 | **0.035** | 0.343 | **0.022** | 0.139 | **0.025** | 0.372 | 0.095 |
| | Peer problems | 0.582 | 0.246 | 0.167 | 0.664 | 0.216 | 0.247 | 0.078 | 0.463 | 0.118 |
| | Internalizing | 0.293 | 0.119 | **0.044** | 0.469 | **0.04** | 0.072 | **0.011** | 0.282 | **0.041** |
| | Total SDQ | 0.922 | 0.315 | 0.38 | 0.412 | 0.155 | 0.435 | **0.024** | 0.091 | **0.035** |
| | Pro-social | 0.191 | 0.058 | 0.13 | **0.038** | **0.025** | **0.037** | 0.952 | 0.607 | 0.816 |
| | Social competence | 0.467 | **0.013** | 0.599 | 0.089 | 0.056 | 0.097 | 0.693 | 0.968 | 0.973 |
| | Emotional competence | 0.162 | **0.044** | 0.297 | **0.029** | **0.043** | **0.046** | 0.699 | 0.9 | 0.823 |
| Falling reactivity | Conduct problems | 0.939 | 0.393 | 0.252 | 0.906 | 0.469 | 0.947 | 0.746 | 0.595 | 0.753 |
| | Externalizing | 0.673 | 0.396 | 0.435 | 0.903 | 0.553 | 0.752 | 0.895 | 0.173 | 0.337 |
| | Hyperactivity | 0.283 | 0.975 | 0.678 | 0.908 | 0.95 | 0.573 | 0.792 | 0.055 | 0.252 |
| | Emotional difficulties | 0.447 | 0.172 | 0.762 | 0.482 | **0.033** | 0.253 | 0.324 | 0.493 | 0.342 |
| | Peer problems | 0.455 | 0.18 | 0.254 | 0.57 | 0.43 | 0.379 | 0.081 | 0.708 | 0.247 |
| | Internalizing | 0.815 | 0.074 | 0.393 | 0.409 | **0.046** | 0.147 | 0.104 | 0.64 | 0.162 |
| | Total SDQ | 0.793 | 0.134 | 0.864 | 0.647 | 0.148 | 0.621 | 0.316 | 0.152 | 0.178 |
| | Pro-social | 0.638 | 0.632 | 0.426 | 0.985 | 0.943 | 0.715 | 0.978 | 0.178 | 0.395 |
| | Social competence | 0.158 | 0.286 | 0.338 | 0.101 | 0.7 | 0.242 | 0.491 | 0.304 | 0.363 |
| | Emotional competence | 0.264 | 0.489 | 0.878 | 0.315 | 0.73 | 0.327 | 0.237 | 0.309 | 0.223 |
| Fear | Conduct problems | 0.994 | 0.774 | 0.828 | 0.774 | 0.766 | 0.74 | 0.527 | 0.359 | 0.424 |
| | Externalizing | 0.586 | 0.345 | 0.596 | 0.751 | 0.618 | 0.792 | 0.991 | 0.141 | 0.518 |
| | Hyperactivity | 0.392 | 0.167 | 0.193 | 0.276 | 0.56 | 0.47 | 0.676 | 0.223 | 0.87 |
| | Emotional difficulties | 0.339 | 0.084 | 0.257 | **0.046** | 0.074 | 0.054 | 0.345 | 0.559 | 0.281 |
| | Peer problems | 0.128 | **0.031** | **0.019** | 0.067 | **0.03** | **0.009** | 0.721 | 0.233 | 0.376 |
| | Internalizing | 0.11 | **0.02** | **0.023** | **0.014** | **0.013** | **0.006** | 0.353 | 0.279 | 0.215 |
| | Total SDQ | 0.286 | 0.059 | 0.148 | 0.142 | 0.123 | 0.074 | 0.495 | 0.17 | 0.276 |
| | Pro-social | 0.715 | 0.732 | 0.142 | 0.458 | 0.453 | 0.49 | 0.399 | 0.355 | 0.359 |
| | Social competence | 0.838 | 0.738 | 0.808 | 0.894 | 0.563 | 0.899 | 0.834 | 0.78 | 0.992 |

**Table 5.5**: continued

| Temperament | Behaviour | Sensitivity | Positive regard | Stimulation | Animation | Detachment | Sensitive parenting | Negative regard | Intrusiveness | Negative-intrusive |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Parenting** | | | | | | | | |
| Fear | Emotional competence | 0.615 | 0.958 | 0.435 | 0.941 | 0.862 | 0.977 | 0.257 | 0.404 | 0.555 |
| Negative affect | Conduct problems | 0.664 | 0.693 | 0.505 | 0.848 | 0.668 | 0.684 | 0.673 | 0.466 | 0.539 |
| | Externalizing | 0.46 | 0.688 | 0.674 | 0.937 | 0.969 | 0.737 | 0.572 | 0.447 | 0.444 |
| | Hyperactivity | 0.446 | 0.597 | 0.873 | 0.569 | 0.832 | 0.974 | 0.546 | 0.481 | 0.481 |
| | Emotional difficulties | 0.222 | **0.04** | 0.141 | 0.11 | **0.009** | **0.022** | 0.351 | 0.659 | 0.562 |
| | Peer problems | 0.632 | **0.031** | **0.035** | 0.177 | 0.073 | **0.047** | 0.154 | 0.961 | 0.444 |
| | Internalizing | 0.239 | **0.01** | **0.023** | 0.075 | **0.005** | **0.006** | 0.162 | 0.728 | 0.382 |
| | Total SDQ | 0.846 | 0.057 | 0.285 | 0.18 | 0.091 | 0.24 | 0.224 | 0.433 | 0.295 |
| | Pro-social | 0.417 | 0.326 | 0.086 | 0.294 | 0.282 | 0.227 | 0.733 | 0.203 | 0.377 |
| | Social competence | 0.715 | 0.462 | 0.921 | 0.835 | 0.443 | 0.979 | 0.972 | 0.888 | 0.822 |
| | Emotional competence | 0.934 | 0.757 | 0.516 | 0.836 | 0.592 | 0.798 | 0.961 | 0.972 | 0.72 |
| Orienting | Conduct problems | 0.625 | **0.018** | 0.764 | 0.068 | 0.055 | 0.125 | 0.862 | 0.153 | 0.364 |
| | Externalizing | 0.707 | 0.113 | 0.643 | 0.368 | 0.128 | 0.318 | 0.244 | 0.302 | 0.139 |
| | Hyperactivity | 0.944 | 0.869 | 0.777 | 0.662 | 0.623 | 0.932 | 0.073 | 0.662 | 0.126 |
| | Emotional difficulties | 0.446 | 0.706 | 0.986 | 0.613 | 0.824 | 0.563 | 0.332 | 0.555 | 0.453 |
| | Peer problems | 0.415 | 0.436 | 0.614 | 0.368 | 0.784 | 0.396 | 0.209 | 0.597 | 0.236 |
| | Internalizing | 0.391 | 0.872 | 0.896 | 0.986 | 0.911 | 0.802 | 0.179 | 0.57 | 0.267 |
| | Total SDQ | 0.821 | 0.373 | 0.922 | 0.623 | 0.302 | 0.602 | 0.126 | 0.284 | 0.1 |
| | Pro-social | 0.177 | 0.054 | 0.09 | 0.15 | 0.304 | 0.069 | 0.843 | 0.22 | 0.46 |
| | Social competence | 0.755 | 0.59 | 0.291 | 0.356 | 0.31 | 0.291 | 0.546 | 0.664 | 0.944 |
| | Emotional competence | 0.463 | 0.429 | 0.174 | 0.258 | 0.125 | 0.107 | 0.147 | 0.919 | 0.195 |

**Table 5.6**: 15 month parenting p-values for assessed interactions in the FLP sample. No interactions significant after adjusting for multiple comparisons. Light grey shading *p*<0.05, dark grey shading *p*<0.01.

| | | Parenting | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Temperament** | **Behaviour** | Sensitivity | Positive regard | Stimulation | Animation | Detachment | Sensitive parenting | Negative regard | Intrusiveness | Negative-intrusive |
| Approach | Conduct problems | 0.734 | 0.491 | 0.133 | 0.298 | 0.341 | 0.368 | **0.038** | 0.515 | 0.102 |
| | Externalizing | 0.678 | 0.524 | 0.194 | 0.62 | 0.449 | 0.413 | 0.375 | 0.778 | 0.523 |
| | Hyperactivity | 0.671 | 0.788 | 0.613 | 0.688 | 0.739 | 0.825 | 0.548 | 1 | 0.753 |
| | Emotional difficulties | 0.09 | 0.118 | 0.343 | 0.193 | 0.276 | 0.106 | 0.256 | 0.813 | 0.308 |
| | Peer problems | 0.581 | 0.63 | 0.969 | 0.665 | 0.494 | 0.989 | 0.052 | 0.05 | **0.014** |
| | Internalizing | 0.14 | 0.177 | 0.472 | 0.464 | 0.68 | 0.218 | 0.058 | 0.224 | **0.04** |
| | Total SDQ | 0.261 | 0.714 | 0.632 | 0.893 | 0.74 | 0.923 | 0.099 | 0.327 | 0.104 |
| | Pro-social | 0.989 | 0.87 | 0.579 | 0.332 | 0.478 | 0.963 | 0.8 | 0.687 | 0.837 |
| | Social competence | 0.381 | 0.241 | 0.439 | 0.367 | 0.524 | 0.372 | 0.153 | 0.734 | 0.244 |
| | Emotional competence | 0.886 | 0.124 | 0.526 | 0.282 | 0.663 | 0.381 | 0.413 | 0.379 | 0.226 |
| Distress to limitations | Conduct problems | 0.641 | 0.75 | 0.672 | 0.502 | 0.279 | 0.64 | 0.72 | 0.504 | 0.572 |
| | Externalizing | 0.898 | 0.684 | 0.604 | 0.386 | 0.654 | 0.93 | 0.571 | 0.816 | 0.744 |
| | Hyperactivity | 0.861 | 0.709 | 0.571 | 0.48 | 0.896 | 0.852 | 0.517 | 0.896 | 0.696 |
| | Emotional difficulties | 0.318 | 0.198 | 0.162 | 0.289 | 0.101 | 0.177 | 0.225 | 0.88 | 0.451 |
| | Peer problems | 0.948 | 0.362 | 0.455 | 0.915 | 0.894 | 0.549 | 0.08 | 0.674 | 0.184 |
| | Internalizing | 0.502 | 0.134 | 0.162 | 0.471 | 0.25 | 0.178 | 0.05 | 0.657 | 0.206 |
| | Total SDQ | 0.688 | 0.361 | 0.697 | 0.394 | 0.334 | 0.38 | 0.087 | 0.434 | 0.138 |
| | Pro-social | 0.226 | 0.495 | 0.282 | 0.201 | 0.132 | 0.099 | 0.651 | 0.583 | 0.536 |
| | Social competence | 0.359 | 0.077 | 0.983 | 0.381 | 0.512 | 0.219 | 0.321 | 0.477 | 0.378 |
| | Emotional competence | 0.122 | **0.038** | 0.972 | 0.126 | 0.2 | 0.128 | 0.357 | 0.922 | 0.566 |
| Falling reactivity | Conduct problems | 0.609 | 0.773 | 0.584 | 0.765 | 0.878 | 0.959 | 0.755 | 0.808 | 0.839 |
| | Externalizing | 0.568 | 0.967 | 0.915 | 0.752 | 0.822 | 0.931 | 0.65 | 0.629 | 0.974 |
| | Hyperactivity | 0.695 | 0.786 | 0.633 | 0.928 | 0.762 | 0.892 | 0.647 | 0.316 | 0.809 |
| | Emotional difficulties | 0.208 | 0.057 | 0.211 | **0.013** | **0.009** | **0.02** | 0.205 | 0.387 | 0.277 |
| | Peer problems | 0.323 | 0.485 | 0.108 | 0.453 | 0.785 | 0.816 | 0.233 | 0.825 | 0.537 |
| | Internalizing | 0.668 | 0.091 | 0.069 | **0.047** | 0.103 | 0.064 | 0.146 | 0.715 | 0.326 |
| | Total SDQ | 0.533 | 0.376 | 0.385 | 0.337 | 0.254 | 0.31 | 0.561 | 0.629 | 0.5 |
| | Pro-social | 0.422 | 0.881 | 0.376 | 0.945 | 0.962 | 0.685 | 0.291 | 0.125 | 0.109 |
| | Social competence | 0.777 | 0.063 | 0.456 | 0.154 | 0.347 | 0.288 | 0.89 | 0.825 | 0.999 |
| | Emotional competence | 0.894 | 0.073 | 0.771 | 0.7 | 0.958 | 0.576 | 0.752 | 0.559 | 0.731 |
| Fear | Conduct problems | 0.96 | 0.851 | 0.677 | 0.305 | 0.624 | 0.628 | 0.385 | 0.355 | 0.262 |
| | Externalizing | 0.416 | 0.569 | 0.639 | 0.995 | 0.949 | 0.79 | 0.859 | 0.151 | 0.343 |
| | Hyperactivity | 0.13 | 0.299 | 0.176 | 0.304 | 0.442 | 0.315 | 0.582 | 0.125 | 0.591 |
| | Emotional difficulties | 0.484 | 0.239 | 0.174 | **0.049** | 0.207 | 0.121 | 0.293 | 0.414 | 0.143 |
| | Peer problems | 0.325 | 0.119 | 0.058 | 0.367 | 0.609 | 0.15 | 0.912 | 0.372 | 0.723 |
| | Internalizing | 0.261 | 0.096 | **0.032** | 0.059 | 0.265 | 0.054 | 0.619 | 0.27 | 0.232 |
| | Total SDQ | 0.2 | 0.198 | 0.112 | 0.318 | 0.408 | 0.233 | 0.639 | 0.126 | 0.173 |
| | Pro-social | 0.842 | 0.85 | 0.115 | 0.851 | 0.837 | 0.658 | 0.29 | 0.377 | 0.331 |
| | Social competence | 0.762 | 0.786 | 0.794 | 0.724 | 0.898 | 0.881 | 0.678 | 0.369 | 0.811 |

**Table 5.6:** continued

| Temperament | Behaviour | Parenting: Sensitivity | Positive regard | Stimulation | Animation | Detachment | Sensitive parenting | Negative regard | Intrusiveness | Negative-intrusive |
|---|---|---|---|---|---|---|---|---|---|---|
| Fear | Emotional competence | 0.847 | 0.978 | 0.934 | 0.251 | 0.393 | 0.746 | 0.503 | 0.434 | 0.456 |
| Negative affect | Conduct problems | 0.984 | 0.922 | 0.556 | 0.5 | 0.93 | 0.819 | 0.899 | 0.948 | 0.809 |
| | Externalizing | 0.577 | 0.783 | 0.997 | 0.982 | 0.774 | 0.945 | 0.969 | 0.67 | 0.863 |
| | Hyperactivity | 0.487 | 0.765 | 0.592 | 0.611 | 0.706 | 0.883 | 0.795 | 0.82 | 0.933 |
| | Emotional difficulties | 0.093 | **0.033** | 0.051 | **0.009** | **0.006** | **0.007** | 0.597 | 0.774 | 0.928 |
| | Peer problems | 0.855 | 0.231 | 0.092 | 0.61 | 0.867 | 0.339 | 0.172 | 0.671 | 0.655 |
| | Internalizing | 0.247 | **0.031** | **0.015** | **0.037** | 0.069 | **0.022** | 0.293 | 0.646 | 0.725 |
| | Total SDQ | 0.404 | 0.171 | 0.198 | 0.269 | 0.263 | 0.162 | 0.476 | 0.726 | 0.981 |
| | Pro-social | 0.265 | 0.972 | 0.104 | 0.761 | 0.511 | 0.42 | 0.298 | 0.127 | 0.13 |
| | Social competence | 0.809 | 0.776 | 0.779 | 0.35 | 0.668 | 0.879 | 0.854 | 0.396 | 0.575 |
| | Emotional competence | 0.387 | 0.902 | 0.95 | 0.848 | 0.875 | 0.866 | 0.604 | 0.841 | 0.803 |
| Orienting | Conduct problems | 0.768 | 0.233 | 0.622 | 0.145 | 0.433 | 0.579 | 0.474 | **0.031** | 0.06 |
| | Externalizing | 0.661 | 0.663 | 0.481 | 0.765 | 0.502 | 0.789 | 0.202 | **0.024** | **0.039** |
| | Hyperactivity | 0.815 | 0.798 | 0.495 | 0.486 | 0.642 | 0.696 | 0.217 | 0.269 | 0.104 |
| | Emotional difficulties | 0.596 | 0.697 | 0.44 | 0.872 | 0.815 | 0.662 | 0.549 | 0.421 | 0.511 |
| | Peer problems | 0.601 | 0.533 | 0.568 | 0.369 | 0.958 | 0.48 | 0.278 | 0.658 | 0.322 |
| | Internalizing | 0.566 | 0.515 | 0.441 | 0.562 | 0.882 | 0.467 | 0.31 | 0.525 | 0.296 |
| | Total SDQ | 0.52 | 0.932 | 0.365 | 0.838 | 0.678 | 0.735 | 0.169 | 0.066 | 0.051 |
| | Pro-social | 0.101 | 0.157 | **0.026** | 0.251 | 0.464 | 0.05 | 0.586 | 0.059 | 0.15 |
| | Social competence | 0.954 | 0.801 | 0.19 | 0.648 | 0.479 | 0.566 | 0.311 | 0.836 | 0.376 |
| | Emotional competence | 0.48 | 0.877 | **0.03** | 0.267 | 0.295 | 0.173 | 0.397 | 0.467 | 0.192 |

**Table 5.7**: 24 month parenting p-values for assessed interactions in the FLP sample. No interactions significant after adjusting for multiple comparisons. Light grey shading $p<0.05$, dark grey shading $p<0.01$.

| | | Parenting | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Temperament** | **Behaviour** | Sensitivity | Positive regard | Stimulation | Animation | Detachment | Sensitive parenting | Negative regard | Intrusiveness | Negative-intrusive |
| Approach | Conduct problems | 0.672 | 0.116 | **0.023** | 0.209 | 0.175 | 0.068 | 0.382 | 0.588 | 0.494 |
| | Externalizing | 0.542 | 0.136 | **0.033** | 0.186 | 0.316 | 0.135 | 0.52 | 0.989 | 0.779 |
| | Hyperactivity | 0.731 | 0.748 | 0.401 | 0.57 | 0.915 | 0.455 | 0.561 | 0.643 | 0.856 |
| | Emotional difficulties | 0.888 | 0.782 | 0.443 | 0.405 | 0.876 | 0.896 | 0.947 | 0.631 | 0.683 |
| | Peer problems | 0.204 | **0.029** | 0.855 | 0.171 | 0.22 | 0.086 | 0.521 | 0.129 | 0.291 |
| | Internalizing | 0.524 | 0.211 | 0.675 | 0.738 | 0.354 | 0.351 | 0.847 | 0.677 | 0.631 |
| | Total SDQ | 0.869 | 0.999 | 0.19 | 0.635 | 0.885 | 0.66 | 0.369 | 0.708 | 0.604 |
| | Pro-social | 0.579 | 0.436 | 0.266 | 0.616 | 0.355 | 0.731 | 0.592 | 0.966 | 0.772 |
| | Social competence | 0.111 | 0.761 | 0.084 | 0.114 | 0.092 | 0.156 | 0.532 | 0.274 | 0.225 |
| | Emotional competence | 0.485 | 0.468 | 0.3 | 0.818 | 0.478 | 0.609 | 0.086 | 0.278 | 0.104 |
| Distress to limitations | Conduct problems | 0.219 | 0.858 | 0.715 | 0.686 | 0.748 | 0.9 | 0.08 | 0.129 | 0.143 |
| | Externalizing | 0.114 | 0.961 | 0.794 | 0.844 | 0.552 | 0.956 | 0.236 | 0.162 | 0.277 |
| | Hyperactivity | 0.201 | 0.825 | 0.536 | 0.782 | 0.446 | 0.761 | 0.645 | 0.45 | 0.637 |
| | Emotional difficulties | 0.116 | 0.118 | **0.021** | 0.531 | 0.064 | **0.012** | 0.227 | 0.807 | 0.977 |
| | Peer problems | 0.916 | 0.53 | 0.09 | 0.788 | 0.35 | 0.589 | 0.356 | 0.49 | 0.148 |
| | Internalizing | 0.329 | 0.133 | **0.015** | 0.475 | 0.072 | **0.02** | 0.078 | 0.612 | 0.275 |
| | Total SDQ | 0.733 | 0.375 | 0.179 | 0.72 | 0.548 | 0.29 | 0.081 | 0.271 | 0.187 |
| | Pro-social | 0.213 | **0.018** | 0.105 | 0.1 | **0.039** | 0.07 | 0.84 | 0.86 | 0.809 |
| | Social competence | 0.969 | **0.027** | 0.522 | 0.065 | 0.055 | 0.229 | 0.994 | 0.431 | 0.626 |
| | Emotional competence | 0.589 | 0.232 | 0.284 | 0.099 | 0.233 | 0.134 | 0.751 | 0.782 | 0.595 |
| Falling reactivity | Conduct problems | 0.488 | 0.246 | 0.316 | 0.831 | 0.696 | 0.722 | 0.456 | 0.386 | 0.4 |
| | Externalizing | 0.093 | 0.414 | 0.101 | 0.833 | 0.753 | 0.392 | 0.328 | 0.08 | 0.137 |
| | Hyperactivity | **0.038** | 0.996 | 0.168 | 0.508 | 0.257 | 0.197 | 0.404 | 0.053 | 0.108 |
| | Emotional difficulties | 0.741 | 0.59 | 0.671 | 0.215 | 0.263 | 0.688 | 0.462 | 0.822 | 0.711 |
| | Peer problems | 0.964 | 0.059 | 0.47 | 0.453 | 0.286 | 0.404 | 0.169 | 0.769 | 0.179 |
| | Internalizing | 0.931 | 0.14 | 0.948 | 0.759 | 0.188 | 0.409 | 0.111 | 0.573 | 0.299 |
| | Total SDQ | 0.324 | 0.157 | 0.364 | 0.742 | 0.573 | 0.778 | 0.204 | 0.191 | 0.122 |
| | Pro-social | 0.907 | 0.138 | 0.533 | 0.861 | 0.321 | 0.833 | 0.636 | 0.395 | 0.88 |
| | Social competence | **0.036** | 0.774 | 0.143 | **0.046** | 0.264 | **0.046** | 0.43 | 0.18 | 0.132 |
| | Emotional competence | 0.05 | 0.715 | 0.614 | 0.071 | 0.057 | 0.052 | 0.244 | 0.799 | 0.233 |
| Fear | Conduct problems | 0.524 | 0.514 | 0.733 | 0.972 | 0.806 | 0.578 | 0.685 | 0.994 | 0.691 |
| | Externalizing | 0.618 | 0.365 | 0.723 | 0.875 | 0.938 | 0.941 | 0.802 | 0.866 | 0.649 |
| | Hyperactivity | 0.742 | 0.347 | 0.321 | 0.698 | 0.655 | 0.684 | 0.625 | 0.974 | 0.643 |
| | Emotional difficulties | 0.485 | 0.073 | 0.791 | 0.267 | 0.161 | 0.321 | 0.136 | 0.68 | 0.387 |
| | Peer problems | 0.101 | **0.037** | **0.017** | **0.004** | **0.001** | **0.016** | 0.301 | 0.303 | 0.342 |
| | Internalizing | 0.142 | **0.01** | 0.106 | **0.012** | **0.002** | **0.031** | 0.152 | 0.409 | 0.366 |
| | Total SDQ | 0.726 | **0.037** | 0.251 | 0.152 | 0.146 | 0.216 | 0.28 | 0.601 | 0.663 |
| | Pro-social | 0.354 | 0.142 | 0.283 | 0.404 | 0.333 | 0.284 | 0.258 | 0.551 | 0.188 |
| | Social competence | 0.424 | 0.441 | 0.84 | 0.872 | 0.973 | 0.923 | 0.343 | 0.77 | 0.968 |

**Table 5.7**: continued

| Temperament | Behaviour | Parenting: Sensitivity | Positive regard | Stimulation | Animation | Detachment | Sensitive parenting | Negative regard | Intrusiveness | Negative-intrusive |
|---|---|---|---|---|---|---|---|---|---|---|
| Fear | Emotional competence | 0.902 | 0.871 | 0.658 | 0.817 | 0.849 | 0.763 | 0.33 | 0.539 | 0.349 |
| Negative affect | Conduct problems | 0.179 | 0.644 | 0.551 | 0.788 | 0.977 | 0.764 | 0.497 | 0.323 | 0.213 |
| | Externalizing | 0.053 | 0.521 | 0.468 | 0.809 | 0.43 | 0.567 | 0.286 | 0.159 | 0.15 |
| | Hyperactivity | 0.086 | 0.607 | 0.73 | 0.937 | 0.255 | 0.498 | 0.551 | 0.244 | 0.281 |
| | Emotional difficulties | 0.232 | 0.073 | 0.427 | 0.784 | 0.06 | 0.103 | 0.998 | 0.755 | 0.787 |
| | Peer problems | 0.529 | **0.03** | **0.039** | **0.048** | **0.012** | 0.108 | 0.477 | 0.937 | 0.301 |
| | Internalizing | 0.197 | **0.015** | **0.048** | 0.166 | **0.002** | **0.038** | 0.523 | 0.952 | 0.643 |
| | Total SDQ | 0.631 | **0.048** | 0.542 | 0.473 | 0.256 | 0.331 | 0.424 | 0.399 | 0.264 |
| | Pro-social | 0.41 | **0.029** | 0.14 | 0.316 | 0.493 | 0.311 | 0.772 | 0.426 | 0.662 |
| | Social competence | 0.193 | 0.166 | 0.86 | 0.925 | 0.744 | 0.558 | 0.928 | 0.314 | 0.36 |
| | Emotional competence | 0.602 | 0.358 | 0.568 | 0.888 | 0.984 | 0.89 | 0.87 | 0.651 | 0.876 |
| Orienting | Conduct problems | 0.524 | **0.039** | 0.215 | 0.149 | 0.066 | 0.161 | 0.347 | 0.884 | 0.874 |
| | Externalizing | 0.604 | 0.176 | 0.346 | 0.392 | 0.219 | 0.278 | 0.927 | 0.946 | 0.839 |
| | Hyperactivity | 0.886 | 0.89 | 0.592 | 0.976 | 0.908 | 0.771 | 0.219 | 0.904 | 0.541 |
| | Emotional difficulties | 0.673 | 0.405 | 0.54 | 0.357 | 0.518 | 0.725 | 0.713 | 0.975 | 0.799 |
| | Peer problems | 0.331 | 0.184 | 0.351 | 0.35 | 0.412 | 0.162 | 0.194 | 0.645 | 0.334 |
| | Internalizing | 0.387 | 0.957 | 0.863 | 0.749 | 0.932 | 0.801 | 0.311 | 0.882 | 0.585 |
| | Total SDQ | 0.83 | 0.427 | 0.562 | 0.485 | 0.443 | 0.514 | 0.513 | 0.969 | 0.575 |
| | Pro-social | 0.123 | 0.05 | 0.223 | 0.162 | 0.217 | 0.096 | 0.517 | 0.495 | 0.946 |
| | Social competence | 0.483 | 0.342 | 0.601 | 0.235 | 0.291 | 0.325 | 0.489 | 0.604 | 0.86 |
| | Emotional competence | 0.079 | 0.107 | 0.272 | 0.114 | **0.037** | 0.144 | 0.35 | 0.984 | 0.906 |

### 5.9.8 Exploration of theoretical models

The two interactions found in the LSAC sample indicate one case of differential susceptibility and one case of two-way contrastive effects. Differential susceptibility implies the effects of parenting are stronger for reactive children and weaker or absent for non-reactive children. In the current data, the specific interaction was that hostile parenting interacted with irritability to predict emotional difficulties consistent with differential susceptibility (Figure 5.2). Two-way contrastive effects implies that the direction of effect is different for reactive and non-reactive children. In the current analysis, internalizing behaviour outcomes evidenced two-way contrastive effects. Children with high temperamental irritability showed greater internalising response to hostile parenting while low irritability children actually showed less internalising behaviours under conditions of hostile parenting (Figure 5.3). Additionally, highly irritable children had fewer internalising behaviours with low hostile parenting, while low irritability children had more internalising behaviours when hostile parenting was low. Evidence confirming differential susceptibility and two-way contrastive effects is presented in Table 5.8. Note that some ROS lower bounds are below 0. However, imputation can make numbers outside the scale range and maintain viable regression inference (von Hippel 2009). Additionally, the two-way contrastive effects lower ROS bound is within 2.06SD (not 2SD) and has a PA below the crossover of 0.15, not 0.16. These are interpreted as close enough to evidential cut points.

