The Emergence of Sensitivity to Normative Expectations in British and Ugandan Children

Sophie Marie Marshall

Doctor of Philosophy

University of York

Psychology

April 2025

Abstract

Why do we tip in restaurants, join queues at the back, or lower our voices in libraries? These behaviours reflect our sensitivity to "normative expectations" – our beliefs about what is approved and disapproved of by others (Bicchieri, 2016). Yet, how this sensitivity emerges in childhood, and what drives it, remains unclear. Does normativity unfold similarly across cultures, or in both boys and girls? This thesis investigated the cognitive foundations of children's sensitivity to normative expectations, along with the role of culture and gender in development. Using a novel sharing paradigm, I tested how 4- to 11-year-old children from the UK and Uganda responded to an experimenter's approval of generosity under varying levels of observation. If children were guided by normative expectations, it was expected that they would share most when an experimenter (i) approved of generosity and (ii) was observing their sharing behaviour. From 5 years of age, approval consistently influenced children's sharing, yet only 4-year-old boys specifically showed a pattern consistent with a sensitivity to normative expectations. No cultural differences in children's sensitivity to normative expectations were observed, suggesting that this aspect of normative behaviour may be less influenced by cultural attitudes, and more by developmental factors or gender-related socialisation. To explore these factors, Theory of Mind and executive functioning were assessed, but neither related to 4-year-olds' normative prosociality. Similarly, longitudinal processes which have been theorised to shape children's normativity, such as joint attention and imitation in infancy, did not predict normative sharing. However, general cognitive ability in infancy was associated with later sensitivity to normative expectations, highlighting the importance of domain-general processes. Taken together, these findings challenge existing theoretical models that prioritise specific cognitive mechanisms as foundational for normativity, and highlight instead the role of general cognition and social context in shaping the early emergence of normative behaviour.

List of Contents

Abstract 2

List of Contents 3

List of Tables 6

List of Figures 9

Acknowledgements 11

Authors Declaration 13

Chapter 1 : General Introduction 15

1.1 The Development of Normativity 18

1.2 Reputation and Normative Expectations 19

1.2.1 Reputation Management and Sensitivity to Observation in Adults 20

1.2.2 Reputation Management and Sensitivity to Observation in Adolescence and Late Childhood 21

1.2.3 Reputation Management and Sensitivity to Observation in Early and Middle Childhood 22

1.3 Developmental and Cultural Factors that Influence the Emergence of Normative Expectations 24

1.4 Current Thesis 27

1.4.1 Methodological Approach 28

1.4.2 Outline of Thesis 29

Chapter 2 : Normative Expectations in Early and Middle Childhood: The Influence of Culture and Gender 31

2.1 Abstract 31

2.2 Introduction 32

2.2.1 Factors Influencing the Development of Normative Expectations about Sharing 34

2.2.2 Motivation of the Current Research 38

2.2.3 Study 1 Research Questions 40

2.3 Study 1 42

2.3.1 Methods 42

2.3.2 Procedure and Materials 43

2.3.3 Results 59

2.3.4 Discussion 72

2.4 Study 2 75

2.4.1 Introduction 75

2.4.2 Methods 77

2.4.3 Results 83

2.4.4 Discussion 98

2.5 Chapter Discussion 100

Chapter 3 : Concurrent Predictors of Children's Sensitivity to Normative Expectations 109
3.1 Abstract 109
3.2 Introduction 110
3.2.1 Theory of Mind and the Development of Normative Expectations 110
3.2.2 Executive Functioning and the Development of Normative Expectations 113
3.2.3 The Current Study 116
3.3 Part 1: Research Questions 1 and 2 120
3.3.1 Methods 120
3.3.2 Results 131
3.4 Part 2: Research Questions 3 and 4 136
3.4.1 Methods 136
3.4.2 Results 145
3.5 Discussion 151
Chapter 4 : Longitudinal predictors of normative expectations in British and Ugandan children 159
4.1 Abstract 159
4.2 Introduction 160
4.2.1 Joint Attention and the Development of Normative Expectations 160
4.2.2 Overimitation and the Development of Normative Expectations 164
4.2.3 Domain-General Cognition and the Development of Normative Expectations 168
4.2.4 Motivation for the Current Study 169
4.3 Part 1: Research Questions 1 and 2 174
4.3.1 Methods 174
4.3.2 Results 189
4.4 Part 2: Research Questions 3 and 4 191
4.4.1 Methods 191
4.4.2 Results 197
4.5 Part 3: Research Questions 5 and 6 200
4.5.1 Methods 200
4.5.2 Results 205
4.6 Discussion 208
Chapter 5 : General Discussion 214
5.1 Summary of Findings 214
5.2 Theoretical Implications 215
5.3 Normative Expectations Task 217
5.4 Cultural Variation 218
5.5 Key Considerations for Future Cross-Cultural Research 219