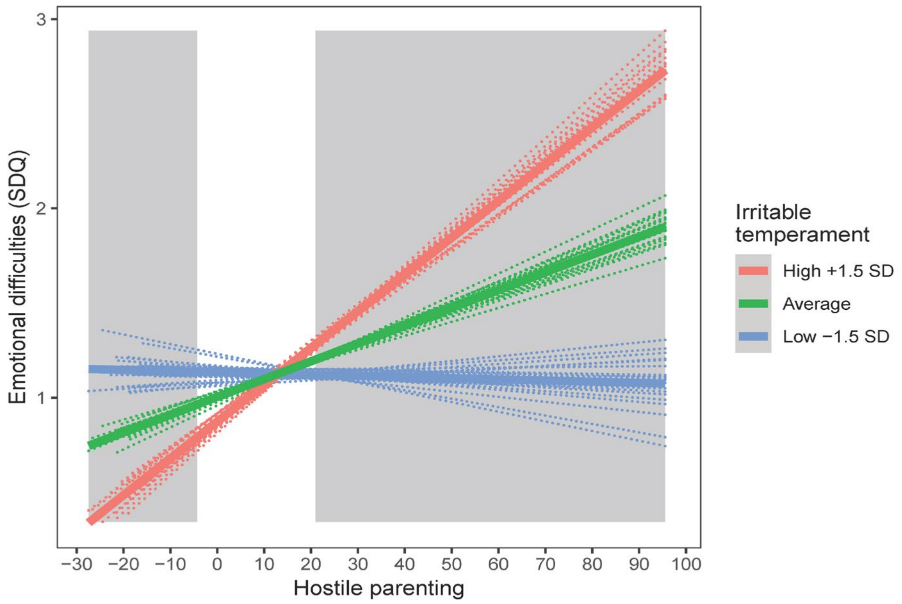



**Figure 5.2:** Conditional interaction plot with regions of significance (grey polygons) where hostile parenting effects on emotional difficulties significantly differ due to temperament (red +1.5 SD, green = average, and blue -1.5 SD). Model is consistent with differential susceptibility. Solid line is a regression of the dashed lines obtained from each imputed data set, thus the cross-over point is not identical to Table 5.8 that contains correctly combined regression coefficients.

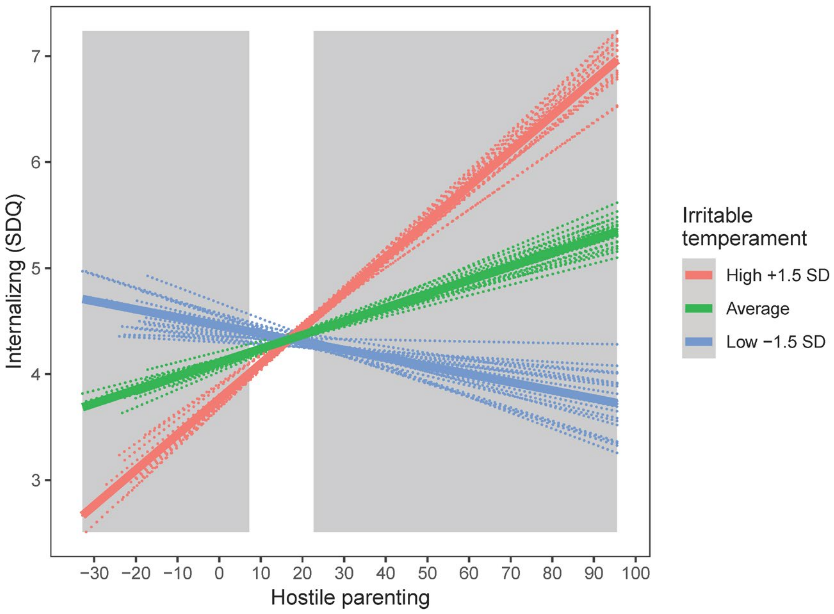



**Figure 5.3:** Conditional interaction plot with regions of significance (grey polygons) where hostile parenting effects on internalizing behaviour significantly differ due to temperament (red +1.5 SD, green = average, and blue -1.5 SD). Model is consistent with two-way contrastive effects. Solid line is a regression of the dashed lines obtained from each imputed data set, thus the cross-over point is not identical to Table 5.8 that contains correctly combined regression coefficients.

**Table 5.8**: Criteria to evaluate theoretical models. Columns 2–5 indicate regression coefficients for the intercept, parenting environment, temperament and interaction, respectively, p-value refers to interaction p-value. All interactions are significant after adjusting for multiple comparisons. CO = crossover point, PA = proportion affected. Outcome codes; EmotD = emotional difficulties, Internal = internalizing, Hostile P = hostile parenting. Temperament codes; Diff = difficult, React = reactivity, Intro = introversion. Model codes; CE2 = two-way contrastive effects, DS = differential susceptibility. * = $p<0.05$, ** <0.01, ***<0.001. < and > indicate significance occurred outside observed values. Nearly all ROS bounds within 2SD of mean, # indicates exception which was within 2.06SD, see results text. Mean and SD used for ROS lower and upper approximate (as model specific data was used) those for imputed data in Appendix B, Table B.1.

| | | | | | | ROS temperament on parent | | | | ROS parent on temperament | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome | Intercept | Hostile P | Irritability | Inter | p-value | Lower | Upper | CO | PA > | Lower | Upper | CO | PA > | Model |
| EmotD | 1.76 | -0.003 | -0.005 | 0.000416 | 0.000415 | -4.29 | 20.96 | 13.03 | 0.77 | < | 19.88 | 7.20 | 0.92 | DS |
| Internal | 6.34 | -0.012 | -0.014 | 0.00085 | 0.00003 | 7.19 | 22.71 | 16.83 | 0.63 | -3.32# | 22.68 | 14.57 | 0.85 | CE2 |

## 5.10 Discussion

The aim of this chapter was to extend chapter 4 and systematically explore how different measures may influence the empirical support for *individual × environment* theoretical models. Specifically, measures of parenting, temperament, and behaviour were systematically varied to assess the number, type and reliability of *temperament × parenting* interactions explaining behaviour. Two large data sets, the Family Life Project (United States of America) and Longitudinal Study of Australian Children, were analysed to assess four key questions centred on measurement:

1) Do more symmetric distributions of parenting and behaviour measures identify more interactions than skewed distributions?
2) Does a longer (more items) measurement of parent reported temperament identify more interactions than a shorter one?
3) Are measures derived from observations of parenting behaviour more revealing of interaction effects compared to self-reported parenting measures?
4) Are more interactions identified when temperament is measured at a younger age compared to when temperament is assessed at a later age?

Results showed that measurement variation did not increase empirical support for any of the *individual × environment* models. Only two interactions were observed as statistically reliable thus no viable evidence *supporting or rejecting* any of these research questions was found. Though systematic changes in measurement had a viable rationale for testing, they did not yield a coherent evidence profile. The discussion proceeds by discussing each research question, describing the observed interactions, and concluding with strengths and limitations and future directions.

### 5.10.1 Research question 1 – Do more symmetric distributions of parenting and behaviour identify more interactions?

Although parenting and the social and emotional competence scales in the FLP had more coverage and less skew compared to parenting and behavioural measures applied in LSAC, no more statistically reliable interactions were observed in the FLP sample. The finding did not support the hypothesised increased number of reliable interactions deriving from improved coverage. This is unexpected, as Ellis et al. (2011) & Roisman et al. (2012) noted that sufficiently high and low parenting and child behavioural outcomes may be requisite for identifying theoretical models of individual variation to the environment. Moreover, bivariate relationships between parenting and behaviour were observed suggesting the variance in parenting could be meaningful for behavioural development (Table 5.2 and 5.3). These findings may indicate the temperament scales measured do not index a sensitivity to parenting that influences behavioural development. On the other hand, high measurement error in parenting, behaviour, and temperament may have prevented detection of interactions (Wong et al. 2004). The very low relative efficiency of the samples (Table 5.2 and 5.3) compared to an optimal design to detect interaction effects (e.g., max 0.081, equivalent to an optimal sample of 80 children) suggests that there was a very low likelihood of detecting interactions in the sample (McClelland & Judd 1993). Thus, sampling more children from distributional limits of the parenting and temperament dimensions might better rule out, or show, the moderating effects of temperament.

### 5.10.2 Research question 2 – Do longer (more items) measures of temperament identify more interactions?

Although the IBQ-R employed in the FLP dataset had considerably more items and subscales assessing temperament than the STSI, no statistically reliable interactions were observed in the FLP sample. This indicates increasing the number of temperament items and subscales does not necessarily identify temperamental variation more indicative of sensitivity to parenting experiences. It may be more important, therefore, to focus on selecting the 'right' temperament items that indicate developmentally meaningful sensitivity to parenting (Moore & Depue 2016). A-priori item identification is difficult, however, as sensitivity is most easily inferred retrospectively once differential responses to parenting are observed. Nonetheless, researchers have moved towards novel coding of temperament and emotional reactivity (Slagt et al. 2019) to identify sensitivity in children that may better identify the proposed variation driving differential responses to parenting (Greven et al. 2019). That is, the focus is to identify salient aspects of temperament that capture individual variation in response to parenting.

### 5.10.3 Research question 3 – Does measures derived from observations of parenting identify more interactions than parent reported parenting?

Although observations of parenting in the FLP sample had more variation than self-reported parenting in LSAC, there were no statistically reliable interactions identified when using observations of parenting and only a trivial amount identified with parent reported parenting. The merits and disadvantages of parent report and observed parenting have been noted (e.g., Gardner 2000; Bornstein et al. 2015b), yet the method of parenting assessment did not influence the results of this study. Other authors examining *temperament* × *parenting* interactions have found observed parenting improved identification of interactions (Slagt et al. 2016a), whilst others found little difference in identification of interactions (Slagt et al. 2016b). In and of itself, therefore, changes in the way in which parenting is measured do not always lead to identifying more or fewer interactions. However, the results in this study are limited as parenting was not measured by observation and parent report in the same study. Nonetheless, it is likely much more important to have a parenting measure that is valid and reliable for testing the interaction over a measure of parenting collected by particular format (e.g., observation vs. survey). When either method provides more validity and reliability it is to be preferred.

### 5.10.4 Research question 4 – are more interactions identified when temperament is measured at a younger age compared to when temperament is assessed at a later age

Compared to the analysis of LSAC presented in Chapter 4 which assessed temperament at 4-5 years, this study used a measure of infant temperament and did not find a meaningfully larger number of statistically reliable interactions. Assessing temperament earlier may have a closer link to genetically based sensitivity given the reduced time for environmental drivers to have an effect (Goldstein et al. 1987). Though, a change from earlier temperament may also be problematic if sensitivity changes and the younger measurement becomes redundant (Roberts & DelVecchio 2000; Shiner et al. 2012). As so few interactions were observed in this chapter, it suggests infant temperament was unable to identify consistent and systematic differential responses to parenting in the LSAC sample.

### 5.10.5 Notes on observed interactions

Although no evidence for the research questions was found, two (of 1684) interactions were statistically robust after adjustment. One was consistent with differential susceptibility. Children with high irritability, compared to children with low irritability, had more peer problems when experiencing more hostile parenting and fewer peer problems when exposed to lower hostile parenting. This is theoretically grounded, as hostile parenting may have a stronger effect on irritable children (Belsky & Pluess 2009; Kiff et al. 2011). Specifically, hostile parenting could model low emotion regulation and encourage hostile responding to peers (Barker et al. 2008; Dodge et al. 1995; Patterson 2002). Conversely, low parenting hostility may teach children how to regulate behaviour in challenging social interactions (Yamagata et al. 2013). If irritability indexed sensitivity, these children may be more receptive to parenting hostility and consequently develop social skills that alleviate or exacerbate inter-peer problems (Bradley & Corwyn 2008; Kim & Kochanska 2012).

The other interaction was consistent with two-way contrastive effects. Specifically, children with high irritability were found to react to hostile parenting by exhibiting more peer problems, but fewer peer problems in the absence of hostile parenting. In contrast, those with low irritability had more peer problems when hostile parenting was lower and less peer problems, compared to high irritability children, when exposed to hostile parenting. Thus, the effects of parenting are contrastive and reverse dependent on child irritability. There is little theoretical rationale for why two-way contrastive effects would occur, as noted by other studies reporting two-way contrastive effects (Leerkes et al. 2009; Slagt et al. 2018; Stocker et al. 2017). Nonetheless, it may be that neurologically different children recognise and respond to different levels of parenting hostility. Specifically, if less/more irritable children respond positively/negatively to parental hostility, in terms of behavioural regulation, parenting hostility or low hostility may indicate a tailored parenting strategy to assist children with behavioural regulation (Chess & Thomas 1991). The converse parenting strategy for less/more irritable children thus indicates non-tailored, less effective parenting.

A final explanation for these interactions, though they may have an empirical reality beyond chance, is that they are the result of chance and spurious correlation (Steegen et al. 2016). Indeed, it seems unlikely that they warrant serious consideration given they were not present across different parenting measures, behavioural measures, or similar measures of temperament. Thus, until research finds strong and systematic evidence otherwise, we might conclude these findings are sporadic.

Thus, there are two possible conclusions to be drawn from the interactions observed in this study. The first is that interactions between temperament and parenting have limited effect on behavioural development, such that models of individual differences in response to parenting are empirically invalid in this context. The second conclusion is that individual differences do matter in terms of the effects of parenting on behavioural development, but the measures and study design were inadequate to detect them. Given the extensive theoretical and empirical demonstration that children have individual neurological differences that influence their sensitivity and susceptibility to experience (Boyce 2016), it seems more likely to accept the second conclusion and more rigorously evaluate the hypotheses in future research. Some suggestions for these improvements are detailed below.

### 5.11 Strengths and limitations

This research was strengthened by a systematic and thorough exploration of analytical decisions, robust imputation methods to handle missing data, the use of two data sources, and longitudinal separation of measurement that may have reduced confounding. However, a major limitation is that analyses were dependent on available data and limited to exploration of measurement variation within each data source. For example, observed and parent reported parenting were not simultaneously collected for comparison. Likewise, longer and shorter versions of temperament were not simultaneously collected and comparatively evaluated. In addition, although using the latest techniques, missing data imputation may have introduced bias if the data was not missing at random and too few covariates were included to accurately estimate missing values (Bartlett et al. 2015; Enders et al. 2018). A future sensitivity analysis could quantify this potential bias. Moreover, as only general release data was available for the Family Life Project analysis, some demographic information could not be included. Finally, the biggest limitation was that the non-causal design increases the risk of incorrect inference and selection bias, and inhibits concluding any observed associations or non-associations as empirical (Gelman & Hill 2006; Mitchell et al. 2013). Additional research would therefore be needed to conclusively evaluate the research questions.

### 5.12 Conclusion and future directions

This chapter assessed 1684 interactions and found no empirically robust evidence for temperament moderating the effects of parenting on behavioural development. Systematically varying the measure of temperament, parenting, and behaviour did not reveal consistent evidence, and neither did more symmetric distributions of parenting and behaviour, the use of observed parenting compared to parent reported parenting, or the use of temperament measured at younger ages. Future research to validate the systematic approach to measurement effects in *temperament* × *parenting* interactions could be improved by collecting or examining multi-method measures of temperament, behaviour, and parenting in the same study combined with simulation studies to evaluate effects of methodological decisions (Morris et al. 2019). Additionally, longitudinal, within-person and experimental design to accommodate more causal analysis would strengthen the ability to infer if and how temperament moderates the effect of parenting, whilst reducing the risk of selection bias and sporadic associations. These improvements and others are discussed in Chapter 6, where extensive consideration is given to ways forward when testing *individual* × *environment* interactions.

## Chapter 6: Future directions and recommendations for research

The empirical results of this thesis (Chapters 3, 4, and 5) show that the evidence for *individual* × *environment* interactions, where individual child characteristics moderate the effects of parenting on child behavioural development, is inconsistent and sporadic and, therefore, of tenuous quality. In this chapter, future directions and recommendations for improving the quality and consistency of evidence in testing *individual* × *environment* models are explained and examined in terms of three key phases of quantitative research: design, analysis, and results.

There are seven approaches, identified here, that future research might adapt to improve the consistency and quality of evidence for *individual* × *environment* interactions. These approaches include changes and choices regarding design, analysis, and interpretation of results. It is important to note these approaches are interdependent, yet implementing all seven can also be impractical. For example, a causal design with poor measurement may be of limited value, and an optimum design for a specific analytical model can preclude systematically exploring many analytical models optimally. Thus, each approach improves an aspect of the research process, but may not be applicable to all studies. A flow chart of the changes and choices is presented in Figure 6.1. Each of these approaches is discussed in detail.

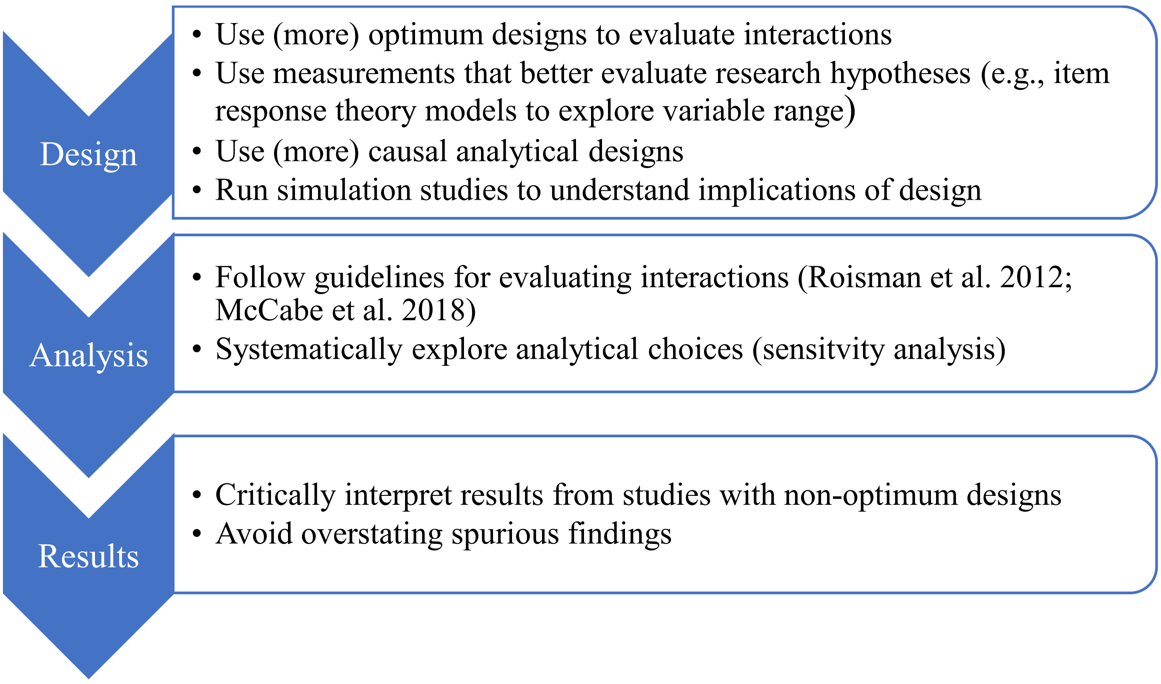



**Figure 6.1:** Future directions and recommendations for research.

## 6.1 Design

### 6.1.1 Optimum design

Using optimum design is one way in which research into *individual* × *environment* interactions may identify more valid evidence. Optimum (also known as optimal) design, in this context, refers to sampling to obtain the maximum power to test an effect (McClelland 1997). It developed out of work on designed experiments where researchers wanted to improve the precision of estimated effects and use fewer experimental resources (Atkinson et al. 2007). Thus, as optimum must always occur in context, the aim of optimum design is efficient and unbiased estimation of an effect. Once the effect of interest is specified, along with all other study constraints and statistical modelling choices, the optimum design can be calculated either analytically for simple designs or using computer based algorithms for more complex designs (Berger & Wong 2009). Therefore, the optimum sample in context of this thesis refers to the values of the *individual* and *environment* that best evaluate an interaction effect.

The use of optimum design, though applicable to any form of quantitative research, is uncommon in the *individual* × *parenting* literature. For example, none of the papers reviewed in this thesis identified how close their sample was to the optimum for testing of a moderating effect. Calls for using optimum design in developmental psychology are not new (e.g., McClelland & Judd 1993), but they appear to have been overlooked. Limited uptake might be best explained by optimum design being less easy to implement when studying the development of children in complex social environments compared with studies conducted in the physical sciences. In the physical sciences experimental designs are typical and data sampling less complex (more 'pure' or 'hard'), labour intensive, and costly (Atkinson et al. 2007). In contrast, methods in developmental science are commonly observational, labour intensive, and subject to less 'pure' measurement. Thus, attendant data collection costs are likely high (Bornstein et al. 2013; Duncan et al. 2014). For example, it is relatively easy to create a 3×3 design of pH and salinity in aquarium water (Torquemada et al. 2005). In comparison finding, sampling, and then manipulating, nine different combinations of hostile parenting and child difficult temperament is complex and likely subject to multiple levels of compromise and bias in measurement decisions. Nonetheless, by using and evaluating optimum design, researchers can evaluate interaction hypotheses with more power and be aware of possible deficiencies in the sample of the *individual* and *environment*. That is, optimum design shows the bivariate distribution best able to test for an interaction effect and allows a comparison of the observed sample to the specified ideal. Sampling with optimum design provides additional power to increase the detection rate of meaningful interaction effects. Comparison to the optimum increases the certainty that an unobserved interaction is trivial. Optimum design highlights that the sample size, magnitude of effect, and the distribution of variables dictate the power to detect interactions.

As an example of optimum design, consider a sample of 1000 with uncorrelated measurements (for simplicity) of parenting and temperament (range 1 to 5, intervals of 1). In the optimum sampling framework, to detect a linear *temperament* × *parenting* interaction a researcher would strive to split the sample of 1000 into four cells (n=250, ¼) at each end of the bivariate distribution (1,1 = 250; 1,5 = 250; 5,5 = 250; 5,1 = 250). This sample maximises the residual variance of *parenting* × *temperament* which minimises the variance of the interaction parameter. Thus, the sample has the most power to detect a *linear* × *linear* interaction (the standard interaction in developmental psychology) and is the optimal design (McClelland 1997; McClelland & Judd 1993). On the other hand, if parenting and temperament are bivariate-normally distributed and a random sample is chosen the efficiency declines to 0.06 of the optimum (McClelland & Judd 1993). That is, 16.7 times more observations (n=16700) would be required to get the same amount of power to detect the interaction. Additionally, although the ¼ split at extremes design can detect linear effects and interactions, such a design is not optimal to evaluate quadratic effects and interactions (McClelland 1997). Testing for quadratic effects is part of the criteria for evaluating *individual* × *environment* models (Roisman et al. 2012). Box 6.1 expands on the requirements for optimum design for quadratic effects and interactions. This shows that additional design considerations are necessary if quadratic effects and interactions are expected and, therefore, describes the challenges of detecting moderation effects.

**Box 6.1:** *Optimum design for quadratic effects and interactions (McClelland 1997).*

Although a ¼ split at the bivariate extremes is optimum for standard *linear* × *linear* interactions, quadratic effects and interactions have different requirements.

To test for quadratic effects, researchers also have to sample individuals from the middle of the distributions. For example, the optimum design to detect a single quadratic effect is ¼, 0, ½, 0, ¼.

Optimum design becomes more complex when quadratic interactions are involved. Specifically, changes to the sampling proportions trade-off the efficiencies to detect linear effects, interactions, quadratic effects, and quadratic interactions. For example, an equal nine point design (1,1; 1,3; 1,5; 3,1; 3,3; 3,5; 5,1; 5,3; 5,5) can detect *linear* × *quadratic* and *quadratic* × *quadratic* interactions, but the relative efficiency for detecting the *linear* × *linear* interaction (e.g., the prototypical differential susceptibility) is only 0.44.