5.6 Key Considerations for Future Social Norms Research 223
5.7 Conclusion 225
Appendix 2 : Supplementary Material for Chapter 2 226
Appendix 3 : Supplementary Materials for Chapter 3 280
Appendix 4 : Supplementary Materials for Chapter 4 320
References 353

List of Tables

Table 2.1 *Video coding criteria for the Normative Expectations task* *52*

Table 2.2 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final models exploring the relationship between baseline sharing (Model 1), and sensitivity to normative expectations controlling for gender (Model 2) and exploring gender (Model 3).* *58*

Table 2.3 *The mean number of tokens given across the Training DG (familiarity phase) and the experimental test trials.* *59*

Table 2.4 *Summary statistics of whether children attended to and comprehended the checkerboard videos shown in the Video Comprehension phase of the Normative Expectation task, broken down by cultural group and gender.* *61*

Table 2.5 *Descriptive statistics of how many 4-year-old children attended to the Generous Prime video prior to each test trial condition in the Normative Expectations task, broken down by cultural group and gender.* *62*

Table 2.6 *Summary statistics of how 4-year-old Ugandan children answered questions about the Generous Prime in the Normative Expectations task.* *62*

Table 2.7 *Results of final Model 1 exploring 4-year-old children's baseline sharing behaviour in the Training DG (estimates and standard errors, together with confidence limits, and significance tests).* *63*

Table 2.8 *Results of final Model 2 exploring 4-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests).* *65*

Table 2.9 *Results of final Model 3 examining 4-year-old children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests).* *67*

Table 2.10 *Phases of the Normative Expectations task with descriptions* *78*

Table 2.11 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final models exploring the relationship between baseline sharing (Model 1), and sensitivity to normative expectations controlling for gender (Model 2) and exploring gender (Model 3) for the 5- to 11-year-old children.* *82*

Table 2.12 *The mean number of tokens given across the DG (familiarity phase) and the experimental test conditions for 5–11-year-old children.* *83*

Table 2.13 *Summary statistics of whether 5- to 11-year-old children attended to and comprehended the checkerboard videos shown in the Video Comprehension phase of the Normative Expectation task, broken down by cultural group and gender.* *85*

Table 2.14 *Descriptive statistics of how many children attended to the Generous Prime video prior to each test trial condition in the Normative Expectations task, broken down by cultural group and gender (5-11-year-olds).* *86*

Table 2.15 *Summary statistics of how 5- to 11-year-old children answered questions about the Generous Prime in the Normative Expectations task.* *86*

Table 2.16 *Results of final Model 4 exploring 5- to 11-year-old children's baseline sharing in the Training DG in the UK and Uganda (estimates and standard errors, together with confidence limits, and significance tests).* *87*

Table 2.17 *Results of final Model 5 exploring 5- to 11-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests are shown).* *90*

Table 2.18 *Results of final Model 6 examining 5- to 11-year-olds children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests).* *95*

Table 3.1 *The N and mean age of the participants with valid data for the Normative Expectations Task and the ToM task in each cultural group.* *121*

Table 3.2 *Description of the 7 Stories Used to Assess Theory of Mind (Wellman & Liu, 2004) and the Materials Used in the UK and Uganda* *124*

Table 3.3 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between performance on the Theory of Mind task and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 and 3).* *130*

Table 3.4 *Number and percentage of children who passed each Theory of Mind item, split by cultural group and gender* *131*

Table 3.5 *Results of final Model 1 examining whether gender or cultural group predicted 4-year-old children's ToM scores (estimates and standard errors, together with confidence limits and significance tests).* *132*

Table 3.6 *Results of final Model 2 examining whether 4-year-old children's ToM scores predict sensitivity to normative expectations in girls and boys in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).* *134*

Table 3.7 *The N and mean age of the participants with valid data for the Normative Expectations Task and the DCCS task in each cultural group.* *137*

Table 3.8 *Exclusion reasons for the participants who were excluded from the final analysis related to the DCCS.* *141*

Table 3.9 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between performance on the DCCS task and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 and 3).* *144*

Table 3.10 *Number and percentage of children who passed the DCCS, split by cultural group and gender.* *145*

Table 3.11 *Results of final Model 4 examining whether gender or cultural group predicted 4-year-old children's DCCS scores (estimates and standard errors, together with confidence limits and significance tests).* *146*

Table 3.12 *Results of final Model 5 examining whether 4-year-old children's DCCS scores predict sensitivity to normative expectations in girls and boys in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).....148*

Table 4.1 *The N and mean age of the participants with valid data for the Auditory Joint Attention task in each cultural group.....175*

Table 4.2 *Joint attention levels coded for the Auditory Joint Attention task.....177*

Table 4.3 *Overview of the Auditory Joint Attention task coding procedure 179*

Table 4.4 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between Joint Attention and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 – 5). 187*

Table 4.5 The N and mean age of the participants with valid data for the Overimitation task in each cultural group.....*192*

Table 4.6 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between overimitation and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 and 3). 196*

Table 4.7 *Results of final Model 6 examining whether overimitation is predicted by gender or cultural group (estimates and standard errors, together with confidence limits and significance tests).....197*

Table 4.8 *The N and mean age of the participants with valid data for the Bayley Cognitive Scale in each cultural group.....201*

Table 4.9 *Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between cognition and (i) gender and cultural group (Model 9) and (ii) sensitivity to normative expectations (Model 10). 204*

Table 4.10 *Results of final Model 10 examining whether sensitivity to normative expectations at 4 years is predicted by cognition at 11 months in girls and boys, controlling for cultural group (estimates and standard errors, together with confidence limits and significance tests). 206*

List of Figures

Figure 2.1 *Example of each Experimental Phase of the Normative Expectations Task: A) Pre-Test Phase, B) First Test Phase, C) Second Test Phase. The top row shows a UK example, whilst the bottom row shows a Ugandan example.* *45*

Figure 2.2 *Probability that a token is given in the Training DG for per level of Cultural Group (UK/Uganda). Bar plots show the raw data of the mean proportion of tokens given. Horizontal lines with error bars depict fitted model and its 95% confidence interval.*..... *64*

Figure 2.3 *Probability that a token is given for per level of Observation (Observed/Unobserved) and Approval (Approval/Indifference) with gender centred. Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 1 to 28). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits*..... *66*

Figure 2.4 *Probability that a token is given for per level of Observation (Observed/Unobserved) and Approval (Approval/Indifference) for girls (orange) and boys (blue). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 1 to 18). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.*..... *68*

Figure 2.5 *Probability that a token is given for per level of Approval (Approval/Indifference) and Cultural Group (UK/Uganda) for girls (orange) and boys (blue). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 1 to 25). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.*..... *69*

Figure 2.6 *Probability that a token is given for per level of Observation (Observed/Unobserved) and Cultural Group (UK/Uganda) for girls (orange) and boys (blue). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range:1 to 24). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.*..... *70*

Figure 2.7 *Probability that a token is given for per level of Cultural Group (UK/Uganda) across 5-11 years in the Training DG. Points depict the raw data (proportion of tokens given), whereby point area is proportionate to the number of children with the same probability a token is given (range:1 to 9). Lines with coloured areas depict the fitted model and its bootstrapped 95% confidence limits.*..... *88*

Figure 2.8 *Probability that a token is given per level of Observation (Observed/Unobserved) across ages. Points depict the raw data (proportion of tokens given), whereby point area is proportionate to the number of children with the same probability a token is given (range: 1 to 9). Lines with greyed-out areas depict the fitted model and its bootstrapped 95% confidence limits.*..... *91*

Figure 2.9 *Probability that a token is given per level of Approval (Approval/Indifference). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children who gave the same proportion of tokens (range: 5 to 130). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.*..... *92*

Figure 2.10 *Probability that a token is given per level of Cultural Group (Uganda/UK). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 4 to 220). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.*..... *93*

Figure 2.11 *Probability that a token is given for levels of Observation (Observed/Unobserved) across the ages of 5- to 11-years, separately for A) UK boys B) UK girls, C) Ugandan boys and D) Ugandan girls. Points depict the raw data (proportion of tokens given), whereby point area is proportionate to the number of children with the same probability a token is given (range: 1 to 9). Lines with greyed-out areas depict the fitted model and its bootstrapped 95% confidence limits* *97*

Figure 3.1 *Probability that a ToM item is passed for girls and boys in Uganda and the UK. Bubbles depict the raw data (proportion of tokens given) whereby bubble area is proportionate to the number of children with the same proportion of ToM items correct (range: 1 to 22). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.* *133*

Figure 3.2 *The pictures of the red flower and blue car on the target cards that were affixed to the sorting trays during the DCCS task.* *138*

Figure 3.3 *The pictures of the red car and blue flower on the test cards that were presented to the children during the DCCS task.* *139*

Figure 3.4 *Probability that the DCCS is passed in Uganda and the UK. Bars depict the raw data (proportion of DCCS passes). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.* *147*

Figure 3.5 *Probability that a token is given for children in Uganda and the UK. Bubbles depict the raw data (proportion of tokens given) whereby bubble area is proportionate to the number of children with the same proportion of tokens given (range: 1 to 49). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.* *149*

Figure 4.1 *Three example timelines of Auditory Joint Attention event coding* *183*

Figure 4.2 *The set-up of the Overimitation task in the UK (left) and Uganda (right).* *194*

Figure 4.3 *Probability that overimitation was present in Uganda and the UK. Bars depict the raw data (proportion of participants that overimitated). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.* *198*