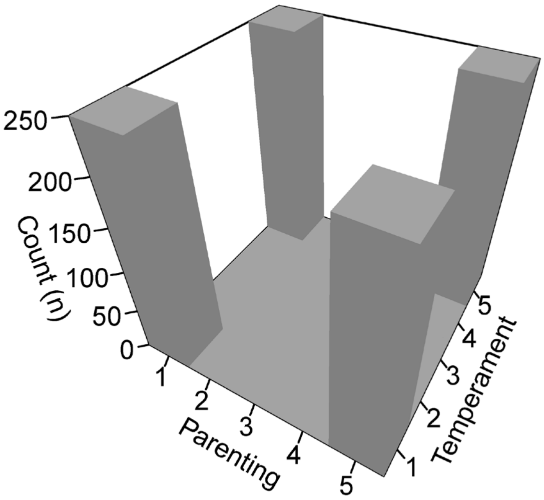
An alternative is a 1,1; 1,5; 3,3; 5,1; 5,5 design that tests both linear and quadratic effects and *linear* × *linear* interaction with equal efficiency of 0.75. A comparison of the predicted and observed mean at the middle (3,3) can then indicate if a higher quadratic trend may be present. Follow up studies could then re-sample to evaluate the higher order interaction effect.

This example shows that evaluating quadratic effects and interactions requires additional considerations and adds complexity to design choices.

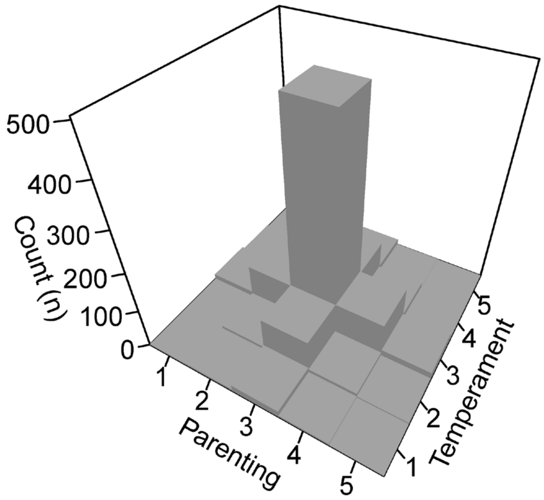
To further illustrate the idea of optimum design, four different bivariate distributions and efficiencies for detecting interactions are presented in Box 6.2. These show how sampling design substantially impacts the ability to test hypothesized *linear* × *linear* interaction effects.

***Box 6.2:*** *Sampling efficiency to detect linear × linear interactions using different designs (A, B, C, and D; McClelland 1997, McClelland & Judd 1993).*

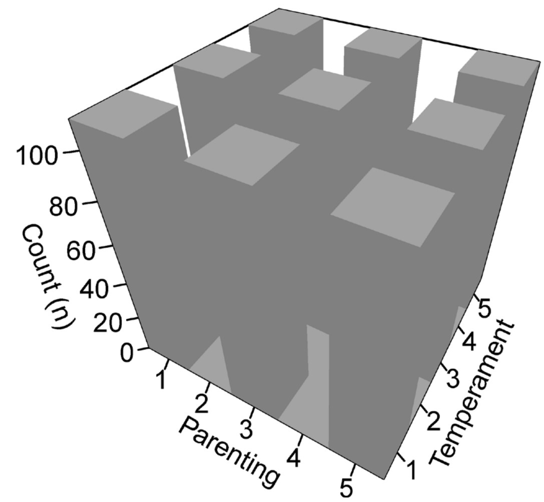
A) 1. Sample of 1000 split into quarters, can test linear effects and linear × linear interactions, cannot evaluate quadratic effects.



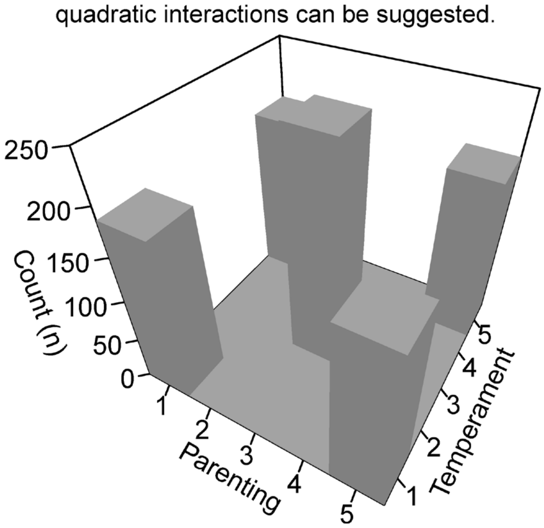

B) 0.042. Sample of 1000, similar to uncorrelated normal distribution (random allocation to each variable approximating 1=0.025; 2=0.135; 3=0.68; 4=0.135; 5=0.025). Very low efficiency to test linear effects (parenting=0.11; temperament=0.119) and linear × linear interactions.




C) 0.445. Equal split of 999 at nine points. Can test higher order interactions such as quadratic × quadratic, but has low efficiency for linear × linear interactions.




D) 0.751. Sample of 998 with ¼ in centre cell, ¾ split equally at the four extreme cells. Can test linear and quadratic effects and linear × linear interactions at 0.751 of optimum. Presence of quadratic interactions can be suggested.




In summary, a practical improvement to detect more interactions is to use optimum design and actively sample from areas of the distributions that increase statistical power. Likewise, in the case of studies that have not specifically employed optimal design strategies at inception, results should be evaluated against optimal design criteria as there is a likelihood of false positives (type I error) and false negatives (type II error). The historical lack of consistent evidence for *individual* × *environment* interactions is not surprising given that the majority of studies did not have optimum design.

### 6.1.2 Improved measurement

One straightforward recommendation to improve the detection of interaction effects is to reduce measurement error. For example, consider environment and moderator measures with acceptable internal reliability (α=0.8). If the interaction between the variables was to be explored, the internal reliability drops to 0.64, the product of the reliabilities (Aiken et al. 1991). Although Cronbach α does not necessarily mean a measure is reliable (Dunn et al. 2014; DeVellis 2016), this example illustrates that interaction effects multiply inconsistencies in measurement. Thus, assuring measurements have low error can improve the power to detect interaction effects (Whisman & McClelland 2005).

Another often repeated recommendation to test *individual* × *environment* theories is to have measurements of the individual, environment, and outcome that represent the positive and negative limits of the distribution (Belsky & Pluess 2009; Rabinowitz & Drabick 2017; Roisman et al. 2012). This recommendation exists because some theoretical models cannot be adequately tested if a measure of the individual, environment, or outcome has ceiling or floor effects. For example, differential susceptibility theory assumes that the outcome can be more positive in a high quality environment. If the outcome measure has a ceiling effect and therefore does not capture higher performance the theory cannot be evaluated. The challenge to evaluating and implementing this recommendation is that common psychometric techniques do not demonstrate whether a multi-item measure can adequately distinguish low and high performance (van der Linden 2018) or capture the range of performance. More specifically, the approach documented in the review of papers, presented in Chapter 3, was to calculate internal reliability using inter-item correlation (e.g., Cronbach's (1951) alpha) and/or perform an exploratory and confirmatory factor analysis. Though these analyses are an index of internal-reliability, they do not illustrate the ability of a scale to distinguish low and high performance.

Item response theory models can demonstrate how well items on a scale distinguish low and high levels of performance (van der Linden 2018). Specifically, item response theory models incorporate parameters for each item and the characteristics of people scoring the item (Baker & Kim 2017). Because of this, each item can be allocated a 'difficulty' and the ability of the item to distinguish different scores on the scale can be calculated (Gordon 2015). That is, the ability of an item to distinguish low and high performance can be evaluated. This contrasts to classical test theory which assumes items replicate each other and can be summed or averaged to reduce error (van Alphen et al. 1994). As with any statistical method, there are assumptions of item response theory models that can be problematic. For example, item response theory assumes unidimensionality in which one dimension is being measured per item. Such an assumption may be unrealistic in the measurement of temperament (e.g., Gartsein & Rothbart 2003). Item response theory also generally assumes local independence in which the answer on one item is independent of answer on another item, even though they may represent a latent trait (Edwards et al. 2018). If these assumptions can be reasonably met or adjusted for, item response theory can be used to understand if a scale measures low and high performance (Gordon 2015).

In summary, a recommendation for future research, therefore, is to use item response theory models to better understand the ability of their multi-item measures to distinguish low and high performance. For example, Box 6.3 illustrates an item response theory model evaluating an outcome measure used in tests of *individual* × *environment* theories. Using item response theory models can thus help researchers establish if their measurements are likely to adequately measure low and high performance.

There is, however, some complexity in that measure validation requires a normative sample, whilst optimum design seeks to sample from pre-specified ranges (Blinkhorn 1997). Researchers will be challenged to determine the available range of a measure and sample from parts of that measure at the same time. Pilot studies may therefore be necessary to first validate measures. Also, utilising existing data to find measures that have desirable ranges may help select variables to include in an upcoming study.

***Box 6.3:*** *Example of an item response theory model: Strengths and Difficulties Questionnaire*

Item response theory models can be used to estimate item and person parameters of a scale which demonstrate how well items distinguish people with different scores on the overall scale.

In this example of item response theory, the Strengths and Difficulties Questionnaire, the behaviour outcome measure used in Chapter 4 and 5, is evaluated using a partial credit model estimated via conditional marginal likelihood with the eRm package (version 1.0-0; Mair et al. 2019; Mair & Hatzinger 2007) in R (R Core Team 2018). The partial credit model is used as an example because it is the simplest (least complex) method when the rating scale is polytomous (ordered categories). See van der Linden (2018) for additional modelling options and considerations outside the scope of this example.

Specifically, Figure 6.2 shows a person-item map of conduct problems reported in the Longitudinal Study of Australian Children (see Chapter 5 for sample selection). The upper panel shows the person parameter distribution, a score for each child on the scale. It has an observed mean of -1.24 and standard deviation of 1.24. The lower panel shows the item thresholds and overall item location (or difficulty) on the standardised item logit scale. This demonstrates the person mean is a full logit below the scale mean (0). The thresholds indicate when the probability of endorsing an item with a particular score (e.g., 0=not true, 1=somewhat true, or 2=certainly true) changes as a function of the overall score. For example, up until -1.86 on the overall scale parents were more likely to rate their child a 0 for the item *often has temper tantrums or hot tempers*, they were then more likely to rate the child a 1 up until overall scores of 0.95 where they were more likely to rate their child a 2 for that item. If there are no item thresholds that cross an overall score, it indicates the measure is unable to distinguish differences in performance at that score e.g., the end-left and end-right bars in the histogram. Additionally, scores of 0 and 10, the minimum and maximum, are excluded as they do not contribute any information to the ability of items to discriminate between scores.

As can be seen in Figure 6.2, many children score below the mean (1515, 42.1%) and there are no items with thresholds that distinguish the large number of children in the lowest histogram category (799, 22.2%). Further to this, 729 (20.2%) children scored a complete 0 and 7 (>1%) scored a complete 10. The items cannot discriminate between these cases and they are not included in the histogram. Instead, most items distinguish children with higher conduct problems (7 of 10 thresholds 1SD higher than the person mean). These issues are not a fault of the measure as it was designed to screen for clinical behaviour disorders and not to measure exceptional conduct (Goodman 1997). However, it does indicate the measure is inadequate for determining low levels of conduct problems.

This example thus illustrates that the conduct problem measure might be problematic to test differential susceptibility models where low and high scores are expected. A better measure for testing differential susceptibility using behaviour would be able to distinguish both low and high scores (e.g., Davidson et al. 2018).

**Box 6.3:** *Example of an item response theory model continued*

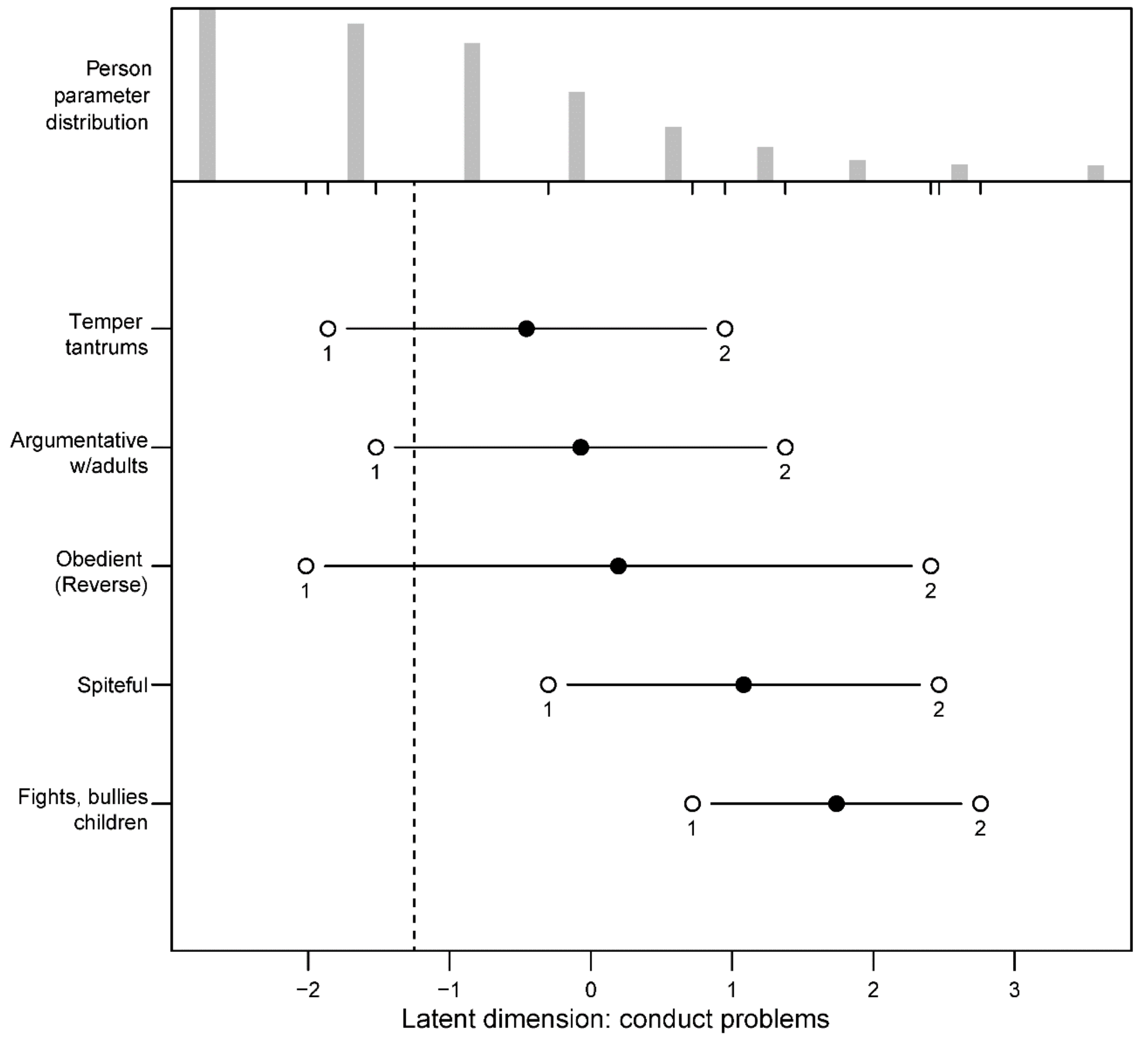



**Figure 6.2:** Partial credit model item-probability map for the conduct problems subscale of the Strengths and Difficulties Questionnaire collected for 3062 children from the Longitudinal Study of Australian Children (wave 3, birth cohort). Open circles are thresholds, solid circles are the item locations, dashed line is the person parameter average (-1.24), and histogram tick marks indicate thresholds. Note children with a raw score of 0 (-3.82, n=729) and 10 (4.6, n=7) are not included as their scores cannot provide item discrimination information.

### 6.1.3 Use (more) causal designs

The use of longitudinal, quasi-experimental, and randomised experimental designs instead of cross-sectional designs, though not a solution to measurement and sample issues, is another way to enhance the quality of research on *individual* × *environment* interactions (Campbell & Stanley 2015; Mitchell et al. 2013). By being able to control exposure to an intervention or evaluate exposure to the environment at a within-person level researchers can evaluate how the same children, with different sensitivities, respond to changes in the environment across time (Moore & Depue 2016). This approach can reduce endogeneity (correlation between the covariates and error in the model), confounding, and sampling bias, improve internal validity, and provide stronger causal evidence (Angrist & Pischke 2008; Bailey et al. 2018; Barker & Milivojevich 2016; Card & Little 2007; Dunning 2008; Gelman & Hill 2006; Rubin 1974; Rubin 2008). The drawback of longitudinal, quasi-experimental, and randomised experimental approaches is the logistical feasibility of implementing, validating, and collecting data. Furthermore, there is no guarantee of elucidating causal effects and the study may lack external validity if the sample does not generalise to the population of interest. However, the inferential ability is substantially greater than the use of cross-sectional data (Evans 2003; Hoppe et al. 2009; Merlin et al. 2009). Therefore, when possible, using experimental and longitudinal designs can increase the quality of evidence for *individual* × *environment* interactions.

### 6.1.4 Run simulation studies to understand design implications

Another improvement for identifying *individual* × *environment* moderating effects would be to undertake simulation studies to illustrate how choices in research design and analysis can influence the results before collecting data or after pilot studies (Morris et al. 2019; Wu et al. 2015). For example, this thesis qualitatively compared the ability to detect interactions using observations of parenting and surveys of parenting. An alternative to this qualitative comparison would be to simulate data where measurement error is reduced or increased by measurement modality and quantify the relative ability to estimate the interaction effect (Jaccard & Wan 1995). Likewise, this thesis qualitatively evaluated if temperament assessed earlier detected more interactions than temperament assessed later. A simulation study could prescribe different rates of change in temperament to determine at what age an interaction effect would be become impractical to detect (Rast & Hofer 2014). Simulations such as these can be used to understand how changes to research design can influence the ability to detect interactions. If this understanding is illuminated before full data collection researchers can make adjustments to improve power (Moskowitz et al. 2017; Timmons & Preacher 2015). Whilst simulations do not provide the whole picture, they can illustrate features of the statistical landscape that may be particularly important (Morris et al. 2019; e.g., the bivariate distribution of the moderator and environment variables, Shieh 2009). Thus, using simulation studies to guide research design is another way to improve the ability to detect *individual* × *environment* interactions.

## 6.2 Analysis

### 6.2.1 Use pre-existing guidelines for exploring interactions

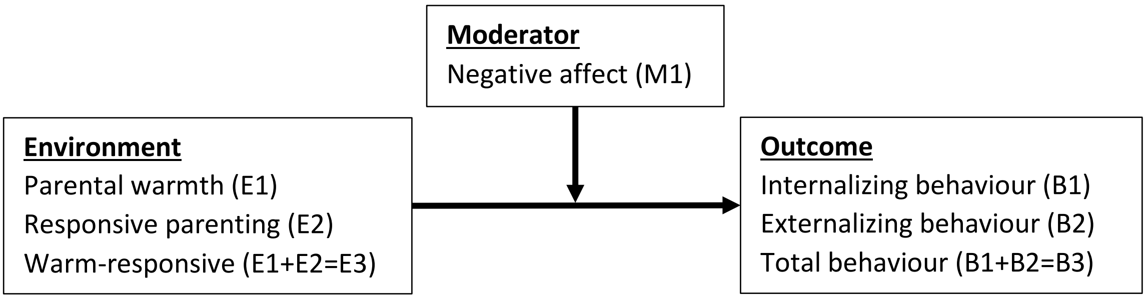
Researchers should justify the robustness of their results in light of the recommendations of well-developed guidelines to assess interaction effects. Chapter 2 provides an overview and explanation of these criteria in detail, but they are summarised in Table 6.1. For example, regions of significance testing illustrate the values at which significant moderation effects occur, whilst the location of cross-over points and proportion of values above and below these points indicate the amount of evidence for different theoretical propositions. There are slight variations on these criteria that can be used to evaluate interactions (Belsky et al. 2013; Del Giudice 2017; McCabe et al. 2018; Widaman et al. 2012) and the value of different criteria remain subject to debate. For example, some literature considers within-person effects more robust than between-person effects (Gelman & Hill 2006). Nonetheless, analytic guidelines provide substantive tests of interactions and inform interpretation of results.

**Table 6.1:** Criteria of Roisman et al. 2012 to explore an interaction.

| Criteria | General explanation |
|---|---|
| Regions of significance (Dearing & Hamilton 2006) | Regions of significance on the environment test shows that the outcome (e.g., child behaviour) and moderator (e.g., temperament) are correlated at high and low ends of the distribution of the environment (e.g., parenting). Tests for contrastive effects also examine regions of significance on the moderator. It is recommended that a significant correlation should be observed within a range ±2 SD from the mean of the environment/moderator. |
| Proportion of interaction and proportion affected | When predicted values crossover (e.g., prediction for low temperament group vs high temperament group given parenting) two metrics can be calculated to indicate how well the sample and interaction support different theoretical models. First, the proportion of interaction indicates the relative amount of space (±2 SD from the mean) above and below the cross-over point. Values of 0.5, for example, are indicative of differential susceptibility, whilst values close to 0.0 or 1.0 suggest diathesis stress or vantage sensitivity. Second, the proportion affected calculates how much of the sample is above and below the crossover point. A greater proportion on each side of the cross-over is suggestive of differential susceptibility. |
| Adjust for false positives (Benjamini & Hochberg 1995) | Because multiple interactions may be tested (e.g., using slightly different measures or different ages) a method to adjust *p-value* cut-offs should be used to reduce false-positives (Type I error). These adjustments make it less likely that chance interactions are detected. |
| Evaluate average effect over time | When possible, instead of only examining interactions separately at each age, between person effects (average across time) should be estimated using multilevel models. |
| Quadratic environment and quadratic environment ×moderator | Interaction estimate may mask quadratic effects and quadratic interactions. Quadratic effects and interactions should thus be tested to rule out alternative explanations. |

### 6.2.2 Systematic exploration of analytical choices (sensitivity analysis)

The multiverse analytical plan (Steegen et al. 2016) utilised in this thesis was valuable to illustrate how different analytical choices could change the observed interactions. A multiverse analysis, which is a form of statistical sensitivity analysis (French 2003), proceeds by examining how all (or most) of the different analytical choices a researcher could make may affect the results. For example, instead of only using a composite of two parenting and behaviour subscales to test a *temperament* × *parenting* interaction, the composite and subscales would be assessed and reported on. A comparison of the standard and multiverse design is contrasted in Figure 6.3. This hypothetical example shows a standard design would indicate an interaction between negative affect and warm-responsive parenting predicting total internalizing and externalizing behaviour. In contrast, the multiverse design shows the aforementioned interaction and also that the parenting warmth component was a non-significant predictor in the moderation model. Additionally, it shows that internalizing behaviour on its own was not predicted by any interactions between parenting and negative affect. Thus, the multiverse can highlight how seemingly innocent or justified choices (e.g., using one behaviour or parenting measure over another) affect (or do not affect) the number and consistency of interactions detected.




| Standard composite design | |
|---|---|
| | Negative affect |
| | Total behaviour |
| Warm-responsive | B3 = M1*E3 ✓ |

| Multiverse design | | | |
|---|---|---|---|
| | Negative affect | Negative affect | Negative affect |
| | Internalizing | Externalizing | Total |
| Warmth | B1 = M1*E1 ✕ | B2 = M1*E1 ✕ | B3 = M1*E1 ✕ |
| Responsive | B1 = M1*E2 ✕ | B2 = M1*E2 ✓ | B3 = M1*E2 ✓ |
| Warm-responsive | B1 = M1*E3 ✕ | B2 = M1*E3 ✓ | B3 = M1*E3 ✓ |

**Figure 6.3:** Comparison of standard composite design and multiverse design to evaluate moderation model. In green cells with a ✓ there is an interaction between the moderator and environment predicting behaviour, whilst in blue cells with a ✕ there is no interaction.

In summary, researchers should be aware of the multiple analytical choices they had available and systematically model how sensitive their results are (Simmons et al. 2011). This statistical sensitivity testing will help identify consistent evidence and outline the boundaries of evidence for interactions.

## 6.3 Interpretation of results

### 6.3.1 Critical assessment to improve the quality of evidence

A possible future research improvement would be to increase the burden-of-proof for researchers claiming interaction effects. Specifically, as *individual* × *environment* interactions are a developmental principle, or scientific paradigm (Kuhn 2012), there may be a tendency to accept weak evidence that confirms what is presumed to be true (formally known as confirmation bias; Nickerson 1998). That is, interactions are accepted as evidence, even if they come from low powered designs and sporadically vary in direction, because they do not challenge what is assumed to be true.

Although accepting low-quality evidence for interactions may not re-write developmental paradigms, there is a risk that an accumulation of low quality evidence may mislead ongoing funding decisions and research enterprise (Dick et al. 2015; Freedman et al. 2015; Munafò et al. 2017). Specifically, when interactions produce unanticipated results (e.g., two-way contrastive effects; Slagt et al. 2018; Stocker et al. 2017) these results are dismissed given the impracticality of an interpretation. However, by accepting inconsistent, yet explainable evidence the research field may undertake research that is not replicable and of little empirical value. For example, Luijk et al. (2011) and Roisman et al. (2013) investigated the association between genetic-polymorphisms and infant attachment security based on initial findings from small, low-powered studies. Their comparatively larger power revealed the initial effects to be trivial and thus absent of empirical utility. It is in the interests of the research community, and those in policy and practice who take up these results, to accumulate consistent and high quality evidence.

In summary, researchers should be more critical of research findings when evidence is inconsistent and sporadic. Accepting that low power from non-optimal design and measurement is one reason *interaction effects are ubiquitous in nature*, [yet] *so difficult to detect* (Ellis et al. 2011; McClelland & Judd 1993; Mitchell et al. 2013; Wachs & Plomin 1991) the onus is on researchers to increase the burden-of-proof. High quality and consistent evidence that confirms, or refutes, evidence for *individual* × *environment* interactions (LeBel et al. 2018; Nosek et al. 2015) is required. This necessitates attention to design, measurement, analytic consistency, and analytic results.

### 6.4 Conclusion

This chapter has outlined seven approaches future research may be able to implement as a means of improving the consistency and quality of evidence for *individual* × *environment* interactions: using more optimum design, improving measurement, using more causal designs, using simulation studies to understand implications of design, using existing guidelines for evaluating interactions, systematically exploring analytical choices, and critically assessing and accurately portraying results. Chapter 7 concludes the thesis by summarizing the results and outlining implications for policy and practice.