Figure 4.4 *Administration set-up of the Bayley Cognitive Scale in the UK (left) and Uganda (right).* *202*

Figure 4.5 *Probability that a token is given in the Approval Observed condition and other conditions. Bubbles depict the raw data (proportion of tokens given) whereby bubble area is proportionate to the number of observations (range: 1 to 17). Lines depict the fitted model, and grey areas depict the bootstrapped confidence limits.* *207*

Acknowledgements

This thesis would not have been possible without the support of many people, for whom I am deeply grateful. Firstly, I would like to extend a special thanks to my supervisors, Dr Bailey House and Professor Katie Slocombe. Your guidance, patience and kindness throughout this whole process have been second to none. While it is bittersweet that many years of supervision have come to an end (10 years for you, Katie!), I take comfort in knowing I will carry with me all the lessons, knowledge and encouraging words you've shared. The unwavering belief you have had in me every step of the way is something I will be forever grateful for.

I would also like to thank the members of my Thesis Advisory Panel, Professor Harriet Over and Dr Tom Hartley, and my Progression Panel chair, Professor Beth Jefferies, for their helpful insights, advice and support. Thank you also to Roger Mundry for all his statistical wisdom, and for passing on his infectious excitement for statistical models.

Thank you to the Economic Social Research Council for funding my PhD, and the Leverhulme Trust for funding this project. A special thanks to all the families from both the UK and Uganda who participated and dedicated their time to this project – none of this would have been possible without you.

I would also like to express my gratitude to my lab mates, Dr Joanna Buryn-Weitzel and Dr Charlotte Knapper. Joanna, thank you for always listening, understanding, and teaching me how to set healthy boundaries. Charlotte, we have shared so much laughter (and a few tears), and I couldn't have wished for a better person to be on this journey with. Thank you for your ongoing kindness, support and friendship.

I would also like to extend my thanks to those people who made this research possible in the UK and Uganda, including Agnes Ayikoru, Beatrice Peringa, Caitlin Woods, Charlie Ives, Charlotte Knapper, Daisy Whitwood, Eleanor Upham, Ellen Lavender, Eve Holden, Florence Tusiime, Georgia Birchenough, Gideon Salter, Hellen Biroch, Joan Isabella, Joanna Buryn-Weitzel, Josephine Paricia, Kirsty Furnell, Lauren Stanton, Lucy Whiting, Michael Jurua, Olivia Lake, Prezzy Eneku, Rachel Bordom, Rebecca Willis, Rhiannon Pearce, Santa Atim and Simon Mua. This project would not have been possible without all the hours collecting, translating and coding data, so thank you all for your help along the way. Also, thank you to the whole Ugandan team and everyone at Nyabyeya Forestry College for providing a home away from home in Uganda – I will cherish those memories forever.

Thank you to my Psychology friends, Anna Crossland, Hannah Brown and Dr Will Strawson. Anna, thank you for being such a reliable, caring and loving friend. I'm so happy the Masters brought us

together, and I cherish our long voice note catchups more than I can say. Hannah, I knew from the moment we met that I'd made a friend for life. Thank you for making every day in the office (and outside of the office) a joy with your positivity, jokes and genuine interest. Will, when I think back to those early revision days, it truly fills me with joy. Thank you for all fun gaming times, the dinners full of nutritional yeast, the crammed revisions sessions, and most importantly, for being an amazing friend.

Thank you to my best friend, Dr Brontë Mckeown. I never thought I'd be so grateful for a mini-group project, but looking back, I count my lucky stars that it brought us together. You have been by my side every step of the way – always ready to talk through the biggest of ideas and the smallest of worries. Thank you for cheering me on during my brightest moments and helping me see the light at the end of the tunnel in my darkest moments. I couldn't wish for a better friend.

I would also like to thank my beautiful family, whose faith in who I am, and who I can become, means more than words can say. To my mum, Mandy, thank you for always being there to help me figure out how to let go of the things I can't control and for reminding me that fun isn't something extra, but essential. To my dad, Martin, thank you for teaching me the quiet strength of kindness and for never doubting my capabilities. To my sister, Jessie, thank you for always leading with your heart and being in my corner – your support makes every situation lighter.

Finally, thank you to my wonderful partner, Ross, for being the first to lift me up and the last to let me fall. Through all the practice talks and many vents about statistical models and writing, your interest has never wavered. Thank you for being the voice reminding me I can do it, and for all the care, patience and love along the way. I am truly blessed to have you by my side.

Authors Declaration

I declare that I am the sole author of this thesis and, unless stated otherwise, all work presented in this thesis was completed by myself. This work has not previously been presented for a degree or other qualification at the University of York or elsewhere. All sources are acknowledged as references.