## Chapter 7: Integrative discussion

### 7.1 Summary of motivation and aims of thesis

Extensive evidence documents that individuals, both human and non-human, exhibit differences in response to the same experience and, further, that these experiences can shape ongoing differences in response (Aron et al. 2012; Boyce 2016; Ellis et al. 2011). In developmental science, such evidence has given rise to *individual* × *environment* models that hypothesise that individual child characteristics moderate the effects of early experiences on developmental outcomes. However, testing of *individual* × *environment* models of development has yielded disappointingly inconsistent results showing extensive variation within and between studies (Rabinowitz & Drabick 2017). This variation relates to the presence or absence of a moderating effect, and, if present, the nature of that moderating effect. Across and within studies there is evidence for the predictions of the three dominant *individual* × *environment* theoretical models: diathesis stress (Zuckerman 1991), differential susceptibility (Belsky 1997; Ellis et al. 2011), and vantage sensitivity (Pluess & Belsky 2013). However, the evidence is sporadic. The aim of this thesis was to ask "Why?" and investigate possible methodological explanations.

This thesis sought to understand the gap between sound, biologically-based theoretical models and weak and inconsistent empirical evidence in the study of child development. Specifically, the thesis systematically explored existing and new evidence for child characteristics moderating the effects of parenting on behavioural development to identify measurement, design and analytic factors systematically associated with support for the theoretical models. The underlying hypothesis was that complexity in child development, alongside the infeasibility of controlling for multiple sources of variation in research studies may explain the unexpected gap. The aim was to test this hypothesis through intense systematic analysis.

At twofold approach was taken in addressing this aim. First, systematic review and analysis of the literature was undertaken (Chapters 2 and 3). The literature review examined variation in theory-testing focused on measurement and analytic approaches. A range of genetic, biological, and observational measures of child characteristics have been applied in the empirical literature to define child characteristics. Parenting and behavioural outcomes have also operationalised by employing a range of different measures. Second, empirical analyses that systematically tested measurement and analytic variation within and across studies (Chapters 4 and 5) were undertaken. The substantive focus was the role of child characteristics in moderating the association of parenting environment and child behavioural development. Empirical analysis of the Longitudinal Study of Australian Children, presented in Chapter 4, investigated how childhood temperament moderated the effects of parenting on parent and teacher reported behavioural development across ages 4 to 8. This analysis included a systematic exploration of 512 model specifications. This systematic analysis was extended in Chapter 5 in further analysis of the Longitudinal Study of Australian Children. Specifically, measurement of temperament in childhood was substituted with measurement in infancy, theoretically a less environmentally contaminated measure of genetic predisposition (Bornstein et al. 2019). Additionally, through comparison with the Family Life Project (United States of America) systematic evaluation of analytical variations were examined. These included child age at assessment; form and extent of temperament assessment (short vs full-item); self-reported parenting compared to observed parenting; and frequency-distribution and ranges of behaviour and parenting data. The empirical findings, reflecting those in the literature review, found sporadic and inconsistent outcomes.

The following discussion presents a summary of key outcomes, directions, and recommendations for future research, and implications of current research for application in child, family, and educational policy and practice.

**7.2 Summary of key findings**

Systematic analysis of the current literature: Chapter 2 and 3 reviewed studies that have applied *individual* × *environment* methodology to understand how child characteristics moderate the effects of parenting on behavioural development. The review of the literature quantified the measurement and statistical methodology associated with findings consistent with diathesis stress, differential susceptibility and vantage sensitivity. The systematic review in Chapter 3 examined 542 *individual* × *parenting* interactions and 86 (16%) were described as statistically reliable. However, within- and between-studies, these reliable interactions inconsistently supported different theoretical models and sporadically varied with measures of parenting, behaviour, and individual characteristics. This inconsistency within and between extant studies directed empirical investigation. The investigation involved systematic selection and testing of measures in *individual* × *parenting* interactions to identify robust areas of evidence and sources of variation.

Systematic empirical analyses: Chapters 4 and 5 thus presented sequential empirical analyses of two longitudinal cohort studies to systematically evaluate multiple analytical choices with the aim of identifying consistent evidence of *temperament* × *parenting* interactions predicting behavioural development. In an analysis of the Longitudinal Study of Australian Children, 13 (2.5%) of 512 interactions were statistically reliable. In further analysis of the Longitudinal Study of Australian Children and comparison with the Family Life Project, two (3%) of 64, and none (0%) of 1620 interactions were statistically reliable. These few interactions were very sporadic and inconsistent across measures of behaviour, temperament, and parenting. As such, the systematic empirical analyses indicate that systematic changes to the reporting source (parents and teachers), the measure of behaviour and parenting, the method of measuring parenting (observation and survey), child age at assessment, and the measure of temperament do not improve the identification of consistent interactions. The systematic empirical exploration demonstrated the importance of analytical decisions and testing the consistency and validity of moderating effects.

### 7.2.1 Key outcomes

The thesis identifies that;

1) The evidence for *individual* × *parenting* and *temperament* × *parenting* interactions predicting behavioural development is sporadic and inconsistent.
2) Systematic changes to measures of behaviour, temperament, and parenting may not achieve more consistent evidence for interactions.
3) Detailed exploration of analytical choices is needed to establish the consistency of interactions.

In combination, the three outcomes converge on the conclusion that *if the **past** approach to individual × parenting interactions remains the **future** approach, researchers are unlikely to find more consistent evidence.*

## 7.3 Implications for research

To increase the consistency and quality of evidence for individual environment interactions researchers will need to change their practices. Seven suggested improvements for research were outlined in Chapter 6. These include using more optimum design, improving measurement, using more causal designs, using simulation studies to understand implications of design, using existing guidelines for evaluating interactions, systematically exploring analytical choices, and critically assessing and accurately portraying results. Researchers are more likely to find consistent evidence if they can improve the quality of design, measurement, and analysis.

## 7.4 Implications for policy and practice

For parents, professionals working with children, and policymakers a fundamental question in supporting children's development relates to the allocation of the finite resources of time, effort, and money. Parents, in making decisions about whether to engage in paid work in the early years of their child's life, for example, must assess the relative impact of family income against the experiences that can be provided for their child in home and childcare settings. Policymakers, similarly, must judge whether to provide targeted provisions for those who stand to gain most or provision broad reaching universal services that avert the potential for stigma (Diderichsen et al. 2019). More recently, the acknowledgement that the same environments are experienced differently by individual children (Boyce 2016) has moved focus in design and translation of developmental science from asking “what works?” to asking ‘what works for whom?” (Belsky & van IJzendoorn 2015). Such approaches present the opportunity for more accurate targeting of resources (Bakermans-Kranenburg & Van Ijzendoorn 2015; Velderman et al. 2006) or , alternatively, an approach in which children who are more reactive to environment function as barometers of environmental quality for all children (Davis et al. 2012). The finding of this thesis, however, suggest that the evidence is yet to deliver these applications. *Individual* × *environment* interaction models to assess outcome effects have failed to provide sufficiently robust evidence to identify a broadly applicable theory and have, instead, directed attention to pragmatic approaches in which individual response to environments, or manipulations of environments (interventions) are directly observed and adjusted to meet individual need.

Applying pragmatic approaches, three key lessons and implications for application emerge from the current status of empirical research. These relate to exposure to the environment, individualised effect of the environment, and evaluation and adjustment of the environment;

1. *Exposure: Individual* × *environment* models powerfully show that children are active agents in their exposure to environments, whether directly or through the mechanism of parent choice. They highlight that individual characteristics not only influence the environment (Kiff et al. 2011), but can direct the type of environments an individual self-selects into (Plomin & Bergeman 1991; Wiggins et al. 2014). While family and social characteristics influence the quality of developmental experience, so too does the preferences and behaviours of a child. Davis (2014), for example, reports that children with more difficult temperament enter early education and care later then less difficult counterparts, but are placed in higher quality provision. Such decisions relate to family capacity for choice, but also the individual child’s characteristics and parent understanding of child need. These findings direct attention to understanding not only type of exposure, but whether there are thresholds of exposure that are required to support positive outcomes (Early Child Care Research Network 2003; Howes et al. 1992). Interventions, whether at an individual (e.g. clinical intervention), group (targeted policies), or whole population level need to consider thresholds of exposure as well as type. Such considerations have already been voiced in the field of public health (Diderichsen et al. 2019).

2. *Effect*: *Individual* × *environment* models have been powerful in demonstrating that not all people encounter an environment in the same way (Boyce 2016). In the absence of pure measures

of ‘sensitivity’, ‘reactivity’, or ‘vulnerability’ to the environment new bottom-up, *insitu* measures are emerging that focus on the child in environmental contexts (Greven et al. 2019). For example, the Individualized Classroom Assessment Scoring System (inCLASS) is an observational measure of the child within the early education and care environment that aims to understand the child’s individual experience of early education (Downer et al. 2010). Similarly, Slagt et al. (2019), instead of pre-defining a measure of reactivity, assessed and defined child reactivity to parenting though direct observation of individual difference in parent-child interaction. Such *insitu* measurement approaches may well deliver improved understanding of effective practice for children. That is, understanding specific effects for specific children for specific reasons can guide practice to better meet the needs of the individual (Moore & Depue 2016).

3. *Environmental adjustment and evaluation*: The evidence base for *individual* × *environment* models highlights the need for pragmatic approaches where practitioners and policy-makers recognise individual differences in response to the environment and attend not only to average outcomes, but to individual response profiles (DeRubeis et al. 2014). This is akin to clinical cycles where interventions are applied, individual responses are observed, and the clinician learns and adjusts the intervention to meet the needs of the individual (Shimokawa et al. 2010; Spring 2007). By continually learning, and adjusting to, the needs of individual children (e.g., responsive parenting; Landry et al. 2006), more equitable care may be provided to support positive developmental (Fuchs & Fuchs 2019; Kristal 2005).

In summary, although the evidence for *individual* × *environment* models falls short of definitive policy and practice guidance, it demonstrates the importance of understanding and working with individual differences. Pragmatic, bottom-up approaches to understand how individuals respond differently to developmental experiences represents a foundational point for policy and practice while efforts to deliver robust, top-down evidence proceed.

### 7.5 Concluding statement

This thesis systematically reviewed and empirically tested evidence for *individual* × *environment* interaction effects and theoretical models. The specific context was individual child characteristics, including temperament, moderating the effects of parenting on behavioural development in childhood. The thesis demonstrated that evidence for *individual* × *parenting* interactions is inconsistent and sporadic. Systematically exploring analytical decisions and the myriad of measurements did not illuminate consistency or clarify differential findings. Instead, the thesis suggested that a variety of ‘moving parts’ coupled with non-optimum design and low quality measurement has hindered the evaluation of *individual* × *parenting* interaction effects. Future research should ensure use of more optimum and causal designs, undertake simulation studies to understand implications of design, have measurements with low error and sufficient variability, use established statistical criteria to evaluate interactions, systematically explore analytical decisions to ensure robustness, and critically evaluate research findings that are inconsistent, sporadic, and use low powered, non-optimum samples. Researchers have long proposed individual characteristics interact with the environment to cause differential child outcomes. Well-designed evaluation of *individual* × *environment* interactions can light the way to understanding what works, and what works for whom.

**List of References**

*included in systematic review, Chapter 3

Achenbach, T. M. (1991). Manual for the Child Behavior Checklist/4-18 and 1991 profile. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin, 85(6), 1275-1301. doi:10.1037/0033-2909.85.6.1275

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101(2), 213-232.

Aiken, L. S., West, S. G., & Reno, R. R. (1991). Multiple regression: Testing and interpreting interactions. Thousand Oaks, CA, US: Sage Publications, Inc.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. N. (1978). Patterns of attachment: A psychological study of the strange situation: Psychology Press.

Alexander, N., Osinsky, R., Schmitz, A., Mueller, E., Kuepper, Y., & Hennig, J. (2010). The BDNF Val66Met polymorphism affects HPA-axis reactivity to acute stress. Psychoneuroendocrinology, 35(6), 949-953. doi:10.1016/j.psyneuen.2009.12.008

Allison, P. D. (2009). Fixed effects regression models (Vol. 160): SAGE publications.

An, X., Curby, T. W., & Brock, L. L. (2018). Is the Child Really What's Being Rated? Sources of Variance in Teacher Ratings of Socioemotional Skills. Journal of Psychoeducational Assessment, doi:10.1177/0734282918808618

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Angus, L. (2015). School choice: neoliberal education policy and imagined futures. British Journal of Sociology of Education, 36(3), 395-413. doi:10.1080/01425692.2013.823835

Appel, K., Schwahn, C., Mahler, J., Schulz, A., Spitzer, C., Fenske, K., . . . Grabe, H. J. (2011). Moderation of adult depression by a polymorphism in the FKBP5 gene and childhood physical abuse in the general population. Neuropsychopharmacology, 36(10), 1982-1991. doi:10.1038/npp.2011.81

Archer, J., Graham-Kevan, N., & Davies, M. (2005). Testosterone and aggression: A reanalysis of book, Starzyk, and Quinsey's (2001) study. Aggression and Violent Behavior, 10(2), 241-261. doi:10.1016/j.avb.2004.01.001

Arcos-Burgos, M., Velez, J. I., Solomon, B. D., & Muenke, M. (2012). A common genetic network underlies substance use disorders and disruptive or externalizing disorders. Human Genetics, 131(6), 917-929. doi:10.1007/s00439-012-1164-4

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. Journal of Personality and Social Psychology, 73(2), 345-368. doi:10.1037/0022-3514.73.2.345

Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory Processing Sensitivity: A Review in the Light of the Evolution of Biological Responsivity. Personality and Social Psychology Review, 16(3), 262-282. doi:10.1177/1088868311434213

Atkinson, A., Donev, A., & Tobias, R. (2007). Optimum experimental designs, with SAS (Vol. 34). Oxford University Press.

Australian Bureau of Statistics (2018a). 5518.0.55.001 – Government Finance Statistics, Education, Australia, 2016-2017.

Australian Bureau of Statistics (2018b). Socio-Economic Indexes for Areas. Census; data and analysis. Available from: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa

Avinun, R., & Knafo, A. (2014). Parenting as a reaction evoked by children's genotype: A meta-analysis of children-as-twins studies. Personality and Social Psychology Review, 18(1), 87-102. doi:10.1177/1088868313498308

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D. H., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. American Psychologist, 73(1), 81-94.

Baker, F. B., & Kim, S. H. (2017). The basics of item response theory using R. New York, NY: Springer.

Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2011). Differential susceptibility to rearing environment depending on dopamine-related genes: New evidence and a meta-analysis. Development and Psychopathology, 23(1), 39-52. doi:10.1017/S0954579410000635

Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2014). A sociability gene? Meta-analysis of oxytocin receptor genotype effects in humans. Psychiatric Genetics, 24(2), 45-51. doi:10.1097/YPG.0b013e3283643684

Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2015). The hidden efficacy of interventions: Gene× environment experiments from a differential susceptibility perspective. Annual Review of Psychology, 66, 381-409.

Banaschewski, T., Becker, K., Scherag, S., Franke, B., & Coghill, D. (2010). Molecular genetics of attention-deficit/hyperactivity disorder: an overview. European Child & Adolescent Psychiatry, 19(3), 237-257. doi:10.1007/s00787-010-0090-z

Barber, B. K., & Harmon, E. L. (2002). Violating the self: Parental psychological control of children and adolescents. In Intrusive parenting: How psychological control affects children and adolescents. (pp. 15-52). Washington, DC, US: American Psychological Association.

Barker, E. D., Boivin, M., Brendgen, M., Fontaine, N., Arseneault, L., Vitaro, F., ... & Tremblay, R. E. (2008). Predictive validity and early predictors of peer-victimization trajectories in preschool. Archives of General Psychiatry, 65(10), 1185-1192.

Barker, T. B., & Milivojevich, A. (2016). Quality by experimental design. Chapman and Hall/CRC.

Barnard, J. and Rubin, D.B. (1999). Small sample degrees of freedom with multiple imputation. Biometrika, 86, 948-955.

Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., & Alzheimer's Disease Neuroimaging Initiative. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. Statistical Methods in Medical Research, 24(4), 462-487. doi:10.1177/0962280214521348

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Bauer, A. M., Quas, J. A., & Boyce, W. T. (2002). Associations between physiological reactivity and children's behavior: Advantages of a multisystem approach. Journal of Developmental & Behavioral Pediatrics, 23(2), 102-113.

Baumrind, D. (1971). Current Patterns of Parental Authority. Developmental Psychology, 4(1), 1-103. Doi:10.1037/h0030372

Bayer, J. K., Ukoumunne, O. C., Lucas, N., Wake, M., Scalzo, K., & Nicholson, J. M. (2011). Risk factors for childhood mental health symptoms: national longitudinal study of Australian children. Pediatrics, 128(4), e865-e879.

Beauchaine, T. (2001). Vagal tone, development, and Gray's motivational theory: toward an integrated model of autonomic nervous system functioning in psychopathology. Development and Psychopathology, 13(2), 183-214. doi:undefined

Behar, L., & Stringfield, S. (1974). Manual for the Preschool Behavior Questionnaire.

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. Psychological Review, 75(2), 81-95.

Belsky, J. (1997). Theory testing, effect-size evaluation, and differential susceptibility to rearing influence: The case of mothering and attachment. Child Development, 68(4), 598-600. doi:10.1111/j.1467-8624.1997.tb04221.x

Belsky, J. (2005). Differential susceptibility to rearing influence. Origins of the social mind: Evolutionary psychology and child development, 139-163.

Belsky, J. A. Y., Hsieh, K.-H., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3 years: Differential susceptibility to rearing experience? Development and Psychopathology, 10(2), 301-319. doi:undefined

Belsky, J., & Fearon, R. M. (2002). Early attachment security, subsequent maternal sensitivity, and later child development: does continuity in development depend upon continuity of caregiving? Attachment & Human Development, 4(3), 361-387. doi:10.1080/14616730210167267

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: differential susceptibility to environmental influences. Psychological Bulletin, 135(6), 885-908. doi:10.1037/a0017376

Belsky, J., & van Ijzendoorn, M. H. (2015). What works for whom? Genetic moderation of intervention efficacy. Development and Psychopathology, 27(01), 1-6.

Belsky, J., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. Current Directions in Psychological Science, 16(6), 300-304. doi:10.1111/j.1467-8721.2007.00525.x

*Belsky, J., Newman, D. A., Widaman, K. F., Rodkin, P., Pluess, M., Fraley, R. C., . . . Roisman, G. I. (2015). Differential susceptibility to effects of maternal sensitivity? A study of candidate plasticity genes. Development and Psychopathology, 27(3), 725-746. doi:10.1017/S0954579414000844

Belsky, J., Pluess, M., & Widaman, K. F. (2013). Confirmatory and competitive evaluation of alternative gene-environment interaction hypotheses. Journal of Child Psychology and Psychiatry, 54(10), 1135-1143. doi:10.1111/jcpp.12075

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological, 57(1), 289-300.

Bennett, D. S., Sullivan, M. W., & Lewis, M. (2006). Relations of Parental Report and Observation of Parenting to Maltreatment History. Child Maltreatment, 11(1), 63-75. doi:10.1177/1077559505283589

Berger, M. P., & Wong, W. K. (2009). An introduction to optimal designs for social and biomedical research (Vol. 83). John Wiley & Sons.

Berntson, G. G., Bigger, J. T., Jr., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., . . . van der Molen, M. W. (1997). Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology, 34(6), 623-648.

Bertoletti, E., Zanoni, A., Giorda, R., & Battaglia, M. (2012). Influence of the OPRM1 gene polymorphism upon children's degree of withdrawal and brain activation in response to facial expressions. Developmental Cognitive Neuroscience, 2(1), 103-109. doi:10.1016/j.dcn.2011.05.001

Binder, E. B. (2009). The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders. Psychoneuroendocrinology, 34 Suppl 1, S186-195. doi:10.1016/j.psyneuen.2009.05.021

Bindman, S. W., Pomerantz, E. M., & Roisman, G. I. (2015). Do children’s executive functions account for associations between early autonomy-supportive parenting and achievement through high school? Journal of Educational Psychology, 107(3), 756-770.

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., . . . Grantham-McGregor, S. (2017). Early childhood development coming of age: science through the life course. The Lancet, 389(10064), 77-90. doi:https://doi.org/10.1016/S0140-6736(16)31389-7

Blair, B. L., Perry, N. B., O'Brien, M., Calkins, S. D., Keane, S. P., & Shanahan, L. (2015). Identifying developmental cascades among differentiated dimensions of social competence and emotion regulation. Developmental Psychology, 51(8), 1062.

Blair, C., Granger, D. A., Kivlighan, K. T., Mills-Koonce, R., Willoughby, M., Greenberg, M. T., . . . Family Life Project, I. (2008). Maternal and child contributions to cortisol response to emotional arousal in young children from low-income, rural communities. Developmental Psychology, 44(4), 1095-1109. doi:10.1037/0012-1649.44.4.1095

Blair, C., Granger, D., & Peters Razza, R. (2005). Cortisol reactivity is positively related to executive function in preschool children attending head start. Child Development, 76(3), 554-567. doi:10.1111/j.1467-8624.2005.00863.x

Blandon, A. Y., Calkins, S. D., Keane, S. P., & O'Brien, M. (2010). Contributions of child's physiology and maternal behavior to children's trajectories of temperamental reactivity. Developmental Psychology, 46(5), 1089-1102.

Block, J. (1957). A study of affective responsiveness in a lie-detection situation. The Journal of Abnormal and Social Psychology, 55(1), 11-15. doi:10.1037/h0046624

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. Child Development, 75(5), 1523-1537. doi:10.1111/j.1467-8624.2004.00755.x

Bongers, I. L., Koot, H. M., Van der Ende, J., & Verhulst, F. C. (2008). Predicting young adult social functioning from developmental trajectories of externalizing behaviour. Psychological Medicine, 38(7), 989-999. doi:10.1017/S0033291707002309

Book, A. S., & Quinsey, V. L. (2005). Re-examining the issues: A response to Archer et al. Aggression and Violent Behavior, 10(6), 637-646. doi:10.1016/j.avb.2005.01.001

Book, A. S., Starzyk, K. B., & Quinsey, V. L. (2001). The relationship between testosterone and aggression: a meta-analysis. Aggression and Violent Behavior, 6(6), 579-599. doi:10.1016/S1359-1789(00)00032-X

Bornstein, M. H., Putnick, D. L., Gartstein, M. A., Hahn, C. S., Auestad, N., & O'Connor, D. L. (2015a). Infant temperament: Stability by age, gender, birth order, term status, and socioeconomic status. Child development, 86(3), 844-863. doi:10.1111/cdev.12367

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Pearson, R. (2019). Stability of child temperament: Multiple moderation by child and mother characteristics. British Journal of Developmental Psychology, 37(1), 51-67.

Bornstein, M. H., Haynes, O. M., O'Reilly, A. W., & Painter, K. M. (1996). Solitary and Collaborative Pretense Play in Early Childhood: Sources of Individual Variation in the Development of Representational Competence. Child Development, 67(6), 2910-2929. doi:10.2307/1131759

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. Developmental Review, 33(4), 357-370.

Bornstein, M. H., Putnick, D. L., Lansford, J. E., Pastorelli, C., Skinner, A. T., Sorbring, E., ... & Al-Hassan, S. M. (2015b). Mother and father socially desirable responding in nine countries: Two kinds of agreement and relations to parenting self-reports. International Journal of Psychology, 50(3), 174-185.

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. Developmental Psychology, 44(3), 867-874. doi:10.1037/0012-1649.44.3.867

Botvin, G. J., & Griffin, K. W. (2014). Life skills training: preventing substance misuse by enhancing individual and social competence. New directions for youth development, 2014(141), 57-65. doi:10.1002/yd.20086

Boyce, W. T. (2016). Differential Susceptibility of the Developing Brain to Contextual Adversity and Stress. Neuropsychopharmacology, 41(1), 142-162. doi:10.1038/npp.2015.294

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. Development and Psychopathology, 17(2), 271-301.

Boyce, W. T., & Kobor, M. S. (2015). Development and the epigenome: the ‘synapse’of gene–environment interplay. Developmental Science, 18(1), 1-23. doi:10.1111/desc.12282

Boyce, W. T., Alkon, A., Tschann, J. M., Chesney, M. A., & Alpert, B. S. (1995). Dimensions of psychobiologic reactivity: Cardiovascular responses to laboratory stressors in preschool children. Annals of Behavioral Medicine, 17(4), 315. doi:10.1007/BF02888596

Bradley, R. H., & Caldwell, B. M. (1984). The HOME Inventory and family demographics. Developmental Psychology, 20(2), 315.

Bradley, R. H., & Corwyn, R. F. (2008). Infant temperament, parenting, and externalizing behavior in first grade: A test of the differential susceptibility hypothesis. Journal of Child Psychology and Psychiatry, 49(2), 124-131.

Brambilla, P., Perez, J., Barale, F., Schettini, G., & Soares, J. C. (2003). GABAergic dysfunction in mood disorders. Molecular Psychiatry, 8(8), 721-737. doi:10.1038/sj.mp.4001362

Breheny, P., & Burchett, W. (2017). Visualization of Regression Models Using visreg. The R Journal, 9: 56-71.

Brock, L. L., & Curby, T. W. (2016). The role of children's adaptability in classrooms characterized by low or high teacher emotional support consistency. School Psychology Review, 45(2), 209-225.

Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. In Handbook of child psychology: Theoretical models of human development, Vol. 1, 6th ed. (pp. 793-828). Hoboken, NJ, US: John Wiley & Sons Inc.

Brummett, B. H., Krystal, A. D., Siegler, I. C., Kuhn, C., Surwit, R. S., Züchner, S., ... & Williams, R. B. (2007). Associations of a regulatory polymorphism of monoamine oxidase-A gene promoter (MAOA-uVNTR) with symptoms of depression and sleep quality. Psychosomatic Medicine, 69(5), 396-401. doi:10.1097/PSY.0b013e31806d040b

Bueller, J. A., Aftab, M., Sen, S., Gomez-Hassan, D., Burmeister, M., & Zubieta, J.-K. (2006). BDNF Val66Met allele is associated with reduced hippocampal volume in healthy subjects. Biological Psychiatry, 59(9), 812-815.

Buss, A. H., & Plomin, R. (2014). Temperament (PLE: Emotion): Early developing personality traits. Psychology Press.