Most of the data presented in this thesis was collected as part of Dr. Bailey House and Prof. Katie Slocombe's Leverhulme grant exploring 'The Development of Normative Prosociality across Diverse Societies'. Data presented with participants younger than 2 years of age was collected as part of Prof. Katie Slocombe's European Research Council grant exploring 'The Evolutionary and Developmental Origins of Joint Attention: A Longitudinal Cross-Species and Cross-cultural Comparison'.

In the UK, Bailey House and Katie Slocombe contributed to designing experimental protocols, outlining research questions, and reviewing the present thesis. Charlotte Knapper and Joanna Buryn-Weitzel contributed to the design of experimental protocols and data collection. Ellen Lavender, Rebecca Willis, Georgia Birchenough, Olivia Lake, Lauren Stanton, Kirsty Furnell, Rachel Bordom, Eve Holden, Rhiannon Peace, Daisy Whitwood, Charlie Ives and Caitlin Woods all contributed to data collection.

In Uganda, Florence Tusiime, Joseph Rusenjule, Charlotte Knapper, Claudia Wilke, Katie Slocombe and Bailey House contributed to the management of the Ugandan team of research assistants and translators. Hellen Biroch, Florence Tusiime, Agnes Ayikoru, Beatrice Peringa, Santa Atim, Josephine Paricia and Michael Jurua contributed to designing experimental protocols, data collection and participant recruitment. Joan Isabella, Nesta Gune, Lucy Stafford and Georgia Birchenough contributed to data collection. Florence Tusiime, Agnes Ayikoru, Beatrice Peringa, Joseph Rusenjule, Joan Isabella, Simon Mua, Michael Jurua and Prezzy Eneku contributed to the translation of experimental protocols.

Katie Slocombe and Bailey House contributed to the design of the coding schemes presented in Chapters 2 – 4. Gideon Salter contributed to the design of the Auditory Joint Attention task, presented in Chapter 4. Olivia Jones and Bailey House designed the coding scheme for the Overimitation task at 18 months presented in Chapter 4. Eleanor Upham contributed to the coding of the Normative Expectations task presented in Chapter 2, and the coding of the Theory of Mind task presented in Chapter 3. Kirsty Furnell and Rachel Bordom contributed to the coding of the Dimensional Change Card Sort task presented in Chapter 3. Gideon Salter, Kirsty Furnell, Emily Mason and Lucy Whiting contributed to the coding of the Auditory Joint Attention task presented in Chapter 4. Olivia Jones and Ellen Lavender coded the video data for the Overimitation task presented in Chapter 4.

Roger Mundry advised with data analysis.

Preliminary results from Chapter 2 were presented during a poster session at the Emory-St. Andrews 'Mechanisms of Learning' Conference in June 2023, a poster session at the Budapest CEU Conference on Cognitive Development in January 2024, a talk at the Internation Association for Cross-Cultural Psychology Conference in August 2024 and during the University of York Psychology Departmental Seminar series in May 2024.

Chapter 1: General Introduction

Humans are inherently social, and cooperation with our social partners lies at the heart of our species' success (Henrich, 2016). Though we are not the only cooperative species, we cooperate in large groups of unrelated partners to achieve a wide range of different goals, leading to a diversity and scale of cooperative behaviours that exceeds those of any other species (Boyd & Richerson, 2009). This cooperative behaviour often includes acts of prosociality (i.e., acting in a way that benefits others) towards non-relatives, such as when caring for the vulnerable, participating in division of labour and voluntarily sharing food (e.g., Gurven, 2004; Jaeggi, Burkart & Van Schaik, 2010; Marlowe, 2007). These forms of prosociality – particularly sharing – frequently involve incurring a personal cost to benefit others, sometimes even strangers. Why would someone do such a thing? One mechanism that is believed to be part of the answer is humans' tendency to conform to social norms, as these rules can motivate people to do many things that they might not otherwise be motivated to do. However, though this conformity might be crucial for our understanding of cooperation, there is much that we do not know about the developmental foundations of our psychology for following of social norms for prosociality. Better understanding how our psychology for conforming to prosocial social norms emerges between infancy and adulthood will help us to explain how uniquely human forms of cooperation come to emerge during childhood.

Social norms are rules or principles that are shared – and sometimes enforced – by members of a group (Bicchieri, 2006). Enforcement of *prosocial* social norms can motivate people to share or help others even at a personal cost, potentially helping to explain the puzzle of prosociality in humans (Chudek & Henrich, 2011; Richerson et al. 2016). For example, across very different societies, adults tend to share the amount deemed appropriate by other members of their social group, regardless of their personal preferences about how much they would like to give (Henrich et al., 2006; House et al., 2020a). However, it's important to note that norms aren't only important for sharing behaviour, they have also been reported to affect adult behaviour and promote prosociality across a variety of real-life contexts, such as: charitable donations (Martin & Randal, 2008), preventing littering (Cialdini et al., 1990; Kallgren, Reno & Cialdini, 2000), reducing water usage and energy consumption (Jaeger & Schultz, 2017, Jachimowicz et al., 2018; Nolan et al., 2008) and a number of other pro-environmental behaviours (see Farrow, Grolleau & Ibanez, 2017 for a review). Throughout development, children must learn to navigate different social contexts, which will at least sometimes require an awareness of the behaviour that is approved of within their specific social group, which is at least partly rooted in their culture's norms. Through this process of perceiving and adapting to normative information, children's behaviour will presumably align more closely with that of adults as they develop. However,

little empirical work has explored the very early development of children's sensitivity to normative information about others' approval, and which social or cognitive mechanisms might support it.