Caldarella, P., & Merrell, K. W. (1997). Common dimensions of social skills of children and adolescents: A taxonomy of positive behaviors. School Psychology Review, 26(2), 264-278.

Campbell, D. T., & Stanley, J. C. (2015). Experimental and quasi-experimental designs for research. Ravenio Books.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. Journal of child Psychology and Psychiatry, 36(1), 113-149. doi:10.1111/j.1469-7610.1995.tb01657.x

Camras, L. A., & Halberstadt, A. G. (2017). Emotional development through the lens of affective social competence. Current Opinion in Psychology, 17, 113-117.

*Cao, C., Wang, M., Ji, L., Wei, X., Cao, Y., & Zhang, W. (2016). The MAOA rs6323 polymorphism interacts with maternal supportive parenting in predicting adolescent depression: Testing the diathesis-stress and differential susceptibility hypotheses. Acta Psychologica Sinica, 48(1), 22-35.

Card, N. A., & Little, T. D. (2007). Longitudinal modeling of developmental processes. International Journal of Behavioral Development, 31(4), 297-302.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the Infant Temperament Questionnaire. Pediatrics, 61(5), 735-739.

Caspi, A., & Moffitt, T. E. (2006). Gene–environment interactions in psychiatry: joining forces with neuroscience. Nature Reviews Neuroscience, 7(7), 583.

Caspi, A., Hariri, A. R., Holmes, A., Uher, R., & Moffitt, T. E. (2010). Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. American journal of Psychiatry, 167(5), 509-527. doi:10.1176/appi.ajp.2010.09101452

Caspi, A., Henry, B., McGee, R. O., Moffitt, T. E., & Silva, P. A. (1995). Temperamental origins of child and adolescent behavior problems: from age three to age fifteen. Child Development, 66(1), 55-68. doi:10.2307/1131190

*Chen, F. R., Raine, A., & Granger, D. A. (2018). Testosterone and proactive-reactive aggression in youth: the moderating role of harsh discipline. Journal of Abnormal Child Psychology, 46(8), 1599-1612.

*Chen, J., Yu, J., Liu, Y., Zhang, L., & Zhang, J. (2015). BDNF Val66Met, stress, and positive mothering: Differential susceptibility model of adolescent trait anxiety. Journal of Anxiety disorders, 34, 68-75.

Chen, X., Liu, M., & Li, D. (2000). Parental warmth, control, and indulgence and their relations to adjustment in Chinese children: A longitudinal study. Journal of Family Psychology, 14(3), 401-419. doi:10.1037/0893-3200.14.3.401

Chen, Z. Y., Jing, D., Bath, K. G., Ieraci, A., Khan, T., Siao, C. J., . . . Lee, F. S. (2006). Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. Science, 314(5796), 140-143. doi:10.1126/science.1129663

Chen, Z. Y., Patel, P. D., Sant, G., Meng, C. X., Teng, K. K., Hempstead, B. L., & Lee, F. S. (2004). Variant brain-derived neurotrophic factor (BDNF)(Met66) alters the intracellular trafficking and activity-dependent secretion of wild-type BDNF in neurosecretory cells and cortical neurons. Journal of Neuroscience, 24(18), 4401-4411.

Chess, S., & Thomas, A. (1991). Temperament and the concept of goodness of fit. In Explorations in Temperament (pp. 15-28). Springer, Boston, MA.

*Cho, J. H., Kogan, S. M., & Brody, G. H. (2016). Genetic Moderation of Transactional Relations Between Parenting Practices and Child Self-Regulation. Journal of Family Psychology, 30(7), 780-790. doi:10.1037/fam0000228

Chrousos, G. P., & Gold, P. W. (1992). The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. JAMA, 267(9), 1244-1252. doi:10.1001/jama.1992.03480090092034

Cicchetti, D., & Dawson, G. (2002). Multiple levels of analysis. Development and Psychopathology, 14(3), 417-420.

Clark, K. E., & Ladd, G. W. (2000). Connectedness and autonomy support in parent–child relationships: Links to children's socioemotional orientation and peer relationships. Developmental Psychology, 36(4), 485-498. doi:10.1037/0012-1649.36.4.485

Coghlan, S., Horder, J., Inkster, B., Mendez, M. A., Murphy, D. G., & Nutt, D. J. (2012). GABA system dysfunction in autism and related disorders: from synapse to symptoms. Neuroscience & Biobehavioral Reviews, 36(9), 2044-2055. doi:10.1016/j.neubiorev.2012.07.005

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The Problem of Units and the Circumstance for POMP. Multivariate Behavioral Research, 34(3), 315-346. doi:10.1207/S15327906MBR3403_2

Collett, B. R., Ohan, J. L., & Myers, K. M. (2003). Ten-Year Review of Rating Scales. VI: Scales Assessing Externalizing Behaviors. Journal of the American Academy of Child & Adolescent Psychiatry, 42(10), 1143-1170. doi:10.1097/00004583-200310000-00006

Comings, D. E., Gade-Andavolu, R., Gonzalez, N., Wu, S., Muhleman, D., Blake, H., . . . P MacMurray, J. (2001). Comparison of the role of dopamine, serotonin, and noradrenaline genes in ADHD, ODD and conduct disorder: multivariate regression analysis of 20 genes. Clinical Genetics, 57(3), 178-196. doi:10.1034/j.1399-0004.2000.570304.x

Cox, M., & Crnic, K. (2002). Qualitative ratings for parent–child interaction at 3–12 months of age. Unpublished manuscript, University of North Carolina at Chapel Hill.

Craig, L., & Bittman, M. (2008). The incremental time costs of children: An analysis of children's impact on adult time use in Australia. Feminist Economics, 14(2), 59-88. doi:10.1080/13545700701880999

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334.

Crowell, S. E., Beauchaine, T. P., Gatzke-Kopp, L., Sylvers, P., Mead, H., & Chipman-Chacon, J. (2006). Autonomic correlates of attention-deficit/hyperactivity disorder and oppositional defiant disorder in preschool children. Journal of Abnormal Psychology, 115(1), 174-178. doi:10.1037/0021-843X.115.1.174

Cunha, C., Brambilla, R., & Thomas, K. L. (2010). A simple role for BDNF in learning and memory? Frontiers in Molecular Neuroscience, 3, 1. doi:10.3389/neuro.02.001.2010

Das, J. K., Salam, R. A., Arshad, A., Finkelstein, Y., & Bhutta, Z. A. (2016). Interventions for adolescent substance abuse: An overview of systematic reviews. Journal of Adolescent Health, 59(4), S61-S75. doi:10.1016/j.jadohealth.2016.06.021

Davidov, M., & Grusec, J. E. (2006). Untangling the links of parental responsiveness to distress and warmth to child outcomes. Child Development, 77(1), 44-58. doi:10.1111/j.1467-8624.2006.00855.x

Davidson, L. A., Crowder, M. K., Gordon, R. A., Domitrovich, C. E., Brown, R. D., & Hayes, B. I. (2018). A continuous improvement approach to social and emotional competency measurement. Journal of Applied Developmental Psychology, 55, 93-106.

*Davies, P., Cicchetti, D., & Hentges, R. F. (2015). Maternal unresponsiveness and child disruptive problems: the interplay of uninhibited temperament and dopamine transporter genes. Child Development, 86(1), 63-79. doi:10.1111/cdev.12281

Davis, E. A. (2014). Optimising the education and care experiences of children with 'difficult' temperament (Doctoral dissertation, Queensland University of Technology).

Davis, E., Eivers, A., & Thorpe, K. (2012). Is Quality More Important if You're Quirky? A Review of the Literature on Differential Susceptibility to Childcare Environments. Australasian Journal of Early Childhood, 37(4), 99-106.

Day, K. L., Schmidt, L. A., Vaillancourt, T., Saigal, S., Boyle, M. H., & Van Lieshout, R. J. (2018). Overprotective Parenting and Peer Victimization in Extremely Low Birth Weight Survivors. Journal of Child and Family Studies, 27(3), 907-915. doi:10.1007/s10826-017-0922-2

De Graaf, I., Speetjens, P., Smit, F., de Wolff, M., & Tavecchio, L. (2008). Effectiveness of the Triple P Positive Parenting Program on behavioral problems in children: A meta-analysis. Behavior Modification, 32(5), 714-735. doi:10.1177/0145445508317134

De Pauw, S. S., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. Child Psychiatry & Human Development, 41(3), 313-329.

de Quervain, D. J., Kolassa, I. T., Ertl, V., Onyut, P. L., Neuner, F., Elbert, T., & Papassotiropoulos, A. (2007). A deletion variant of the α2b-adrenoceptor is related to emotional memory in Europeans and Africans. Nature Neuroscience, 10(9), 1137-1139. doi:10.1038/nn1945

De Wolff, M. S., & van Ijzendoorn, M. H. (1997). Sensitivity and attachment: a meta-analysis on parental antecedents of infant attachment. Child Development, 68(4), 571-591. doi:10.1111/j.1467-8624.1997.tb04218.x

Dearing, E., & Hamilton Lawrence, C. (2006). V. Contemporary advances and classic advice for analyzing mediating and moderating variables. Monographs of the Society for Research in Child Development, 71(3), 88-104. doi:10.1111/j.1540-5834.2006.00406.x

Deater-Deckard, K., Mullineaux, P. Y., Beekman, C., Petrill, S. A., Schatschneider, C., & Thompson, L. A. (2009). Conduct problems, IQ, and household chaos: a longitudinal multi-informant study. Journal of child Psychology and Psychiatry, 50(10), 1301-1308. doi:10.1111/j.1469-7610.2009.02108.x

Del Giudice, M. (2017). Statistical tests of differential susceptibility: Performance, limitations, and improvements. Development and Psychopathology, 29(4), 1267-1278. doi:10.1017/S0954579416001292

Denham, S. A. (1998). Emotional development in young children: Guilford Press.

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach–Major, S., & Queenan, P. (2003). Preschool Emotional Competence: Pathway to Social Competence? Child Development, 74(1), 238-256. doi:10.1111/1467-8624.00533

Derogatis, L. R., & Fitzpatrick, M. (2004). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. PloS one, 9(1), e83875.

DeVellis, R. F. (2016). Scale development: Theory and applications (Vol. 26). Sage publications.

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., … Sher, K. J. (2015). Candidate Gene–Environment Interaction Research: Reflections and Recommendations. Perspectives on Psychological Science, 10(1), 37–59.

Dick, D. M., Bierut, L., Hinrichs, A., Fox, L., Bucholz, K. K., Kramer, J., . . . Foroud, T. (2006). The role of GABRA2 in risk for conduct disorder and alcohol and drug dependence across developmental stages. Behavior Genetics, 36(4), 577-590. doi:10.1007/s10519-005-9041-8

Dick, D. M., Latendresse, S. J., Lansford, J. E., Budde, J. P., Goate, A., Dodge, K. A., ... & Bates, J. E. (2009). Role of GABRA2 in trajectories of externalizing behavior across development and evidence of moderation by parental monitoring. Archives of general psychiatry, 66(6), 649-657. doi:10.1001/archgenpsychiatry.2009.48

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychological Bulletin, 130(3), 355-391. doi:10.1037/0033-2909.130.3.355

Diderichsen, F., Hallqvist, J., & Whitehead, M. (2019). Differential vulnerability and susceptibility: how to make use of recent development in our understanding of mediation and interaction to tackle health inequalities. International Journal of Epidemiology, 48(1), 268-274.

Dodge, K. A., Pettit, G. S., Bates, J. E., & Valente, E. (1995). Social information-processing patterns partially mediate the effect of early physical abuse on later conduct problems. Journal of Abnormal Psychology, 104(4), 632-643.

Domitrovich, C. E., Cortes, R., & Greenberg, M. T. (2000). The teacher style rating scale technical report. Unpublished manuscript, Pennsylvania State University.

Domitrovich, C. E., Durlak, J. A., Staley, K. C., & Weissberg, R. P. (2017). Social-Emotional Competence: An Essential Factor for Promoting Positive Adjustment and Reducing Risk in School Children. Child Development, 88(2), 408-416. doi:10.1111/cdev.12739

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. Early Childhood Research Quarterly, 25(1), 1-16.

D'Souza, H., & Karmiloff-Smith, A. (2017). Neurodevelopmental disorders. Wiley Interdisciplinary Reviews: Cognitive Science, 8(1-2). doi:10.1002/wcs.1398

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. Developmental Psychology, 50(11), 2417.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105(3), 399-412.

Dunning, T. (2008). Improving causal inference: Strengths and limitations of natural experiments. Political Research Quarterly, 61(2), 282-293.

Early Child Care Research Network, N.I.O.C.H.A.H.D. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten?. Child Development, 74(4), 976-1005.

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. Psychological Methods, 23(1), 138-149.

Edwards, S. L., Rapee, R. M., & Kennedy, S. (2010). Prediction of anxiety symptoms in preschool-aged children: examination of maternal and paternal perspectives. Journal of child Psychology and Psychiatry, 51(3), 313-321. doi:10.1111/j.1469-7610.2009.02160.x

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., ... & Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proceedings of the National Academy of Sciences, 98(12), 6917-6922. doi:10.1073/pnas.111134598

Elgar, F. J., Waschbusch, D. A., Dadds, M. R., & Sigvaldason, N. (2007). Development and validation of a short form of the Alabama Parenting Questionnaire. Journal of Child and Family Studies, 16(2), 243-259.

Ellis, B. J., & Boyce, W. T. (2011). Differential susceptibility to the environment: toward an understanding of sensitivity to developmental experiences and context. Development and Psychopathology, 23(1), 1-5. doi:10.1017/S095457941000060X

Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary–neurodevelopmental theory. Development and Psychopathology, 23(1), 7-28.

El-Sheikh, M. (2007). Children's skin conductance level and reactivity: are these measures stable over time and across tasks? Developmental Psychobiology, 49(2), 180-186. doi:10.1002/dev.20171

El-Sheikh, M., Keiley, M., & Hinnant, J. B. (2010). Developmental trajectories of skin conductance level in middle childhood: sex, race, and externalizing behavior problems as predictors of growth. Biological Psychology, 83(2), 116-124. doi:10.1016/j.biopsycho.2009.11.009

Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. Psychological Methods, 23(2), 298-317. doi:10.1037/met0000148

Enders, C.K., Du, H. Keller, B.T. (2018). Substantive model-compatible imputation for multilevel regression models with random coefficients, interaction effects, and other nonlinear terms. Paper submitted for publication.

Enoch, M. A., White, K. V., Waheed, J., & Goldman, D. (2008). Neurophysiological and genetic distinctions between pure and comorbid anxiety disorders. Depression and Anxiety, 25(5), 383-392. doi:10.1002/da.20378

Evans, D. (2003). Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. Journal of Clinical Nursing, 12(1), 77-84.

Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. Psychological Bulletin, 139(6), 1342.

Feldman, R. (2007). Parent–Infant Synchrony. Current Directions in Psychological Science, 16(6), 340-345. doi:10.1111/j.1467-8721.2007.00532.x

Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2005). Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. Journal of child Psychology and Psychiatry, 46(8), 837-849. doi:10.1111/j.1469-7610.2004.00387.x

Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2007). Conduct and attentional problems in childhood and adolescence and later substance use, abuse and dependence: results of a 25-year longitudinal study. Drug and Alcohol Dependence, 88 Suppl 1, S14-26. doi:10.1016/j.drugalcdep.2006.12.011

Forster, M., Grigsby, T. J., Bunyan, A., Unger, J. B., & Valente, T. W. (2015). The protective role of school friendship ties for substance use and aggressive behaviors among middle school students. Journal of School Health, 85(2), 82-89. doi:10.1111/josh.12230

Fowles, D. C., Kochanska, G., & Murray, K. (2000). Electrodermal activity and temperament in preschool children. Psychophysiology, 37(6), 777-787.

Fraley, R. C., & Roisman, G. I. (2015). Do early caregiving experiences leave an enduring or transient mark on developmental adaptation?. Current Opinion in Psychology, 1, 101-106.

Fraley, R. C., Roisman, G. I., & Haltigan, J. D. (2013). The legacy of early experiences in development: Formalizing alternative models of how early experiences are carried forward over time. Developmental Psychology, 49(1), 109-126. doi:10.1037/a0027852

Francis, S. M., Kistner-Griffin, E., Yan, Z., Guter, S., Cook, E. H., & Jacob, S. (2016). Variants in Adjacent Oxytocin/Vasopressin Gene Region and Associations with ASD Diagnosis and Other Autism Related Endophenotypes. Frontiers in Neuroscience, 10(195), 195. doi:10.3389/fnins.2016.00195

Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. PLoS Biology, 13(6), e1002165.

French, S. (2003). Modelling, making inferences and making decisions: the roles of sensitivity analysis. Top, 11(2), 229-251.

Frick, P. J., & Morris, A. S. (2004). Temperament and Developmental Pathways to Conduct Problems. Journal of Clinical Child & Adolescent Psychology, 33(1), 54-68. doi:10.1207/S15374424JCCP3301_6

Frick, P. J., Lahey, B. B., Loeber, R., Tannenbaum, L., Vanhorn, Y., Christ, M. A. G., . . . Hanson, K. (1993). Oppositional Defiant Disorder and Conduct Disorder - a Meta-Analytic Review of Factor-Analyses and Cross-Validation in a Clinic Sample. Clinical Psychology Review, 13(4), 319-340. doi:10.1016/0272-7358(93)90016-F

Frodl, T., Schule, C., Schmitt, G., Born, C., Baghai, T., Zill, P., . . . Meisenzahl, E. M. (2007). Association of the brain-derived neurotrophic factor Val66Met polymorphism with reduced hippocampal volumes in major depression. Archives of General Psychiatry, 64(4), 410-416. doi:10.1001/archpsyc.64.4.410

Fuchs, D., & Fuchs, L. S. (2019). On the Importance of Moderator Analysis in Intervention Research: An Introduction to the Special Issue. Exceptional Children, 85(2), 126–128.

Furmark, T., Tillfors, M., Garpenstrand, H., Marteinsdottir, I., Langstrom, B., Oreland, L., & Fredrikson, M. (2004). Serotonin transporter polymorphism related to amygdala excitability and symptom severity in patients with social phobia. Neuroscience Letters, 362(3), 189-192. doi:10.1016/j.neulet.2004.02.070

Furness, J. B. (2006). The organisation of the autonomic nervous system: peripheral connections. Autonomic Neuroscience: Basic and Clinical, 130(1), 1-5. doi:10.1016/j.autneu.2006.05.003

Gabella, G. (2001). Autonomic nervous system. e LS.

*Gallitto, E. (2015). Temperament as a moderator of the effects of parenting on children's behavior. Development and Psychopathology, 27(3), 757-773. doi:10.1017/S0954579414000753

Gardner, F. (2000). Methodological Issues in the Direct Observation of Parent–Child Interaction: Do Observational Findings Reflect the Natural Behavior of Participants? Clinical Child and Family Psychology Review, 3(3), 185-198. doi:10.1023/A:1009503409699

Gardner, F. E. M. (1994). The Quality of Joint Activity between Mothers and Their Children with Behavior Problems. Journal of Child Psychology and Psychiatry and Allied Disciplines, 35(5), 935-948. doi:10.1111/j.1469-7610.1994.tb02303.x

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological Bulletin, 118(3), 392.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. Infant Behavior and Development, 26(1), 64-86.

Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review. Psychological Bulletin, 128(4), 539-579. doi:10.1037/0033-2909.128.4.539

Goenjian, A. K., Bailey, J. N., Walling, D. P., Steinberg, A. M., Schmidt, D., Dandekar, U., & Noble, E. P. (2012). Association of TPH1, TPH2, and 5HTTLPR with PTSD and depressive symptoms. Journal of Affective Disorders, 140(3), 244-252. doi:10.1016/j.jad.2012.02.015

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., ... & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. Child development, 58(2), 505-529.

Goldstein, H., Carpenter James, R., & Browne William, J. (2013). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177(2), 553-564. doi:10.1111/rssa.12022

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. Journal of the American Academy of Child & Adolescent Psychiatry, 48(4), 400-403.

Goodman, R. (1994). A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note. Journal of Child Psychology and Psychiatry, 35(8), 1483-1494.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry, 38(5), 581-586. doi:10.1111/j.1469-7610.1997.tb01545.x

Gottesman, I. I., & Shields, J. (1973). Genetic Theorizing and Schizophrenia. British Journal of Psychiatry, 122(566), 15-30. doi:10.1192/bjp.122.1.15

Gottesman, II, & Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. American Journal of Psychiatry, 160(4), 636-645. doi:10.1176/appi.ajp.160.4.636

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. Prevention Science, 8(3), 206-213. doi:10.1007/s11121-007-0070-9

Graham, P., Rutter, M., & George, S. (1973). Temperamental Characteristics as Predictors of Behavior Disorders in Children. American Journal of Orthopsychiatry, 43(3), 328-339. doi:10.1111/j.1939-0025.1973.tb00802.x

Greven, C. U., Lionetti, F., Booth, C., Aron, E., Fox, E., Schendan, H. E., ... & Homberg, J. (2019). Sensory Processing Sensitivity in the context of Environmental Sensitivity: a critical review and development of research agenda. Neuroscience & Biobehavioral Reviews, 98, 287-305.

Grund, S., Lüdtke, O., & Robitzsch, A. (2017). Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations. Organizational Research Methods, 21(1), 111-149. doi:10.1177/1094428117703686

Guerin, D. W., Gottfried, A. W., & Thomas, C. W. (1997). Difficult temperament and behaviour problems: A longitudinal study from 1.5 to 12 years. International Journal of Behavioral Development, 21(1), 71-90. doi:10.1080/016502597384992

Gunnar, M. R., & Vazquez, D. M. (2001). Low cortisol and a flattening of expected daytime rhythm: potential indices of risk in human development. Development and Psychopathology, 13(3), 515-538.

Gutiérrez, B., Arias, B., Gasto, C., Catalan, R., Papiol, S., Pintor, L., & Fananas, L. (2004). Association analysis between a functional polymorphism in the monoamine oxidase A gene promoter and severe mood disorders. Psychiatric Genetics, 14(4), 203-208. doi:10.1097/00041444-200412000-00007

Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. Social Development, 10(1), 79-119.

Hall, F. S., & Perona, M. T. (2012). Have studies of the developmental regulation of behavioral phenotypes revealed the mechanisms of gene–environment interactions?. Physiology & Behavior, 107(5), 623-640.

Haltigan, J. D., Roisman, G. I., & Fraley, R. C. (2013). The predictive significance of early caregiving experiences for symptoms of psychopathology through midadolescence: Enduring or transient effects?. Development and Psychopathology, 25(1), 209-221.

Haltigan, J. D., Roisman, G. I., Cauffman, E., & Booth-LaForce, C. (2017). Correlates of childhood vs. adolescence internalizing symptomatology from infancy to young adulthood. Journal of Youth and Adolescence, 46(1), 197-212.

Hastings, P. D., Helm, J., Mills, R. S. L., Serbin, L. A., Stack, D. M., & Schwartzman, A. E. (2015). Dispositional and Environmental Predictors of the Development of Internalizing Problems in Childhood: Testing a Multilevel Model. Journal of Abnormal Child Psychology, 43(5), 831-845. doi:10.1007/s10802-014-9951-0

Hattersley, A. T., & McCarthy, M. I. (2005). What makes a good genetic association study? The Lancet, 366(9493), 1315-1323. doi:10.1016/s0140-6736(05)67531-9

Hawes, D. J., & Dadds, M. R. (2006). Assessing Parenting Practices Through Parent-Report and Direct Observation During Parent-Training. Journal of Child and Family Studies, 15(5), 554-567. doi:10.1007/s10826-006-9029-x

Hawi, Z., Matthews, N., Barry, E., Kirley, A., Wagner, J., Wallace, R. H., . . . Bellgrove, M. A. (2013). A high density linkage disequilibrium mapping in 14 noradrenergic genes: evidence of association between SLC6A2, ADRA1B and ADHD. Psychopharmacology (Berl), 225(4), 895-902. doi:10.1007/s00213-012-2875-x

Hedges, L. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. Journal of Educational Statistics, 6(2), 107-128.

Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., & Lesch, K. P. (1996). Allelic variation of human serotonin transporter gene expression. Journal of Neurochemistry, 66(6), 2621-2624. doi:10.1046/j.1471-4159.1996.66062621.x

Henman, P., Percival, R., Harding, A., & Gray, M. (2007). Costs of children: research commissioned by the Ministerial Taskforce on Child Support.

Henricsson, L., & Rydell, A. M. (2006). Children with behaviour problems: The influence of social competence and social relations on problem stability, school achievement and peer acceptance across the first six years of school. Infant and Child Development, 15(4), 347-366. doi:10.1002/icd.448

Herman-Stahl, M., & Petersen, A. C. (1996). The protective role of coping and social resources for depressive symptoms among young adolescents. Journal of Youth and Adolescence, 25(6), 733-753. doi:10.1007/bf01537451

Hines, M., Constantinescu, M., & Spencer, D. (2015). Early androgen exposure and human gender development. Biology of Sex Differences, 6(1), 3. doi:10.1186/s13293-015-0022-1

Hoppe, D. J., Schemitsch, E. H., Morshed, S., Tornetta III, P., & Bhandari, M. (2009). Hierarchy of evidence: where observational studies fit in and why we need them. JBJS, 91(Supplement_3), 2-9.

Howarth, G. Z., Fettig, N. B., Curby, T. W., & Bell, M. A. (2016). Frontal electroencephalogram asymmetry and temperament across infancy and early childhood: An exploration of stability and bidirectional relations. Child Development, 87(2), 465-476.

Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. Child Development, 63(2), 449-460.

Ising, M., Depping, A. M., Siebertz, A., Lucae, S., Unschuld, P. G., Kloiber, S., ... & Holsboer, F. (2008). Polymorphisms in the FKBP5 gene region modulate recovery from psychosocial stress in healthy controls. European Journal of Neuroscience, 28(2), 389-398. doi:10.1111/j.1460-9568.2008.06332.x

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. Psychological Bulletin, 117(2), 348-357.

Jackson, S. E., & Chester, J. D. (2015). Personalised cancer medicine. International Journal of Cancer, 137(2), 262-266.

Janson, H., & Mathiesen, K. S. (2008). Temperament profiles from infancy to middle childhood: development and associations with behavior problems. Developmental Psychology, 44(5), 1314.

*Janssens, A., Van Den Noortgate, W., Goossens, L., Colpin, H., Verschueren, K., Claes, S., & Van Leeuwen, K. (2017). Externalizing problem behavior in adolescence: parenting interacting with DAT1 and DRD4 genes. Journal of Research on Adolescence, 27(2), 278-297. doi:10.1111/jora.12271

Jewell, J. D., & Stark, K. D. (2003). Comparing the family environments of adolescents with conduct disorder or depression. Journal of Child and Family Studies, 12(1), 77-89. doi:10.1023/a:1021310226400

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness. American Journal of Public Health, 105(11), 2283-2290. doi:10.2105/Ajph.2015.302630

Keller, B. T., & Enders, C. K. (2018). Blimp User's Manual (Version 1.1). Los Angeles, CA.

Kendler, K. S., Myers, J. M., & Keyes, C. L. M. (2011a). The Relationship Between the Genetic and Environmental Influences on Common Externalizing Psychopathology and Mental Wellbeing. Twin Research and Human Genetics, 14(6), 516-523. doi:10.1375/twin.14.6.516

Kendler, K. S., Myers, J. M., Maes, H. H., & Keyes, C. L. (2011b). The relationship between the genetic and environmental influences on common internalizing psychiatric disorders and mental well-being. Behavior Genetics, 41(5), 641-650. doi:10.1007/s10519-011-9466-1

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. International Journal of Behavioral Development, 40(1), 64-75. doi:10.1177/0165025415570647

Kessler, R. C., Barker, P. R., Colpe, L. J., & et al. (2003). Screening for serious mental illness in the general population. Archives of General Psychiatry, 60(2), 184-189. doi:10.1001/archpsyc.60.2.184

Kiff, C. J., Lengua, L. J., & Zalewski, M. (2011). Nature and Nurturing: Parenting in the Context of Child Temperament. Clinical Child and Family Psychology Review, 14(3), 251-301. doi:10.1007/s10567-011-0093-4

Kim, H., Carlson, A. G., Curby, T. W., & Winsler, A. (2016). Relations among motor, social, and cognitive skills in pre-kindergarten children with developmental disabilities. Research in Developmental Disabilities, 53, 43-60.

*Kim, S., & Kochanska, G. (2012). Child temperament moderates effects of parent-child mutuality on self-regulation: a relationship-based path for emotionally negative infants. Child Development, 83(4), 1275-1289. doi:10.1111/j.1467-8624.2012.01778.x

*King, A. P., Muzik, M., Hamilton, L., Taylor, A. B., Rosenblum, K. L., & Liberzon, I. (2016). Dopamine Receptor Gene DRD4 7-Repeat Allele X Maternal Sensitivity Interaction on Child Externalizing Behavior Problems: Independent Replication of Effects at 18 Months. PLOS ONE, 11(8), e0160473. doi:10.1371/journal.pone.0160473

Kirley, A., Lowe, N., Mullins, C., McCarron, M., Daly, G., Waldman, I., . . . Hawi, Z. (2004). Phenotype studies of the DRD4 gene polymorphisms in ADHD: association with oppositional defiant disorder and positive family history. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 131B(1), 38-42. doi:10.1002/ajmg.b.30072

Klein, M., Schmoeger, M., Kasper, S., & Schosser, A. (2016). Meta-analysis of the COMT Val158Met polymorphism in major depressive disorder: the role of gender. The World Journal of Biological Psychiatry, 17(2), 147-158. doi:10.3109/15622975.2015.1083615

Knafo, A., & Plomin, R. (2006). Parental discipline and affection and children's prosocial behavior: Genetic and environmental links. Journal of Personality and Social Psychology, 90(1), 147-164. doi:10.1037/0022-3514.90.1.147

Kochanska, G., Barry, R. A., Aksan, N., & Boldt, L. J. (2008). A developmental model of maternal and child contributions to disruptive conduct: the first six years. Journal of child Psychology and Psychiatry, 49(11), 1220-1227. doi:10.1111/j.1469-7610.2008.01932.x

*Kochanska, G., Boldt, L. J., Kim, S., Yoon, J. E., & Philibert, R. A. (2015). Developmental interplay between children's biobehavioral risk and the parenting environment from toddler to early school age: Prediction of socialization outcomes in preadolescence. Development and Psychopathology, 27(3), 775-790. doi:10.1017/S0954579414000777

*Kochanska, G., Brock, R. L., Chen, K. H., Aksan, N., & Anderson, S. W. (2015). Paths from mother-child and father-child relationships to externalizing behavior problems in children differing in electrodermal reactivity: A longitudinal study from infancy to age 10. Journal of Abnormal Child Psychology, 43(4), 721-734. doi:10.1007/s10802-014-9938-x

*Kochanska, G., Kim, S., Barry, R. A., & Philibert, R. A. (2011). Children's genotypes interact with maternal responsive care in predicting children's competence: Diathesis–stress or differential susceptibility?. Development and Psychopathology, 23(2), 605-616. doi:10.1017/S0954579411000071

Kokko, K., Tremblay, R. E., Lacourse, E., Nagin, D. S., & Vitaro, F. (2006). Trajectories of prosocial behavior and physical aggression in middle childhood: Links to adolescent school dropout and physical violence. Journal of Research on Adolescence, 16(3), 403-428. doi:10.1111/j.1532-7795.2006.00500.x

Koss, K. J., & Gunnar, M. R. (2018). Annual Research Review: Early adversity, the hypothalamic-pituitary-adrenocortical axis, and child psychopathology. Journal of Child Psychology and Psychiatry, 59(4), 327-346. doi:10.1111/jcpp.12784

Kravets, J. L., Reyes, B. A., Unterwald, E. M., & Van Bockstaele, E. J. (2015). Direct targeting of peptidergic amygdalar neurons by noradrenergic afferents: linking stress-integrative circuitry. Brain Structure and Function, 220(1), 541-558. doi:10.1007/s00429-013-0674-8

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. Biological Psychology, 84(3), 394-421. doi:10.1016/j.biopsycho.2010.03.010

Kristal, J. (2005). The temperament perspective: Working with children's behavioral styles. Paul H Brookes Publishing.

Kuckertz, J. M., Mitchell, C., & Wiggins, J. L. (2018). Parenting mediates the impact of maternal depression on child internalizing symptoms. Depression and Anxiety, 35(1), 89-97.

Kuhn, T. S. (2012). The structure of scientific revolutions. University of Chicago press.

Ladd, G. W., & Profilet, S. M. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. Developmental Psychology, 32(6), 1008-1024. doi:10.1037/0012-1649.32.6.1008

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. Developmental Psychology, 42(4), 627-642.

Le Magueresse, C., & Monyer, H. (2013). GABAergic interneurons shape the functional maturation of the cortex. Neuron, 77(3), 388-405. doi:10.1016/j.neuron.2013.01.011

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. Advances in Methods and Practices in Psychological Science, 1(3), 389-402.

Lee, L. O., & Prescott, C. A. (2014). Association of the catechol-O-methyltransferase val158met polymorphism and anxiety-related traits: a meta-analysis. Psychiatric Genetics, 24(2), 52-69. doi:10.1097/YPG.0000000000000018

Leerkes, E. M., Blankson, A. N., & O'Brien, M. (2009). Differential effects of maternal sensitivity to infant distress and nondistress on social-emotional functioning. Child Development, 80(3), 762-775.

Lengua, L. J., West, S. G., & Sandler, I. N. (1998). Temperament as a predictor of symptomatology in children: Addressing contamination of measures. Child Development, 69(1), 164-181.

Lenzenweger, M. F. (2013). Thinking clearly about the endophenotype-intermediate phenotype-biomarker distinctions in developmental psychopathology research. Development and Psychopathology, 25(4 Pt 2), 1347-1357. doi:10.1017/S0954579413000655

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., . . . Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science, 274(5292), 1527-1531.

Leschied, A., Chiodo, D., Nowicki, E., & Rodger, S. (2008). Childhood Predictors of Adult Criminality: A Meta-Analysis Drawn from the Prospective Longitudinal Literature. Canadian Journal of Criminology and Criminal Justice, 50(4), 435-467. doi:10.3138/cjccj.50.4.435

Levy, P. (1968). Short-form tests: A methodological review. Psychological Bulletin, 69(6), 410-416.

*Li, Y., Sulik, M. J., Eisenberg, N., Spinrad, T. L., Lemery-Chalfant, K., Stover, D. A., & Verrelli, B. C. (2016). Predicting childhood effortful control from interactions between early parenting quality and children's dopamine transporter gene haplotypes. Development and Psychopathology, 28(1), 199-212. doi:10.1017/S0954579415000383

Lidberg, L., & Wallin, B. G. (1981). Sympathetic skin nerve discharges in relation to amplitude of skin resistance responses. Psychophysiology, 18(3), 268-270. doi:10.1111/j.1469-8986.1981.tb03033.x

Liu, Z., Zhu, F., Wang, G., Xiao, Z., Wang, H., Tang, J., . . . Li, W. (2006). Association of corticotropin-releasing hormone receptor1 gene SNP and haplotype with major depression. Neuroscience Letters, 404(3), 358-362. doi:10.1016/j.neulet.2006.06.016

Locke, L. M., & Prinz, R. J. (2002). Measurement of parental discipline and nurturance. Clinical Psychology Review, 22(6), 895-929.

Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. Psychological Bulletin, 130(4), 531-552. doi:10.1037/0033-2909.130.4.531

Love, T. M., Enoch, M. A., Hodgkinson, C. A., Pecina, M., Mickey, B., Koeppe, R. A., . . . Zubieta, J. K. (2012). Oxytocin gene polymorphisms influence human dopaminergic function in a sex-dependent manner. Biological Psychiatry, 72(3), 198-206. doi:10.1016/j.biopsych.2012.01.033

LSAC (2016). About growing up in Australia: The Longitudinal Study of Australian children. Available at: http://www.growingupinaustralia.gov.au/about/index.html

Luijk, M. P., Roisman, G. I., Haltigan, J. D., Tiemeier, H., Booth-LaForce, C., Van Ijzendoorn, M. H., ... & Verhulst, F. C. (2011). Dopaminergic, serotonergic, and oxytonergic candidate genes associated with infant attachment security and disorganization? In search of main and interaction effects. Journal of Child Psychology and Psychiatry, 52(12), 1295-1307.

Luyckx, K., Soenens, B., Vansteenkiste, M., Goossens, L., & Berzonsky, M. D. (2007). Parental psychological control and dimensions of identity formation in emerging adulthood. Journal of Family Psychology, 21(3), 546-550. doi:10.1037/0893-3200.21.3.546

Maccoby, E. E. (1992). The role of parents in the socialization of children: An historical overview. Developmental Psychology, 28(6), 1006-1017. doi:10.1037/0012-1649.28.6.1006

Maccoby, E. E. (2000). Parenting and its effects on children: On reading and misreading behavior genetics. Annual Review of Psychology, 51(1), 1-27.

MacDonald, K. (1992). Warmth as a Developmental Construct: An Evolutionary Analysis. Child Development, 63(4), 753-773. doi:10.2307/1131231

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. Journal of Statistical Software, 20(9), 1-20.

Mair, P., Hatzinger, R., & Maier M. J. (2019). eRm: Extended Rasch Modeling. 1.0-0. http://erm.r-forge.r-project.org/

Mannuzza, S., Klein, R. G., & Moulton, J. L., 3rd. (2008). Lifetime criminality among boys with attention deficit hyperactivity disorder: a prospective follow-up study into adulthood using official arrest records. Psychiatry Research, 160(3), 237-246. doi:10.1016/j.psychres.2007.11.003

Manuck, S. (2011). Delay discounting covaries with childhood socioeconomic status as a function of genetic variation in the dopamine D4 receptor (DRD4). Paper presented at the Presentazione a congresso, Society for Research in Child Development, Montreal, Canada.

Masten, A. S., & Obradovic, J. (2006). Competence and resilience in development. Annals of the New York Academy of Sciences, 1094(1), 13-27. doi:10.1196/annals.1376.003

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., . . . Tellegen, A. (2005). Developmental Cascades: Linking Academic Achievement and Externalizing and Internalizing Symptoms Over 20 Years. Developmental Psychology, 41(5), 733-746. doi:10.1037/0012-1649.41.5.733

Matson, J. L., Rotatori, A. F., & Helsel, W. J. (1983). Development of a rating scale to measure social skills in children: the Matson Evaluation of Social Skills with Youngsters (MESSY).

Behaviour Research and Therapy, 21(4), 335-340. doi:https://doi.org/10.1016/0005-7967(83)90001-3

McAdams, T. A., Neiderhiser, J. M., Rijsdijk, F. V., Narusyte, J., Lichtenstein, P., & Eley, T. C. (2014). Accounting for genetic and environmental confounds in associations between parent and child characteristics: a systematic review of children-of-twins studies. Psychological Bulletin, 140(4), 1138-1173. doi:10.1037/a0036416

McBurnett, K., Lahey, B. B., Rathouz, P. J., & Loeber, R. (2000). Low salivary cortisol and persistent aggression in boys referred for disruptive behavior. Archives of General Psychiatry, 57(1), 38-43.

McCabe, C. J., Kim, D. S., & King, K. M. (2018). Improving present practices in the visual display of interactions. Advances in Methods and Practices in Psychological Science, 1(2), 147-165.

McClelland, G. H. (1997). Optimal design in psychological research. Psychological Methods, 2(1), 3-19.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin, 114(2), 376-390.

Melhuish Edward, C., Phan Mai, B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the Home Learning Environment and Preschool Center Experience upon Literacy and Numeracy Development in Early Primary School. Journal of Social Issues, 64(1), 95-114. doi:10.1111/j.1540-4560.2008.00550.x

Merlin, T., Weston, A., & Tooher, R. (2009). Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. BMC Medical Research Methodology, 9(1), 34.

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). Unequal in Opportunity, Equal in Process: Parental Sensitivity Promotes Positive Child Development in Ethnic Minority Families. Child Development Perspectives, 6(3), 239-250. doi:10.1111/j.1750-8606.2011.00223.x

Mill, J., Xu, X., Ronald, A., Curran, S., Price, T., Knight, J., . . . Asherson, P. (2005). Quantitative trait locus analysis of candidate gene alleles associated with attention deficit hyperactivity disorder (ADHD) in five genes: DRD4, DAT1, DRD5, SNAP-25, and 5HT1B. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 133B(1), 68-73. doi:10.1002/ajmg.b.30107

Mills-Koonce, W. R., Garrett-Peters, P., Barnett, M., Granger, D. A., Blair, C., & Cox, M. J. (2011). Father contributions to cortisol responses in infancy and toddlerhood. Developmental Psychology, 47(2), 388.

Mills-Koonce, W. R., Wagner, N. J., Willoughby, M. T., Stifter, C., Blair, C., Granger, D. A., & Family Life Project Key, I. (2015). Greater fear reactivity and psychophysiological hyperactivity among infants with later conduct problems and callous-unemotional traits. Journal of child Psychology and Psychiatry, 56(2), 147-154. doi:10.1111/jcpp.12289

Mills-Koonce, W. R., Willoughby, M. T., Garrett-Peters, P., Wagner, N., Vernon-Feagans, L., & Family Life Project Key Investigators. (2016). The interplay among socioeconomic status, household chaos, and parenting in the prediction of child conduct problems and callous–unemotional behaviors. Development and Psychopathology, 28(3), 757-771.

Milner, J. S., & Crouch, J. L. (1997). Impact and detection of response distortions on parenting measures used to assess risk for child physical abuse. Journal of Personality Assessment, 69(3), 633-650.

Miner, J. L., & Clarke-Stewart, K. A. (2008). Trajectories of externalizing behavior from age 2 to age 9: Relations with gender, temperament, ethnicity, parenting, and rater. Developmental Psychology, 44(3), 771-786. doi:10.1037/0012-1649.44.3.771

Mistler, S. A., & Enders, C. K. (2017). A Comparison of Joint Model and Fully Conditional Specification Imputation for Multilevel Missing Data. Journal of Educational and Behavioral Statistics, 42(4), 432-466.

Mitchell, C., McLanahan, S., Brooks-Gunn, J., Garfinkel, I., Hobcraft, J., & Notterman, D. (2013). Genetic differential sensitivity to social environments: Implications for research. American Journal of Public Health, 103(S1), S102-S110.

Mitchell, C., Schneper, L. M., & Notterman, D. A. (2016). DNA methylation, early life environment, and health outcomes. Pediatric Research, 79(1-2), 212-219.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. Archives of General Psychiatry, 62(5), 473-481. doi:10.1001/archpsyc.62.5.473

Möller, E. L., Nikolic, M., Majdandzic, M., & Bogels, S. M. (2016). Associations between maternal and paternal parenting behaviors, anxiety and its precursors in early childhood: A meta-analysis. Clinical Psychology Review, 45, 17-33. doi:10.1016/j.cpr.2016.03.002

Monroe, S. M., & Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: implications for the depressive disorders. Psychological Bulletin, 110(3), 406.doi:10.1037/0033-2909.110.3.406

Moore, S. R., & Depue, R. A. (2016). Neurobehavioral foundation of environmental reactivity. Psychological Bulletin, 142(2), 107-164. doi:10.1037/bul0000028

Morris Amanda, S., Silk Jennifer, S., Steinberg, L., Sessa Frances, M., Avenevoli, S., & Essex Marilyn, J. (2004). Temperamental Vulnerability and Negative Parenting as Interacting Predictors of Child Adjustment. Journal of Marriage and Family, 64(2), 461-471. doi:10.1111/j.1741-3737.2002.00461.x

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Statistics in Medicine, 38(11), 2074-2102

Morsbach, S. K., & Prinz, R. J. (2006). Understanding and improving the validity of self-report of parenting. Clinical Child and Family Psychology Review, 9(1), 1-21.

Moskowitz, A. L., Krull, J. L., Trickey, K. A., & Chorpita, B. F. (2017). Quality vs. quantity: Assessing behavior change over time. Journal of Psychopathology and Behavioral Assessment, 39(3), 514-533.

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.

Muris, P., & Ollendick, T. H. (2005). The Role of Temperament in the Etiology of Child Psychopathology. Clinical Child and Family Psychology Review, 8(4), 271-289. doi:10.1007/s10567-005-8809-y

Murray, D. W., Kollins, S. H., Hardy, K. K., Abikoff, H. B., Swanson, J. M., Cunningham, C., . . . Chuang, S. Z. (2007). Parent versus Teacher Ratings of Attention-Deficit/Hyperactivity Disorder Symptoms in the Preschoolers with Attention-Deficit/Hyperactivity Disorder Treatment Study (PATS). Journal of Child and Adolescent Psychopharmacology, 17(5), 605-619. doi:10.1089/cap.2007.0060

Myers, K., & Winters, N. C. (2002). Ten-Year Review of Rating Scales. II: Scales for Internalizing Disorders. Journal of the American Academy of Child & Adolescent Psychiatry, 41(6), 634-659. doi:10.1097/00004583-200206000-00004

Naaijen, J., Bralten, J., Poelmans, G., consortium, I., Glennon, J. C., Franke, B., & Buitelaar, J. K. (2017). Glutamatergic and GABAergic gene sets in attention-deficit/hyperactivity disorder: association to overlapping traits in ADHD and autism. Translational Psychiatry, 7(1), e999. doi:10.1038/tp.2016.273

National Longitudinal Survey of Children and Youth (2005). Cycle 5 Survey Instruments 2002/03: Book 2 – Kindergarten and Youth Questionnaires. Ottawa, Ontario: Social Development Canada, Statistics Canada.

NICHD Early Child Care Research Network. (1999). Child care and mother-child interaction in the first three years of life. Developmental Psychology, 35, 1399–1413.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), 175-220.

Nigg, J. T. (2006). Temperament and developmental psychopathology. Journal of Child Psychology and Psychiatry, 47(3-4), 395-422. doi:10.1111/j.1469-7610.2006.01612.x

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2007). Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: results from the National Comorbidity Survey Replication. Journal of child Psychology and Psychiatry, 48(7), 703-713. doi:10.1111/j.1469-7610.2007.01733.x

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.

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2008). Female and male antisocial trajectories: from childhood origins to adult outcomes. Development and Psychopathology, 20(2), 673-716. doi:10.1017/S0954579408000333

*Olofsdotter, S., Aslund, C., Furmark, T., Comasco, E., & Nilsson, K. W. (2018). Differential susceptibility effects of oxytocin gene (OXT) polymorphisms and perceived parenting on social anxiety among adolescents. Development and Psychopathology, 30(2), 449-459. doi:10.1017/S0954579417000967

Panja, D., & Bramham, C. R. (2014). BDNF mechanisms in late LTP formation: A synthesis and breakdown. Neuropharmacology, 76 Pt C, 664-676. doi:10.1016/j.neuropharm.2013.06.024

Paterson, G., & Sanson, A. (1999). The Association of Behavioural Adjustment to Temperament, Parenting and Family Characteristics among 5-Year-Old Children. Social Development, 8(3), 293-309. doi:10.1111/1467-9507.00097

Patterson, G. R. (1982). Coercive family process (Vol. 3): Castalia Publishing Company.

Patterson, G. R. (2002). The early development of coercive family process. In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.), Antisocial behavior in children and adolescents: A developmental analysis and model for intervention (pp. 25-44). Washington, DC, US: American Psychological Association.

Patterson, G. R., & Forgatch, M. S. (1995). Predicting future clinical adjustment from treatment outcome and process variables. Psychological Assessment, 7(3), 275-285. doi:10.1037/1040-3590.7.3.275

Paulussen-Hoogeboom, M. C., Stams, G. J. J., Hermanns, J., & Peetsma, T. T. (2007). Child negative emotionality and parenting from infancy to preschool: A meta-analytic review. Developmental Psychology, 43(2), 438.

Pawlak, C. R., Ho, Y. J., & Schwarting, R. K. (2008). Animal models of human psychopathology based on individual differences in novelty-seeking and anxiety. Neuroscience & Biobehavioral Reviews, 32(8), 1544-1568.

Pettit, G. S., Laird, R. D., Dodge, K. A., Bates, J. E., & Criss, M. M. (2001). Antecedents and behavior-problem outcomes of parental monitoring and psychological control in early adolescence. Child Development, 72(2), 583-598. doi:doi:10.1111/1467-8624.00298

Piko, B. F., & Balazs, M. A. (2012). Control or involvement? Relationship between authoritative parenting style and adolescent depressive symptomatology. European Child & Adolescent Psychiatry, 21(3), 149-155. doi:10.1007/s00787-012-0246-0

Pinheiro, J. C., & Bates, D. M. (2000). Mixed-Effects Models in S and S-PLUS.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2018). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1.137. https://CRAN.R-project.org/package=nlme

Pinquart, M. (2016a). Associations of Parenting Dimensions and Styles with Internalizing Symptoms in Children and Adolescents: A Meta-Analysis. Marriage & Family Review, 53(7), 613-640. doi:10.1080/01494929.2016.1247761

Pinquart, M. (2016b). Associations of Parenting Styles and Dimensions with Academic Achievement in Children and Adolescents: A Meta-analysis. Educational Psychology Review, 28(3), 475-493. doi:10.1007/s10648-015-9338-y

Pinquart, M. (2017). Associations of parenting dimensions and styles with externalizing problems of children and adolescents: An updated meta-analysis. Developmental Psychology, 53(5), 873. doi:10.1037/dev0000295

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on “environmental” measures. Behavioral and Brain Sciences, 14(3), 373-386.

Plomin, R., & Rutter, M. (1998). Child Development, Molecular Genetics, and What to Do with Genes Once They Are Found. Child Development, 69(4), 1223-1242. doi:10.1111/j.1467-8624.1998.tb06169.x

Plomin, R., DeFries, J. C., & Fulker, D. W. (2006). Nature and nurture during infancy and early childhood. Cambridge University Press.

Plomin, R., DeFries, J. C., & McClearn, G. E. (2008). Behavioral Genetics: Macmillan.

Pluess, M. (2015). Individual differences in environmental sensitivity. Child Development Perspectives, 9(3), 138-143.

Pluess, M., & Belsky, J. (2013). Vantage sensitivity: Individual differences in response to positive experiences. Psychological Bulletin, 139(4), 901-916.

Pluess, M., & Boniwell, I. (2015). Sensory-Processing Sensitivity predicts treatment response to a school-based depression prevention program: Evidence of Vantage Sensitivity. Personality and Individual Differences, 82, 40-45. doi:10.1016/j.paid.2015.03.011

Pluess, M., Assary, E., Lionetti, F., Lester, K. J., Krapohl, E., Aron, E. N., & Aron, A. (2018). Environmental sensitivity in children: Development of the Highly Sensitive Child Scale and identification of sensitivity groups. Developmental Psychology, 54(1), 51. doi:10.1037/dev0000406

Porges, S. W. (1992). Vagal tone: a physiologic marker of stress vulnerability. Pediatrics, 90(3), 498-504.

Porges, S. W. (1995). Orienting in a Defensive World - Mammalian Modifications of Our Evolutionary Heritage - a Polyvagal Theory. Psychophysiology, 32(4), 301-318. doi:10.1111/j.1469-8986.1995.tb01213.x

Posthumus, J. A., Bocker, K. B., Raaijmakers, M. A., Van Engeland, H., & Matthys, W. (2009). Heart rate and skin conductance in four-year-old children with aggressive behavior. Biological Psychology, 82(2), 164-168. doi:10.1016/j.biopsycho.2009.07.003

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. Journal of Educational and Behavioral Statistics, 31(4), 437-448. doi:10.3102/10769986031004437

Prior, M., Sanson, A. V. & Oberklaid, F. (1989). The Australian Temperament Project. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), Temperament in Childhood. Chichester, England.

Prosser, J., Hughes, C. W., Sheikha, S., Kowatch, R. A., Kramer, G. L., Rosenbarger, N., . . . Petty, F. (1997). Plasma GABA in children and adolescents with mood, behavior, and comorbid mood and behavior disorders: a preliminary study. Journal of Child and Adolescent Psychopharmacology, 7(3), 181-199. doi:10.1089/cap.1997.7.181

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. Journal of Personality Assessment, 87(1), 102-112.