Bicchieri (2006; 2016) argues that an individual's motivation to conform to a social norm is related to their beliefs about other people. Specifically, their beliefs about (i) whether other people follow a norm and (ii) whether those people *approve* of the norm. These beliefs about other people are referred to as 'social expectations' and are crucial for understanding some of the most important aspects of human normative behaviour. Bicchieri's (2016) model proposes that there are two types of social expectations: empirical expectations and normative expectations. Empirical expectations are beliefs about what other members in the group are doing, and are theorised to be salient drivers in situations where it isn't otherwise clear how one should act. For instance, when deciding which side of the road to drive on in a different country, we attend to how the collective group is behaving, and then drive on the same side (Bicchieri, 2016). It is through observation or indirect knowledge in which we ground our empirical expectations about the appropriate behaviour in certain situations. However, in circumstances where an individual might prefer to *not* do what others are doing (such as when self-interest might lead them to act selfishly by sharing less even if others are being prosocial), empirical expectations alone may not be sufficient for norm compliance. In these situations, you need a different kind of expectation to motivate behaviour: normative expectations.

Normative expectations are beliefs about what other members approve of in a given situation. In the examples of social norm following highlighted above, such as charitable donations and pro-environmental behaviours, prosociality often conflicts with self-interested motives, but individuals nonetheless conform to the norm and act prosocially regardless of the personal cost. They likely do this at least partly because their belief about what others approve of in these situations (i.e., normative expectations) can promote behaviour aligned with the group's social norms, and so avoid sanctions (i.e. punishment) that other group members might impose on them for non-conformist behaviour (Bicchieri & Muldoon, 2018; Kallgren, Reno & Cialdini, 2000). Therefore, it has been argued that social norms, and particularly normative expectations, may be deeply linked to prosociality and cooperation, as they can motivate people to cooperate regardless of whether they would want to do so in the short term (Chudek & Henrich, 2011; Richerson et al., 2016).

While the theoretical framework put forward by Bicchieri (2006; 2016) forms the basis of this thesis, there are also other theoretical lenses through which norms have been conceptualised. One prominent example is Social Domain Theory (Turiel, 1983), which differentiates between moral norms and conventional norms. Within this framework, moral norms are based on relatively universal concerns about welfare and fairness, whereas conventional norms are based upon perspectives of 'right' and

'wrong' which are derived from socially agreed rules. When considering the underlying motivations for prosociality, such as generosity, this distinction may be useful, as it is possible that prosociality may reflect both a moral concern for others' welfare, and a sensitivity to social approval. For the purpose of this thesis, while we acknowledge that prosocial behaviour can have moral motivations, we are particularly interested in how prosociality can be driven by social approval – specifically, normative expectations – which can be viewed as conventional norms within the Social Domain Theory framework, and as such, align more closely with Bicchieri (2006; 2016).

Evidence for this relationship between normative expectations and prosociality comes from lab-based economic games like the Dictator Game, in which participants can choose how much of an endowment they would like to donate to another participant. In these games, empirical expectations can be presented by describing the amount most others donate (e.g., "Most players give £0.50"), and normative expectations can be presented by describing the suggested amount an individual should donate (e.g., "It is suggested that players give £0.50"). Previous studies including these kinds of prompts about empirical and normative expectations have reported that adults often attend to the information about what should be donated in comparison to baseline conditions, or information presented about what others actually do, indicating that they are acting in accordance with their beliefs about what other people think the appropriate amount to donate is – their normative expectations (Raihani & McAuliffe, 2014; Smith & Louis, 2008). Cross-cultural research using economic games like the Dictator Game suggests that prosocial behaviour and the willingness to punish selfish behaviour varies between cultural groups (Fehr & Gächter, 2002; Henrich et al., 2006; Hermann, Thoni & Gächter, 2008) and may be linked to cultural differences in normative beliefs (Henrich et al., 2006; House et al., 2020b), again indicating that a sensitivity to what others in our social group approve of may be related to our prosocial behaviour. But, how does sensitivity to normative expectations develop? To answer this question, it is important to consider both (i) when during development children begin to show sensitivity to social norms, and (ii) when their prosocial behaviour becomes more similar to adult behaviour within their society. We might assume that a developing sensitivity to norms would lead to more adult-like behaviour, but is this true?