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014). Development and assessment of short and very short forms of the Infant Behavior Questionnaire–Revised. Journal of Personality Assessment, 96(4), 445-458.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabinowitz, J. A., & Drabick, D. A. (2017). Do children fare for better and for worse? Associations among child features and parenting with child competence and symptoms. Developmental Review, 45, 1-30. doi:http://dx.doi.org/10.1016/j.dr.2017.03.001

*Rabinowitz, J. A., Drabick, D. A. G., Reynolds, M. D., Clark, D. B., & Olino, T. M. (2016). Child temperamental flexibility moderates the relation between positive parenting and adolescent adjustment. Journal of Applied Developmental Psychology, 43, 43-53. doi:http://dx.doi.org/10.1016/j.appdev.2015.12.006

Raine, A. (2002). Annotation: the role of prefrontal deficits, low autonomic arousal, and early health factors in the development of antisocial and aggressive behavior in children. Journal of child Psychology and Psychiatry, 43(4), 417-434. doi:10.1111/1469-7610.00034

Rapee, R. M. (1997). Potential role of childrearing practices in the development of anxiety and depression. Clinical Psychology Review, 17(1), 47-67. doi:https://doi.org/10.1016/S0272-7358(96)00040-2

Rast, P., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: Simulation results based on actual longitudinal studies. Psychological Methods, 19(1), 133-154.

Raver, C. C., Blair, C., & Willoughby, M. (2013). Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility. Developmental Psychology, 49(2), 292.

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. Biological Reviews, 82(2), 291-318.

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: results of a 24-year longitudinal study. Social Psychiatry and Psychiatric Epidemiology, 46(12), 1233-1241. doi:10.1007/s00127-010-0297-9

Rhee, S. H., Lahey, B. B., & Waldman, I. D. (2015). Comorbidity Among Dimensions of Childhood Psychopathology: Converging Evidence from Behavior Genetics. Child Development Perspectives, 9(1), 26-31. doi:10.1111/cdep.12102

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. Psychological Bulletin, 112(3), 485-499.

Rivera, M., Gutierrez, B., Molina, E., Torres-Gonzalez, F., Bellon, J. A., Moreno-Kustner, B., . . . Cervilla, J. A. (2009). High-activity variants of the uMAOA polymorphism increase the risk for depression in a large primary care sample. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150B(3), 395-402. doi:10.1002/ajmg.b.30829

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. Psychological Bulletin, 126(1), 3-25.

Rohner, R. P. (2004). The Parental "Acceptance-Rejection Syndrome": Universal Correlates of Perceived Rejection. American Psychologist, 59(8), 830-840. doi:10.1037/0003-066X.59.8.830

Roisman, G. I., & Fraley, R. C. (2012). A behavior-genetic study of the legacy of early caregiving experiences: Academic skills, social competence, and externalizing behavior in kindergarten. Child Development, 83(2), 728-742.

Roisman, G. I., Booth-Laforce, C., Belsky, J., Burt, K. B., & Groh, A. M. (2013). Molecular-genetic correlates of infant attachment: A cautionary tale. Attachment & Human Development, 15(4), 384-406.

*Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis–stress: Recommendations for evaluating interaction effects. Development and Psychopathology, 24(2), 389-409. doi:10.1017/S0954579412000065

Rose-Krasnor, L. (1997). The Nature of Social Competence: A Theoretical Review. Social Development, 6(1), 111-135. doi:10.1111/j.1467-9507.1997.tb00097.x

Rothbart, M. K. (2007). Temperament, Development, and Personality. Current Directions in Psychological Science, 16(4), 207-212. doi:10.1111/j.1467-8721.2007.00505.x

Rothbart, M., & Bates, J. (2006). Temperament. Handbook of child psychology: Vol. 3, Social, emotional, and personality development (pp. 99–166). Hoboken. In: NJ: Wiley.

Rothbart, M., & Derryberry, D. (1981). Development of individual differences in temperament. Advances in Developmental Psychology. 1, 37-86.

Rothbaum, F., & Weisz, J. R. (1994). Parental caregiving and child externalizing behavior in nonclinical samples: A meta-analysis. Psychological Bulletin, 116(1), 55-74. doi:10.1037/0033-2909.116.1.55

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688-701.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. The Annals of Applied Statistics, 2(3), 808-840.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.

Rubin, K. H., Begle, A. S., & McDonald, K. L. (2012). Peer relations and social competence in childhood. Developmental social neuroscience and childhood brain insult: Theory and practice, 23-44.

Rubin, K. H., Burgess, K. B., & Hastings, P. D. (2002). Stability and Social-Behavioral Consequences of Toddlers' Inhibited Temperament and Parenting Behaviors. Child Development, 73(2), 483-495. doi:10.1111/1467-8624.00419

Rubin, K. H., Coplan, R. J., & Bowker, J. C. (2009). Social withdrawal in childhood. Annual Review of Psychology, 60, 141-171. doi:10.1146/annurev.psych.60.110707.163642

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: preliminary findings. Journal of Child Psychology and Psychiatry, 8(1), 1-11.

Rutter, M. (1981). Stress, coping and development: some issues and some questions. Journal of Child Psychology and Psychiatry, 22(4), 323-356. doi:doi:10.1111/j.1469-7610.1981.tb00560.x

Rutter, M. (1989). Pathways from childhood to adult life. Journal of Child Psychology and Psychiatry, 30(1), 23-51. doi:10.1111/j.1469-7610.1989.tb00768.x

Rutter, M. (2006). Genes and behavior: Nature-nurture interplay explained. Malden: Blackwell Publishing.

Rutter, M. (2007). Gene–environment interdependence. Developmental Science, 10(1), 12-18.

Rutter, M., Dunn, J., Plomin, R., Simonoff, E., Pickles, A., Maughan, B., ... & Eaves, L. (1997). Integrating nature and nurture: Implications of person–environment correlations and interactions for developmental psychopathology. Development and Psychopathology, 9(2), 335-364.

Rutter, M., Silberg, J., O'Connor, T., & Simonoff, E. (1999). Genetics and child psychiatry: II Empirical research findings. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40(1), 19-55.

Rutter, M., Tizard, J., & Whitmore, K. (1970). Education, health and behaviour. Longman Publishing Group.

Saarni, C. (1990). Emotional competence: How emotions and relationships become integrated. Lincoln, NE, US: University of Nebraska Press.

Sameroff, A. (1975). Transactional Models in Early Social-Relations. Human Development, 18(1-2), 65-79. doi:10.1159/000271476

Sameroff, A. J. (1983). Development systems: Contexts and evolution. Handbook of child psychology: formerly Carmichael's Manual of child psychology/Paul H. Mussen, editor.

Sameroff, A. J. (2000). Developmental systems and psychopathology. Development and Psychopathology, 12(3), 297-312.

Sanders, M. R. (2008). Triple P-Positive Parenting Program as a public health approach to strengthening parenting. Journal of Family Psychology, 22(4), 506-517. doi:10.1037/0893-3200.22.3.506

Sanson, A. V., Smart, D., Prior, M., Oberklaid, F., & Pedlow, R. (1994). The structure of temperament from age 3 to 7 years: Age, sex, and sociodemographic influences. Merrill-Palmer Quarterly (1982-), 233-252.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. Social Development, 13(1), 142-170. doi:10.1046/j.1467-9507.2004.00261.x

Sanson, A., Prior, M., Garino, E., Oberklaid, F., & Sewell, J. (1987). The structure of infant temperament: Factor analysis of the Revised Infant Temperament Questionnaire. Infant Behavior and Development, 10(1), 97-104.

Saudino, K. J., & Micalizzi, L. (2015). Emerging Trends in Behavioral Genetic Studies of Child Temperament. Child Development Perspectives, 9(3), 144-148. doi:10.1111/cdep.12123

Schafer, J. L., & Yucel, R. M. (2002). Computational Strategies for Multivariate Linear Mixed-Effects Models With Missing Values. Journal of Computational and Graphical Statistics, 11(2), 437-457. doi:10.1198/106186002760180608

Schmitz, M., Denardin, D., Silva, T. L., Pianca, T., Roman, T., Hutz, M. H., ... & Rohde, L. A. (2006). Association between alpha-2a-adrenergic receptor gene and ADHD inattentive type. Biological Psychiatry, 60(10), 1028-1033. doi:10.1016/j.biopsych.2006.02.035

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. Stata Journal, 13(1), 65-76.

Segrin, C., McNelis, M., & Swiatkowski, P. (2016). Social Skills, Social Support, and Psychological Distress: A Test of the Social Skills Deficit Vulnerability Model. Human Communication Research, 42(1), 122-137. doi:10.1111/hcre.12070

Seifer, R., Sameroff, A. J., Barrett, L. C., & Krafchuk, E. (1994). Infant temperament measured by multiple observations and mother report. Child development, 65(5), 1478-1490.

Seifer, R., Sameroff, A., Dickstein, S., Schiller, M., & Hayden, L. C. (2004). Your own children are special: Clues to the sources of reporting bias in temperament assessments. Infant Behavior and Development, 27(3), 323-341.

Sheikh, H. I., Kryski, K. R., Smith, H. J., Dougherty, L. R., Klein, D. N., Bufferd, S. J., ... & Hayden, E. P. (2013). Catechol-O-methyltransferase gene val158met polymorphism and depressive symptoms during early childhood. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 162(3), 245-252. doi:10.1002/ajmg.b.32141

Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous variables power and sample size considerations. Organizational Research Methods, 12(3), 510-528.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. Journal of Consulting and Clinical Psychology, 78(3), 298-311.

Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., & Zentner, M. (2012). What Is Temperament Now? Assessing Progress in Temperament Research on the Twenty-Fifth Anniversary of Goldsmith et al. (1987). Child Development Perspectives, 6(4), 436-444.

Siddons, H., & Lancaster, S. (2004). Overview of the Use of the Child Behavior Checklist within Australia, An [online]. Camberwell, Vic.

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(11), 1359-1366.

*Slagt, M., Dubas, J. S., & van Aken, M. A. G. (2016b). Differential Susceptibility to parenting in middle childhood: Do impulsivity, effortful control and negative emotionality indicate susceptibility or vulnerability? Infant and Child Development, 25(4), 302-324. doi:10.1002/icd.1929

Slagt, M., Dubas, J. S., Deković, M., & van Aken, M. A. (2016a). Differences in sensitivity to parenting depending on child temperament: A meta-analysis. Psychological Bulletin, 142(10), 1068.

Slagt, M., Dubas, J. S., Ellis, B. J., Van Aken, M. A., & Deković, M. (2019). Linking emotional reactivity "for better and for worse" to differential susceptibility to parenting among kindergartners. Development and Psychopathology, 31(2), 741-758.

*Slagt, M., Dubas, J. S., van Aken, M. A. G., Ellis, B. J., & Deković, M. (2018). Sensory processing sensitivity as a marker of differential susceptibility to parenting. Developmental Psychology, 54(3), 543-558. doi:10.1037/dev0000431

Smoller, J. W., Yamaki, L. H., Fagerness, J. A., Biederman, J., Racette, S., Laird, N. M., . . . Sklar, P. B. (2005). The corticotropin-releasing hormone gene and behavioral inhibition in children at risk for panic disorder. Biological Psychiatry, 57(12), 1485-1492. doi:10.1016/j.biopsych.2005.02.018

Smyth, C. (2016). Getting ahead in the preschool years: An analysis of a preschool enrichment and entertainment market. Sociology, 50(4), 731-747.

Soenens, B., & Vansteenkiste, M. (2010). A theoretical upgrade of the concept of parental psychological control: Proposing new insights on the basis of self-determination theory. Developmental Review, 30(1), 74-99. doi:10.1016/j.dr.2009.11.001

Soenens, B., Elliot, A. J., Goossens, L., Vansteenkiste, M., Luyten, P., & Duriez, B. (2005). The intergenerational transmission of perfectionism: parents' psychological control as an intervening variable. Journal of Family Psychology, 19(3), 358-366. doi:10.1037/0893-3200.19.3.358

Solt, F., & Hu, Y. (2016). interplot: Plot the Effects of Variables in Interaction Terms. R package version 0.1.5. https://CRAN.R-project.org/package=interplot

Solt, F., and Hu, Y. (2018). interplot: Plot the Effects of Variables in Interaction Terms. R package version 0.2.1. https://CRAN.R-project.org/package=interplot

Spera, C. (2005). A review of the relationship among parenting practices, parenting styles, and adolescent school achievement. Educational Psychology Review, 17(2), 120-146. doi:10.1007/s10648-005-3950-1

Spokas, M., & Heimberg, R. G. (2008). Overprotective Parenting, Social Anxiety, and External Locus of Control: Cross-sectional and Longitudinal Relationships. Cognitive Therapy and Research, 33(6), 543. doi:10.1007/s10608-008-9227-5

Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. Journal of Clinical Psychology, 63(7), 611-631.

Stams, G. J. J. M., Juffer, F., & van Ijzendoorn, M. H. (2002). Maternal sensitivity, infant attachment, and temperament in early childhood predict adjustment in middle childhood: The case of adopted children and their biologically unrelated parents. Developmental Psychology, 38(5), 806-821. doi:10.1037//0012-1649.38.5.806

Stattin, H., & Kerr, M. (2000). Parental monitoring: a reinterpretation. Child Development, 71(4), 1072-1085. doi:doi:10.1111/1467-8624.00210

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. Perspectives on Psychological Science, 11(5), 702-712. doi:10.1177/1745691616658637

Steinberg, L., Lamborn, S. D., Darling, N., Mounts, N. S., & Dornbusch, S. M. (1994). Over-time changes in adjustment and competence among adolescents from authoritative, authoritarian, indulgent, and neglectful families. Child Development, 65(3), 754-770.

*Stocker, C. M., Masarik, A. S., Widaman, K. F., Reeb, B. T., Boardman, J. D., Smolen, A., . . . Conger, K. J. (2017). Parenting and adolescents' psychological adjustment: Longitudinal moderation by adolescents' genetic sensitivity. Development and Psychopathology, 29(4), 1289-1304. doi:10.1017/S0954579416001310

*Stoltz, S., Beijers, R., Smeekens, S., & Dekovic, M. (2017). Diathesis stress or differential susceptibility? testing longitudinal associations between parenting, temperament, and children's problem behavior. Social Development, 26(4), 783-796. doi:10.1111/sode.12237

Stoop, R. (2012). Neuromodulation by oxytocin and vasopressin. Neuron, 76(1), 142-159. doi:10.1016/j.neuron.2012.09.025

Stormshak, E. A., Bierman, K. L., McMahon, R. J., & Lengua, L. J. (2000). Parenting Practices and Child Disruptive Behavior Problems in Early Elementary School. Journal of Clinical Child Psychology, 29(1), 17-29. doi:10.1207/S15374424jccp2901_3

*Sulik, M. J., Eisenberg, N., Lemery-Chalfant, K., Spinrad, T. L., Silva, K. M., Eggum, N. D., ... & Stover, D. A. (2012). Interactions between serotonin transporter gene haplotypes and quality of mothers' parenting predict the development of children's noncompliance. Developmental Psychology, 48(3), 740-754. doi:10.1037/a0025938

*Sulik, M. J., Eisenberg, N., Spinrad, T. L., Lemery-Chalfant, K., Swann, G., Silva, K. M., . . . Verrelli, B. C. (2015). Interactions among catechol-O-methyltransferase genotype, parenting, and sex predict children's internalizing symptoms and inhibitory control: Evidence for differential susceptibility. Development and Psychopathology, 27(3), 709-723. doi:10.1017/S0954579414000807

Super, C. M., Axia, G., Harkness, S., Welles-Nystrom, B., Zylicz, P. O., Parmar, P., ... & Palacios, J. (2008). Culture, temperament, and the “difficult child”: A study in seven western cultures. International Journal of Developmental Science, 2(1-2), 136-157.

Sweitzer, M. M., Halder, I., Flory, J. D., Craig, A. E., Gianaros, P. J., Ferrell, R. E., & Manuck, S. B. (2013). Polymorphic variation in the dopamine D4 receptor predicts delay discounting as a function of childhood socioeconomic status: evidence for differential susceptibility. Social Cognitive and Affective Neuroscience, 8(5), 499-508. doi:10.1093/scan/nss020

Symeou, M., & Georgiou, S. (2017). Externalizing and internalizing behaviours in adolescence, and the importance of parental behavioural and psychological control practices. Journal of Adolescence, 60, 104-113. doi:10.1016/j.adolescence.2017.07.007

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. Journal of Affective Disorders, 61(3), 201-216. doi:https://doi.org/10.1016/S0165-0327(00)00338-4

The 1000 Genomes Project, C., Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., . . . Abecasis, G. R. (2015). A global reference for human genetic variation. Nature, 526, 68. doi:10.1038/nature15393

Thomas, A. (1968). Temperament and behavior disorders in children.

Thomas, A., & Chess, S. (1977). Temperament and development: Brunner/Mazel.

Timmons, A. C., & Preacher, K. J. (2015). The importance of temporal design: How do measurement intervals affect the accuracy and efficiency of parameter estimates in longitudinal research?. Multivariate Behavioral Research, 50(1), 41-55.

Todd, R. M., Muller, D. J., Lee, D. H., Robertson, A., Eaton, T., Freeman, N., . . . Anderson, A. K. (2013). Genes for emotion-enhanced remembering are linked to enhanced perceiving. Psychological Science, 24(11), 2244-2253. doi:10.1177/0956797613492423

Torquemada, Y. F., Durako, M. J., & Lizaso, J. L. S. (2005). Effects of salinity and possible interactions with temperature and pH on growth and photosynthesis of Halophila johnsonii Eiseman. Marine Biology, 148(2), 251-260.

* Trucco, E. M., Villafuerte, S., Heitzeg, M. M., Burmeister, M., & Zucker, R. A. (2016). Susceptibility effects of GABA receptor subunit alpha-2 (GABRA2) variants and parental monitoring on externalizing behavior trajectories: Risk and protection conveyed by the minor allele. Development and psychopathology, 28(1), 15-26. doi:10.1017/S0954579415000255

Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. Molecular psychiatry, 15(1), 18.

Ungar, M. (2009). Overprotective Parenting: Helping Parents Provide Children the Right Amount of Risk and Responsibility. The American Journal of Family Therapy, 37(3), 258-271. doi:10.1080/01926180802534247

Ungar, M. (2018). The differential impact of social services on young people's resilience. Child Abuse & Neglect, 78, 4-12. doi:10.1016/j.chiabu.2017.09.024

US Department of Education. (2001). Early Childhood Longitudinal Study, birth cohort. Washington, DC: National Centre for Education Statistics.

Vaillancourt, T., Brittain, H. L., McDougall, P., & Duku, E. (2013). Longitudinal links between childhood peer victimization, internalizing and externalizing problems, and academic functioning: developmental cascades. Journal of Abnormal Child Psychology, 41(8), 1203-1215. doi:10.1007/s10802-013-9781-5

van Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. Journal of Advanced Nursing, 20(1), 196-201.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67.

van den Boom, D. C. (1994). The influence of temperament and mothering on attachment and exploration: An experimental manipulation of sensitive responsiveness among lower-class mothers with irritable infants. Child Development, 65(5), 1457-1477.

* Van der Graaff, J., Meeus, W., de Wied, M., van Boxtel, A., van Lier, P., & Branje, S. (2016). Respiratory sinus arrhythmia moderates the relation between parent-adolescent relationship quality and adolescents' social adjustment. Journal of Abnormal Child Psychology, 44(2), 269-281. doi:10.1007/s10802-015-9989-7

van der Linden, W. J. (2018). Handbook of item response theory, three volume set. Chapman and Hall/CRC.

van Ijzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2015). Genetic differential susceptibility on trial: meta-analytic support from randomized controlled experiments. Development and Psychopathology, 27(1), 151-162. doi:10.1017/S0954579414001369

Vernon-Feagans, L., Cox, M., Willoughby, M., Burchinal, M., Garrett-Peters, P., Mills-Koonce, R., ... & Bauer, P. J. (2013). The Family Life Project: An epidemiological and developmental study of young children living in poor rural communities. Monographs of the Society for Research in Child Development, i-150.

Viding, E., Fontaine, N. M., Oliver, B. R., & Plomin, R. (2009). Negative parental discipline, conduct problems and callous–unemotional traits: monozygotic twin differences study. The British Journal of Psychiatry, 195(5), 414-419.

Villafuerte, S., Heitzeg, M. M., Foley, S., Yau, W. W., Majczenko, K., Zubieta, J. K., ... & Burmeister, M. (2012). Impulsiveness and insula activation during reward anticipation are associated with genetic variants in GABRA2 in a family sample enriched for alcoholism. Molecular Psychiatry, 17(5), 511-519. doi:10.1038/mp.2011.33

Villafuerte, S., Strumba, V., Stoltenberg, S. F., Zucker, R. A., & Burmeister, M. (2013). Impulsiveness mediates the association between GABRA2 SNPs and lifetime alcohol problems. Genes, Brain and Behavior, 12(5), 525-531.doi:10.1111/gbb.12039

Vincent, C. (2017). 'The children have only got one education and you have to make sure it's a good one': parenting and parent–school relations in a neoliberal age. Gender and Education, 29(5), 541-557. doi:10.1080/09540253.2016.1274387

von Hippel, P. T. (2009). 8. How to Impute Interactions, Squares, and other Transformed Variables. Sociological Methodology, 39(1), 265–291.

Wachs, T. D., & Gandour, M. J. (1983). Temperament, environment, and six-month cognitive-intellectual development: A test of the organismic specificity hypothesis. International Journal of Behavioral Development, 6(2), 135-152. doi:10.1177/016502548300600202

Wachs, T. D., & Plomin, R. E. (1991). Conceptualization and measurement of organism-environment interaction. American Psychological Association.

*Wagner, N. J., Mills-Koonce, W. R., Willoughby, M. T., Cox, M. J., & Family Life Project Key, I. (2017). Parenting and Cortisol in Infancy Interactively Predict Conduct Problems and Callous-Unemotional Behaviors in Childhood. Child Development. doi:10.1111/cdev.12900

Waller, R., Gardner, F., Dishion, T., Sitnick, S. L., Shaw, D. S., Winter, C. E., & Wilson, M. (2015). Early Parental Positive Behavior Support and Childhood Adjustment: Addressing Enduring Questions with New Methods. Social Development, 24(2), 304-322. doi:10.1111/sode.12103

Wang, M., & Liu, L. (2018). Reciprocal Relations Between Harsh Discipline and Children's Externalizing Behavior in China: A 5-Year Longitudinal Study. Child Development, 89(1), 174-187. doi:10.1111/cdev.12724

Way, B. M., Taylor, S. E., & Eisenberger, N. I. (2009). Variation in the μ-opioid receptor gene (OPRM1) is associated with dispositional and neural sensitivity to social rejection. Proceedings of the National Academy of Sciences, 106(35), 15079-15084.

Weeland, J., Overbeek, G., de Castro, B. O., & Matthys, W. (2015). Underlying Mechanisms of Gene-Environment Interactions in Externalizing Behavior: A Systematic Review and Search for Theoretical Mechanisms. Clinical Child and Family Psychology Review, 18(4), 413-442. doi:10.1007/s10567-015-0196-4

Wells, G., Shea, B., O'connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2015). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute, 2014.

West, D. W., & Phillips, S. M. (2010). Anabolic processes in human skeletal muscle: restoring the identities of growth hormone and testosterone. The Physician and Sportsmedicine, 38(3), 97-104. doi:10.3810/psm.2010.10.1814

Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. Journal of Family Psychology, 19(1), 111.

Whittle, S., Allen, N. B., Lubman, D. I., & Yücel, M. (2006). The neurobiological basis of temperament: Towards a better understanding of psychopathology. Neuroscience & Biobehavioral Reviews, 30(4), 511-525. doi:https://doi.org/10.1016/j.neubiorev.2005.09.003

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.

Widaman, K. F., Helm, J. L., Castro-Schilo, L., Pluess, M., Stallings, M. C., & Belsky, J. (2012). Distinguishing ordinal and disordinal interactions. Psychological Methods, 17(4), 615-622. doi:10.1037/a0030003

Wiggins, J. L., Mitchell, C., Stringaris, A., & Leibenluft, E. (2014). Developmental trajectories of irritability and bidirectional associations with maternal depression. Journal of the American Academy of Child & Adolescent Psychiatry, 53(11), 1191-1205.

Willoughby, M., Burchinal, M., Garrett-Peters, P., Mills-Koonce, R., Vernon-Feagans, L., & Cox, M. (2013). The Family Life Project: An epidemiological and developmental study of young children living in poor rural communities: II. Recruitment of the Family Life Project sample. Monographs of the Society for Research in Child Development.

Wilson, M. N. (2006). Future Visions of the Black Caucus of Srcd. Monographs of the Society for Research in Child Development, 71(1), 214-217. doi:10.1111/j.1540-5834.2006.00374.x

Wochnik, G. M., Ruegg, J., Abel, G. A., Schmidt, U., Holsboer, F., & Rein, T. (2005). FK506-binding proteins 51 and 52 differentially regulate dynein interaction and nuclear translocation of the glucocorticoid receptor in mammalian cells. Journal of Biological Chemistry, 280(6), 4609-4616. doi:10.1074/jbc.M407498200

Wong, M. Y., Day, N. E., Luan, J. A., & Wareham, N. J. (2004). Estimation of magnitude in gene–environment interactions in the presence of measurement error. Statistics in Medicine, 23(6), 987-998.

Woo, J. M., Yoon, K. S., Choi, Y. H., Oh, K. S., Lee, Y. S., & Yu, B. H. (2004). The association between panic disorder and the L/L genotype of catechol-O-methyltransferase. Journal of Psychiatric Research, 38(4), 365-370. doi:10.1016/j.jpsychires.2004.01.001

Wood Jeffrey, J., McLeod Bryce, D., Sigman, M., Hwang, W.-C., & Chu Brian, C. (2002). Parenting and childhood anxiety: theory, empirical findings, and future directions. Journal of Child Psychology and Psychiatry, 44(1), 134-151. doi:10.1111/1469-7610.00106

Wu, C., Shi, X., Cui, Y., & Ma, S. (2015). A penalized robust semiparametric approach for gene–environment interactions. Statistics in Medicine, 34(30), 4016-4030.

Yamagata, S., Takahashi, Y., Ozaki, K., Fujisawa, K. K., Nonaka, K., & Ando, J. (2013). Bidirectional influences between maternal parenting and children's peer problems: a longitudinal monozygotic twin difference study. Developmental Science, 16(2), 249-259.