This chapter presents an overview of how the human understanding of social norms develops, using empirical research with both children and adults. The aim of this will be to demonstrate that although current research tells us something about the ages at which children begin to show an awareness of social norms, or a tendency to copy others, there is limited research directly on the question of how children's behaviour relates to their beliefs about how others evaluate their behaviour (i.e. normative expectations). However, we can make crucial inferences about the development of this by drawing from prior literature on reputation management, and individuals' sensitivity to being observed. Both

play a pivotal role in our decisions about whether or not to conform to others' expectations. Finally, the factors that may influence variation in children's sensitivity to normative expectations will be detailed, before the current thesis is outlined.

1.1 The Development of Normativity

Developmental studies have found that some foundations of children's understanding of norms emerge in early in life. Recent research suggests that from as young as 11 months, infants show some implicit understanding of the context surrounding social approval, as measured by pupillary surprise (Köster & Hepach, 2024). Infants were presented with animated characters that acted in a way that was either similar or different to the actions of others and were responded to with either approval or disapproval from other characters. Pupil diameter increased (signalling surprise) when conformity was met with disapproval (but not approval), and non-conformity was met with approval (but not disapproval). This suggests that some understanding of social evaluation may develop early in development.

The bulk of developmental studies, however, have focussed on more explicit behavioural markers of normative awareness, which seem to emerge in the third year of life. For instance, from the age of 3 young children begin acting in accordance with a norm and enforcing others to do the same (i.e. norm conformity); they show explicit understanding of conventional norms (Kanngiesser et al., 2022; Rakoczy & Schmidt, 2013), conform to conventional statements (e.g. rules about how "we" behave; Li, Britvan & Tomasello, 2021) and protest non-normative behaviour such as rule-breaking during conventional games (Rakoczy, Warneken & Tomasello, 2008). Furthermore, children as young as age 3 enforce fairness norms of sharing, by indicating that they should share equally and anticipating others will share equally with them, but it may not be until later (between age 7 and 8) that children actually begin to share equally (Smith, Blake & Harris, 2013). Around age 4 to 5 years is also the time at which important socio-cognitive skills which could theoretically support normativity show important developmental changes, such as Theory of Mind (House, 2018) and executive functioning (House & Tomasello, 2018).

Consistent with this, developmental research has also found that at older ages there are further changes in children's understanding of and sensitivity to social norms. For example, studies have found that across a range of societies, prosociality becomes increasingly influenced by social normative information in middle childhood (House et al 2020a; House & Tomasello, 2018), at around the same age that (i) children begin to punish others for being selfish (House et al 2020b), and (ii) societal variation in prosociality and fairness emerges (Amir et al., 2024; Blake et al., 2015; Cowell et al., 2017; House et al., 2013; House, 2017; House, 2018; House et al., 2020a; Huppert et al., 2019;

Rochat et al., 2009). Though societies may vary in the overall frequency with which children punish selfish behaviour, the developmental trajectory of third-party punishment appears to develop similarly across societies – increasing substantially in early to middle childhood (House et al., 2020b). The developmental co-emergence of (i) children's sensitivity to social norms, (ii) children's understanding of punishing others for selfishness, and (iii) culturally variable prosocial behaviour strongly points to the emergence of children's understanding of normative expectations. These findings suggest that at some point during ontogeny, likely between early and middle childhood, humans (i) become increasingly sensitive to normative expectations and act in accordance with their beliefs about what is deemed to be collectively appropriate, which leads to (ii) the emergence of cultural variation in prosociality across societies with different social norms. Though there is very limited direct evidence for the development of children's understanding of normative expectations, there is substantial indirect evidence from research on the development of children's understanding of reputation.

1.2 Reputation and Normative Expectations

'Reputation' is a widely used concept in psychological literature, yet explicit definitions of it vary. For instance, Emler (1990) describes reputation as a social construct, based on our beliefs about how we are seen by others, while Leimgruber, Shaw, Santos and Olson (2012) state that reputation is information about another's character that can be used by others to anticipate that person's future actions. For the purpose of this thesis, 'reputation' will refer to our beliefs about how we are viewed in the eyes of others (aligning with Emler, 1990). The practice of tracking others' reputations, and maintaining one's own reputation, can play an important role in encouraging cooperation in social groups (Miliniski, Semmann & Krambeck, 2002; Nowak & Sigmund, 1998).