Young, S. E., Smolen, A., Corley, R. P., Krauter, K. S., DeFries, J. C., Crowley, T. J., & Hewitt, J. K. (2002). Dopamine transporter polymorphism associated with externalizing behavior problems in children. American Journal of Medical Genetics, 114(2), 144-149. doi:10.1002/ajmg.10155

Yu, Y. W., Yang, C. W., Wu, H. C., Tsai, S. J., Hong, C. J., Chen, M. C., & Chen, T. J. (2005). Association study of a functional MAOA-uVNTR gene polymorphism and personality traits in Chinese young females. Neuropsychobiology, 52(3), 118-121. doi:10.1159/000087556

Zai, C. C., Ehtesham, S., Choi, E., Nowrouzi, B., de Luca, V., Stankovich, L., . . . Beitchman, J. H. (2012). Dopaminergic system genes in childhood aggression: possible role for DRD2. The World Journal of Biological Psychiatry, 13(1), 65-74. doi:10.3109/15622975.2010.543431

*Zarra-Nezhad, M., Kiuru, N., Aunola, K., Zarra-Nezhad, M., Ahonen, T., Poikkeus, A.-M., . . . Nurmi, J.-E. (2014). Social withdrawal in children moderates the association between parenting styles and the children's own socioemotional development. Journal of Child Psychology and Psychiatry, 55(11), 1260-1269. doi:10.1111/jcpp.12251

Zaslow, M. J., Weinfield, N. S., Gallagher, M., Hair, E. C., Ogawa, J. R., Egeland, B., … De Temple, J. M. (2006). Longitudinal prediction of child outcomes from differing measures of parenting in a low-income sample. Developmental Psychology, 42(1), 27-37.

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. International Journal of Developmental Science, 2(1-2), 7-37.

*Zhang, L., Li, Z., Chen, J., Li, X., Zhang, J., & Belsky, J. (2016). The BDNF Val66Met polymorphism interacts with maternal parenting influencing adolescent depressive symptoms: Evidence of differential susceptibility model. Journal of Youth and Adolescence, 45(3), 471-483. doi:10.1007/s10964-015-0378-x

Zhang, Q., & Wang, L. (2017). Moderation analysis with missing data in the predictors. Psychological Methods, 22(4), 649-666.

*Zhang, W., Cao, C., Wang, M., Ji, L., & Cao, Y. (2016). Monoamine oxidase a (MAOA) and catechol-o-methyltransferase (COMT) gene polymorphisms interact with maternal parenting in association with adolescent reactive aggression but not proactive aggression: evidence of differential susceptibility. Journal of Youth and Adolescence, 45(4), 812-829. doi:10.1007/s10964-016-0442-1

*Zhang, W., Cao, Y., Wang, M., Ji, L., Chen, L., & Deater-Deckard, K. (2015). The Dopamine D2 Receptor Polymorphism (DRD2TaqIA) Interacts with Maternal Parenting in Predicting Early Adolescent Depressive Symptoms: Evidence of Differential Susceptibility and Age Differences. Journal of Youth and Adolescence, 44(7), 1428-1440. doi:10.1007/s10964-015-0297-x

Zhang, X., Gainetdinov, R. R., Beaulieu, J. M., Sotnikova, T. D., Burch, L. H., Williams, R. B., . . . Caron, M. G. (2005). Loss-of-function mutation in tryptophan hydroxylase-2 identified in unipolar major depression. Neuron, 45(1), 11-16. doi:10.1016/j.neuron.2004.12.014

Zhou, Q., Eisenberg, N., Losoya, S. H., Fabes, R. A., Reiser, M., Guthrie, I. K., . . . Shepard, S. A. (2002). The relations of parental warmth and positive expressiveness to children's empathy-related responding and social functioning: a longitudinal study. Child Development, 73(3), 893-915. doi:10.1111/1467-8624.00446

Zubrick, S., Taylor, C., Lawrence, D., Mitrou, F., Christensen, D., & Dalby, R. (2009). Round Table: Lifecourse Epidemiology: The development of human capability across the lifecourse: Perspectives from childhood. Australasian Epidemiologist, 16(3), 6-10.

Zuckerman, M. (1999). Diathesis-stress models. In M. Zuckerman, Vulnerability to psychopathology: A biosocial model (pp. 3-23). Washington, DC, US: American Psychological Association.

## Appendix A: Supplementary descriptive statistics for Chapter 4

Table A.1 presents the average, standard deviation, Hedges' *G* effect size (Hedge 1981) comparison of included to excluded children in Chapter 4 where positive numbers mean the excluded children scored higher and Cronbach's alpha where relevant for all variables included in the model.

**Table A.1:** Summary statistics for variables analysed in Chapter 4 by child age. For categorical variables, averages are frequencies and Hedges' *G* (Hs' *G*) is the percent difference. T refers to teacher reported information, all other is parent report. A positive Hedges' G indicates the excluded sample was higher for that measurement.

| Child age | Age 4 | | | | Age 6 | | | | Age 8 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | α | *Hs' G* | Mean | SD | α | *Hs' G* | Mean | SD | α | *Hs' G* |
| Parental warmth | 87.53 | 11.5 | 0.88 | -0.02 | 88.61 | 12.12 | 0.89 | -0.10 | 86.02 | 13.18 | 0.89 | -0.05 |
| Harsh parenting | 22.98 | 12.04 | 0.65 | 0.16 | 22.88 | 12.36 | 0.67 | 0.06 | 23.35 | 12.62 | 0.68 | 0.10 |
| Difficult | 38.09 | 11.72 | 0.71 | 0.22 | 33.47 | 11.81 | 0.71 | 0.13 | | | | |
| Introversion | 43.8 | 21.98 | 0.81 | -0.05 | 37.92 | 21.3 | 0.79 | -0.23 | | | | |
| Reactivity | 29.83 | 16.3 | 0.67 | 0.27 | 24.46 | 16.33 | 0.67 | 0.39 | | | | |
| Unpersistence | 40.65 | 16.48 | 0.78 | 0.35 | 38.02 | 17.36 | 0.79 | 0.17 | | | | |
| Conduct problems | 1.94 | 1.69 | 0.66 | 0.10 | 1.27 | 1.3 | 0.53 | 0.13 | 1.06 | 1.27 | 0.56 | 0.12 |
| Hyperactivity | 2.98 | 1.94 | 0.70 | 0.11 | 3.1 | 2.15 | 0.76 | 0.13 | 3.02 | 2.27 | 0.77 | 0.15 |
| Externalizing | 4.92 | 3.06 | 0.75 | 0.15 | 4.37 | 2.99 | 0.76 | 0.19 | 4.08 | 3.07 | 0.77 | 0.21 |
| Emotional difficulties | 3.29 | 1.35 | 0.52 | 0.07 | 3.51 | 1.54 | 0.59 | 0.03 | 3.56 | 1.67 | 0.64 | 0.04 |
| Peer problems | 1.12 | 1.26 | 0.48 | 0.21 | 1.07 | 1.3 | 0.53 | 0.13 | 1.11 | 1.34 | 0.55 | 0.20 |
| Internalizing | 4.41 | 2.12 | 0.60 | 0.21 | 4.58 | 2.31 | 0.65 | 0.12 | 4.66 | 2.46 | 0.67 | 0.18 |
| Externalizing and internalizing | 9.33 | 4.13 | 0.74 | 0.26 | 8.95 | 4.28 | 0.76 | 0.25 | 8.74 | 4.51 | 0.78 | 0.30 |
| Pro-social | 7.85 | 1.67 | 0.66 | -0.05 | 8.58 | 1.57 | 0.67 | -0.13 | 8.69 | 1.53 | 0.67 | -0.12 |
| T Conduct problems | 0.84 | 1.55 | 0.78 | 0.13 | 0.64 | 1.28 | 0.72 | 0.08 | 0.64 | 1.33 | 0.74 | 0.13 |
| T Hyperactivity | 1.96 | 2.21 | 0.83 | 0.28 | 2.45 | 2.54 | 0.86 | 0.26 | 2.37 | 2.64 | 0.87 | 0.24 |
| T Externalizing | 2.8 | 3.33 | 0.86 | 0.31 | 3.09 | 3.42 | 0.85 | 0.27 | 3.01 | 3.58 | 0.87 | 0.27 |

**Table A.1 continued:** Summary statistics for variables by child age. Summary statistics for variables by child age. For categorical variables, averages are frequencies and Hedges' *G* (Hs' *G*) is the percent difference. T refers to teacher reported information, all other is parent report. A positive Hedges' G indicates the excluded sample was higher for that measurement.

| Child age | Age 4 | | | | Age 6 | | | | Age 8 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | α | *Hs' G* | Mean | SD | α | *Hs' G* | Mean | SD | α | *Hs' G* |
| T Emotional difficulties | 2.83 | 1.38 | 0.67 | 0.14 | 3.09 | 1.61 | 0.73 | 0.06 | 3.1 | 1.73 | 0.77 | 0.02 |
| T Peer problems | 1.19 | 1.53 | 0.61 | 0.13 | 0.92 | 1.33 | 0.58 | 0.22 | 1.09 | 1.54 | 0.64 | 0.10 |
| T Internalizing | 4.02 | 2.43 | 0.72 | 0.20 | 4.01 | 2.43 | 0.73 | 0.22 | 4.19 | 2.78 | 0.78 | 0.08 |
| T Externalizing and internalizing | 6.82 | 4.71 | 0.83 | 0.39 | 7.1 | 4.8 | 0.84 | 0.38 | 7.21 | 5.24 | 0.86 | 0.27 |
| T Pro-social | 7.6 | 2.11 | 0.81 | -0.20 | 8 | 2.07 | 0.81 | -0.26 | 8.01 | 2.08 | 0.81 | -0.11 |
| % Children female | 0.52 | | | -0.05 | | | | | | | | |
| % Indigenous children | 0.02 | | | 0.02 | | | | | | | | |
| % Development delay | 0.02 | | | 0.04 | | | | | | | | |
| % Hear, sight, speech | 0.21 | | | 0.06 | | | | | | | | |
| % Physical | 0.03 | | | 0.03 | | | | | | | | |
| % Other condition | 0.12 | | | 0.02 | | | | | | | | |
| Maternal depression | 12.37 | 12.81 | 0.81 | 0.1 | 11.32 | 12.32 | 0.82 | 0.34 | 10.62 | 12.23 | 0.82 | 0.43 |
| SLE | 0.93 | 1.28 | | 0.01 | 1.4 | 1.53 | | 0.02 | 1.4 | 1.74 | | 0.05 |
| HLE | 56.7 | 17.66 | 0.70 | -0.17 | 47.98 | 16.72 | 0.66 | -0.16 | 38.22 | 16.13 | 0.63 | -0.04 |
| Education year 11 | 0.19 | | | -0.05 | | | | | | | | |
| Education year 12 | 0.38 | | | -0.09 | | | | | | | | |
| Education university | 0.43 | | | -0.1 | | | | | | | | |
| SEIFA | 1018.98 | 72.69 | | -0.83 | 1019.09 | 74.76 | | -0.67 | 1021.29 | 70.01 | | -1.09 |

**Appendix B: Supplementary descriptive statistics for Chapter 5**

Table B.1 (Longitudinal Study of Australian Children) and Table B.2 (Family Life Project) presents the average, standard deviation, Hedges' *G* effect size (Hedge 1981) comparison of included to excluded children from Chapter 5 where positive numbers mean the included children scored higher, Cronbach's alpha where relevant, and the mean and standard deviation of imputed data for all variables included in the model.

**Table B.1:** Summary statistics for the Longitudinal Study of Australian Children variables from Chapter 5. For categorical variables, averages are proportions and Hedges' *G* (Hs' *G*) is the percent difference. # values from imputed data set with pro-social as the outcome and non-approach, un-cooperation, and irritability as temperament parameters, except * which was the same model data but with difficult temperament. A positive Hedges' G indicates the included sample was higher for that measurement.

| | Included sample | | | | Excluded sample | | | | Imputed sample# | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | n | α | Mean | SD | n | *Hs' G* | Imputed mean | Imputed SD |
| Parental warmth | 90.4 | 10.39 | 3448 | 0.846 | 89.32 | 11.4 | 1000 | 0.10 | 90.39 | 10.39 |
| Hostile parenting | 23.28 | 14.39 | 2930 | 0.846 | 23.02 | 15.02 | 552 | 0.02 | 23.42 | 14.43 |
| Irritability | 29.82 | 16.09 | 3231 | 0.57 | 30.44 | 17.38 | 1002 | -0.04 | 29.79 | 16.11 |
| Un-cooperative | 36.55 | 17.51 | 3235 | 0.655 | 37.09 | 18.11 | 1017 | -0.03 | 36.53 | 17.54 |
| Non-approach | 24.97 | 16.53 | 3200 | 0.73 | 26.79 | 17.84 | 992 | -0.11 | 24.94 | 16.54 |
| Difficult | 30.42 | 12.22 | 3145 | 0.744 | 31.31 | 12.76 | 954 | -0.07 | 30.43* | 12.24 |
| Conduct problems | 2.13 | 1.78 | 3602 | 0.67 | 2.38 | 1.8 | 175 | -0.14 | | |
| Hyperactivity | 3.25 | 2.08 | 3602 | 0.726 | 3.44 | 1.94 | 157 | -0.09 | | |
| Externalizing | 5.39 | 3.28 | 3602 | 0.772 | 5.88 | 3.26 | 126 | -0.15 | | |
| Emotional difficulties | 3.4 | 1.48 | 3602 | 0.563 | 3.76 | 1.72 | 195 | -0.24 | | |
| Peer problems | 1.36 | 1.46 | 3602 | 0.535 | 1.68 | 1.48 | 193 | -0.22 | | |
| Internalizing | 4.76 | 2.42 | 3602 | 0.651 | 5.51 | 2.49 | 169 | -0.31 | | |
| Total externalizing and internalizing | 10.15 | 4.68 | 3602 | 0.779 | 11.52 | 4.78 | 87 | -0.29 | | |
| Pro-social | 7.71 | 1.76 | 3584 | 0.685 | 7.59 | 1.8 | 208 | 0.07 | 7.71 | 1.76 |
| Parental depression | 13.99 | 13.68 | 3242 | 0.819 | 16.5 | 16.32 | 1014 | -0.17 | 14.05 | 13.73 |
| Stressful life events | 9.25 | 10.23 | 3180 | 0.495 | 10.63 | 11.64 | 967 | -0.13 | 9.39 | 10.3 |

**Table B.1 continued**

| | Included sample | | | | Excluded sample | | | | Imputed sample# | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | n | α | Mean | SD | n | *Hs' G* | Imputed mean | Imputed SD |
| Financial hardship | 7.33 | 14.84 | 3590 | 0.623 | 12.34 | 19.28 | 1487 | -0.31 | 7.34 | 14.85 |
| Child age at 9 month sample (years) | 0.77 | 0.21 | 3602 | | 0.77 | 0.22 | 1505 | -0.004 | | |
| Proportion male children | 0.51 | | 3602 | | 0.51 | | 1505 | 0.00 | | |
| Proportion Indigenous children | 0.03 | | 3602 | | 0.08 | | 1505 | -0.05 | | |
| Proportion development delay | 0.01 | | 3602 | | 0.02 | | 1505 | -0.01 | | |
| Proportion hear, sight, speech | 0.02 | | 3602 | | 0.03 | | 1505 | -0.01 | | |
| Proportion physical disability | 0.005 | | 3602 | | 0.006 | | 1505 | -0.001 | | |
| Parent age at 9 month sample (years) | 32.02 | 5.16 | 3602 | | 30.31 | 6.05 | 1501 | 0.31 | | |
| Proportion education year 11 | 0.25 | | 3602 | | 0.4 | | 1496 | 0.15 | | |
| Proportion education year 12 | 0.39 | | 3602 | | 0.35 | | 1496 | 0.04 | | |
| Proportion education university | 0.36 | | 3602 | | 0.25 | | 1496 | 0.11 | | |

**Table B.2:** Summary statistics for Family Life Project analytical variables from Chapter 5. For categorical variables, averages are proportions and Hedges' *G* (Hs' *G*) is the percent difference. # values from imputed data set with social competence as the outcome, sensitivity, stimulation, positive regard, animation, detachment, negative regard, and intrusiveness as parenting parameters, and distress to limitations, fear, falling reactivity, approach, and orienting as temperament parameters, except ** which was the same model data but with negative effect for temperament, and * which used sensitive and negative-intrusive parenting. Age specific variables used for parenting, whilst all others use 15 to 24 month parenting average. A positive Hedges' G indicates the included sample was higher for that measurement.

| | Included sample | | | | Excluded sample | | | | Imputed sample# | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | n | α | Mean | SD | n | *Hs' G* | Imputed mean | Imputed SD |
| 15 months | | | | | | | | | | |
| Sensitivity | 44.57 | 24.55 | 973 | | 35.75[a] | 20.32 | 93 | 0.37 | 43.38 | 21.13 |
| Stimulation | 40.15 | 24.61 | 973 | | 29.84[a] | 22.5 | 93 | 0.37 | 39.06 | 25.42 |
| Positive regard | 47.6 | 24.96 | 973 | | 38.71[a] | 22.88 | 93 | 0.35 | 47.06 | 24.1 |
| Animation | 46.73 | 21.83 | 973 | | 37.9[a] | 24.06 | 93 | 0.24 | 43.43 | 23.61 |
| Detachment | 39.31 | 25.12 | 973 | | 52.96[a] | 25.5 | 93 | -0.25 | 46.87 | 24.5 |
| Sensitive parenting | 47.95 | 20.24 | 973 | 0.896 | 37.85[a] | 19.24 | 93 | 0.37 | 45.28* | 19.92 |
| Negative regard | 22.2 | 24.04 | 973 | | 47.85[a] | 20.07 | 93 | -0.17 | 18.69 | 21.28 |
| Intrusiveness | 46.8 | 24.63 | 973 | | 18.55[a] | 20.49 | 93 | 0.00 | 44.41 | 20.15 |
| Negative-intrusive | 34.5 | 21.77 | 973 | 0.571 | 33.2[a] | 16.54 | 93 | -0.10 | 31.48* | 17.27 |
| 24 months | | | | | | | | | | |
| Sensitivity | 43.44 | 21.04 | 976 | | 36.46[a] | 20.11 | 64 | 0.33 | 44.26 | 24.62 |
| Stimulation | 39.09 | 25.37 | 976 | | 34.11[a] | 22.7 | 64 | 0.25 | 39.75 | 24.73 |
| Positive regard | 47.16 | 24.05 | 976 | | 38.8[a] | 24.32 | 64 | 0.35 | 47.21 | 25.04 |
| Animation | 43.47 | 23.48 | 976 | | 40.63[a] | 19.89 | 64 | 0.28 | 46.31 | 21.91 |
| Detachment | 46.77 | 24.44 | 976 | | 48.96[a] | 21.8 | 64 | -0.39 | 47.95 | 20.24 |
| Sensitive parenting | 45.28 | 19.92 | 976 | 0.891 | 40.21[a] | 17.3 | 64 | 0.39 | 22.34* | 24.1 |
| Negative regard | 44.42 | 20.05 | 976 | | 28.13[a] | 29.08 | 64 | -0.24 | 47 | 24.71 |
| Intrusiveness | 18.55 | 21.23 | 976 | | 55.47[a] | 24.68 | 64 | -0.35 | 34.5 | 21.77 |
| Negative-intrusive | 31.48 | 17.27 | 976 | 0.751 | 41.8[a] | 23.31 | 64 | -0.33 | 44.26* | 24.62 |

**Table B.2 continued:**

| | Included sample | | | | Excluded sample | | | | Imputed sample# | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | n | α | Mean | SD | n | *Hs' G* | Imputed mean | Imputed SD |
| Average of 15 and 24 months | | | | | | | | | | |
| Sensitivity | 43.89 | 20.21 | 1048 | | 36.35[b] | 18.66 | 98 | 0.38 | 43.78 | 20.22 |
| Stimulation | 39.62 | 22.35 | 1048 | | 32.4[b] | 21.23 | 98 | 0.32 | 39.49 | 22.37 |
| Positive regard | 47.14 | 21.59 | 1048 | | 39.16[b] | 21.81 | 98 | 0.37 | 47.02 | 21.59 |
| Animation | 44.44 | 20.83 | 1048 | | 38.78[b] | 21.65 | 98 | 0.27 | 44.26 | 20.84 |
| Detachment | 43.46 | 21.56 | 1048 | | 51.53[b] | 23.42 | 98 | -0.37 | 43.58 | 21.58 |
| Sensitive parenting | 46.33 | 18.36 | 1048 | 0.913 | 39.03[b] | 18.12 | 98 | 0.40 | 46.33* | 18.36 |
| Negative regard | 21.1 | 19.11 | 1048 | | 22.32[b] | 21.63 | 98 | -0.06 | 21.12 | 19.17 |
| Intrusiveness | 45.87 | 19.34 | 1048 | | 50[b] | 18.39 | 98 | -0.21 | 45.94 | 19.36 |
| Negative-intrusive | 33.49 | 16.87 | 1048 | 0.702 | 36.16[b] | 17.03 | 98 | -0.16 | 33.49* | 16.87 |
| Distress to limitations | 40.91 | 13.23 | 1014 | 0.77 | 41.19 | 14.31 | 166 | -0.02 | 40.95 | 13.25 |
| Fear | 30.01 | 16.54 | 977 | 0.888 | 47.33 | 17.51 | 166 | -1.04 | 30.08 | 16.47 |
| Falling reactivity | 70.33 | 15.27 | 1003 | 0.526 | 68.6 | 14.73 | 167 | 0.11 | 70.19 | 15.28 |
| Approach | 71.4 | 14.99 | 1011 | 0.846 | 69.69 | 16.06 | 165 | 0.11 | 33.57 | 11.11 |
| Orienting | 44.43 | 16.05 | 1001 | 0.816 | 44.8 | 17.18 | 163 | -0.02 | 71.25 | 15.07 |
| Negative affect | 33.57 | 11.11 | 1009 | 0.874 | 34.35 | 11.7 | 167 | -0.07 | 44.3** | 16.11 |
| Conduct problems | 3.17 | 2.09 | 1093 | 0.679 | | | 0 | | | |
| Hyperactivity | 4.66 | 2.18 | 1093 | 0.696 | | | 0 | | | |
| Externalizing | 7.82 | 3.73 | 1093 | 0.783 | | | 0 | | | |
| Emotional difficulties | 2.25 | 1.71 | 1093 | 0.503 | | | 0 | | | |
| Peer problems | 1.64 | 1.77 | 1093 | 0.661 | | | 0 | | | |
| Internalizing | 3.89 | 2.94 | 1093 | 0.698 | | | 0 | | | |
| Total externalizing and internalizing | 11.72 | 5.78 | 1093 | 0.821 | | | 0 | | | |
| Pro-social | 6.44 | 1.97 | 1093 | 0.673 | | | 0 | | | |

**Table B.2 continued:**

| | Included sample | | | | Excluded sample | | | | Imputed sample# | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | n | α | Mean | SD | n | *Hs' G* | Imputed mean | Imputed SD |
| Emotional competence | 2.5 | 0.66 | 1092 | 0.864 | 2.58 | 0.54 | 9 | -0.11 | 2.5 | 0.66 |
| Social competence | 2.16 | 0.59 | 1093 | 0.86 | 2.43 | 0.57 | 8 | -0.45 | 2.16 | 0.59 |
| Parental brief symptom inventory | 8.15 | 9.14 | 1067 | 0.859[c]/ 0.886[d] | 7.58[e] | 9.74 | 199 | 0.06 | 8.19 | 9.17 |
| Stressful life events | 3.77 | 4.43 | 1052 | 0.734 | 3.57 | 4.75 | 199 | 0.04 | 3.76 | 4.43 |
| Income-to-needs-ratio | 11.13 | 10.43 | 1064 | | 9.86 | 8.14 | 140 | 0.12 | 11.1 | 10.42 |
| Home learning environment | 77.31 | 21.29 | 1017 | 0.775 | 70.91[f] | 27.11 | 140 | 0.29 | 77 | 21.36 |
| Child age at 6 month sample (months)[g] | 7.72 | 1.48 | 1093 | | 7.86 | 1.56 | 140 | -0.09 | | |
| Proportion male children | 0.496 | | 1093 | | 0.600 | | 140 | -0.104 | | |
| Proportion black children | 0.437 | | 1093 | | 0.387 | | 199 | 0.05 | | |
| Parent age at 2 month sample; years (centred at mean) | 26.37 (0.00) | 6.53 (6.53) | 1079 | | 25.36[e] | 5.63 | 199 | 0.16 | (-0.02) | (6.53) |

Notes. [a] Primary parent at each age. [b] Unclear identity of parent across time period to calculate accurate average. [c] 2 months. [d] 6 months. [e] Only primary parent at 2 month sample if there was no Strengths and Difficulties data for the child at 35 months. [f] One observation available for each child, primary or secondary parent may have been present for missing data only. [g] Or age at 2 month sample plus 4 months.

## Appendix C: Ethics information

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

25 July 2019

Professor Karen Thorpe
Institute for Social Science Research

Human Ethics Research Office
Cumbrae-Stewart Building #72
The University of Queensland
St Lucia, QLD 4072

CRICOS PROVIDER NUMBER 00025B

Dear Prof Thorpe,

2019001822:
Project Title: *When theory needs more evidence: a systematic review and analysis of individual × environment models of child development*

This project has been reviewed by the Office of Research Ethics and is deemed to be exempt from ethics review under the National Statement on Ethical Conduct in Human Research and University of Queensland policy.

1. Research that uses only existing collections of data that contain only non-identifiable data about human beings **AND** is of negligible risk, and is exempt from review (National Statement §5.1.22).

| | |
|---|---|
| Chief Investigator | Professor Karen Thorpe |
| Associate Investigators | Mr Peter Rankin, Professor Mark Western (ISSR); Professor Michele Haynes (Institute for Learning Sciences and Teacher Education, ACU) |
| Data Source | The Family Life Project, Phase I, United States, September 2003-January 2008 (ICPSR 34602) https://www.icpsr.umich.edu/icpsrweb/ <br> The Longitudinal Study of Australian Children (general release 6) https://www.dss.gov.au/national-centre-for-longitudinal-data-ncld/access-to-dss-longitudinal-datasets |
| Completion Date | 01/04/2020 |

Thank you for your advice re previous use of data; UQ HREC is not able to provide retrospective ethical approval. There were varying interpretations on the use of publically available data being outside the scope of the National Statement. Current guidance (3.1.52) *Unless a waiver of the requirement for consent is obtained, any research access to or use of publicly available data or information must be in accordance with the consent obtained from the person to whom the data or information relates.*

Please keep a copy of this document for your records.

Yours sincerely

Chris Rose'Meyer

Chris Rose'Meyer
Governance Officer