Humans, like other species, favour partners for joint or cooperative tasks who are good cooperators (Hermann et al., 2013; Melis, Hare & Tomasello, 2006; Subiaul et al., 2008; Wedekind & Milinski, 2000). Humans also understand they can be judged as a social partner, and thus act in ways which promote the positive judgements of others who are watching their behaviour (Engelmann, Herrmann & Tomasello, 2012). When individuals are aiming to positively impact how they are seen by others, it is of obvious importance to attend to whether or not others are watching or observing their behaviour, but it is also crucial to attend to what behaviour others approve of in that situation (i.e. normative expectations). The absence of information about one, or both, of these would limit an individual's ability to act strategically to positively influence how they are viewed in the eyes of others (i.e. reputation), either because (i) they may not act in accordance with what is socially approved of in that situation, or (ii) they don't know if engaging in socially undesirable behaviour could be negatively attributed to their identity. As such, the above model of normative expectations implies that adherence to norms is magnified in situations where reputational concerns are important and people could be

punished for their behaviour (Fehr & Fishbacher, 2004), such as when they are being observed by others.

1.2.1 Reputation Management and Sensitivity to Observation in Adults

Research suggests that (among adults) prosociality is motivated, to some extent, by an implicit desire to promote one's reputation by gaining social approval (Cialdini & Goldstein, 2004; Fehr & Fischbacher, 2004). It has been widely documented that adults often maximise their prosocial acts when in the presence of an audience who can witness their actions. In a study by Rege and Telle (2004), participants were asked to decide how much to either contribute to a public good or keep for themselves. In different manipulations, the authors attempted to explore the relative influence of social approval, specifically, when the participants had to show their decision to the rest of the group, and framing (i.e. using normative terms associated with cooperation), on the contributions to the public good. They found that participants donated the least when there was no chance for social judgement and when no framing was used. The inclusion of normative language alone increased donations slightly, but participants donated a much higher amount when social approval was a possibility, and this was further increased if the chance for social approval was alongside normative language. This suggests the chance of social approval (or disapproval) has a powerful influence on giving behaviour. Additionally, in a similar public goods game conducted by Andreoni and Petrie (2004), participants donated significantly more to a public good when specifics about their identity and their donation were both available to other participants. Importantly, neither identification or information about their donation when presented alone had the same effect as presenting them both together, suggesting that for an increase in generosity, others must be able to firstly, know who you are and secondly, know the amount you have donated to the public good. It is only through both of these that others can evaluate your behaviour and attach this evaluation to your identity, effectively influencing your reputation. This pattern has also been reported across several lab-based economic games (Bohnet & Fret, 1999; Christens et al., 2019; Reinstein & Reiner, 2012; Samek & Sheremeta, 2014). Whilst laboratory experiments have received criticism from some authors, as the perception of an experimental situation alone may increase self-awareness with regards to participant's self-image (List & Levitt, 2005), it can be argued that these are still evidencing the effects of observation on prosocial behaviour - even if that observation is perceived as stronger by participants of a lab-based study due to the nature of these studies. Taken together, these studies provide evidence that observation increases adult prosocial and cooperative behaviour within a lab-setting, and also indicates that reputational concerns may moderate this.

Furthermore, these lab-based effects of observation on prosociality have been complimented by a number of naturalistic field studies. Alpizar, Carlsson and Johansson-Stenman (2008) demonstrated

that, when asked for a monetary contribution, tourists in a Costa Rican National Park donated more money when observed by a solicitor than when donations were completely anonymous. In a similar study by Alpizar and Martinsson (2013), it was found that donations to a Costa Rican National Park were incrementally and positively correlated with visitors traveling in a group (i.e. with fellow companions), and with being exposed to a third party (i.e. other travellers). This tendency for observation to increase prosocial behaviour has been reported in several other contexts, such as other kinds of monetary donations (Izuma, Saito & Sadato, 2010; Soetevent, 2005), blood donations (Lacetara & Macis, 2008) and charitable volunteering opportunities (Bereczkei, Birkas, & Kerekes, 2007). Additionally, contributions have been reported to increase further when participants receive an object which they can use to publicly display their prosocial behaviour. In a recent study by Dannenberg, Johansson-Stenman and Wetzel (2022), when participants were given the option to donate money towards a climate change cause, they donated the most when they received a gift detailing their donation (e.g. a mug or bag which expressed they were supporters of the climate campaign). This gift was believed to act as a status signal to other members of the participant's social environment, indicating that they have acted prosocially. This suggests that the more publicly salient the effects of observation, the bigger the effect on prosocial behaviour. Overall, these studies highlight that adults often act more prosocially during observation, in an attempt to gain social approval and increase their reputation. Conversely, when not observed, individuals may be more inclined to act with self-interest. Although indirect, these results suggest that individuals are assessing what others believe to be the socially desirable behaviour within a given situation (i.e., the normative expectations), and even if it presents a cost to the self, are acting accordingly.

Together, these studies with adults highlight the importance of studying how observation by other group members fosters a situation where reputational concern is high, such as where you believe that others want you to act in a particular way (i.e. you hold normative expectations that others approve of acting in that way). These will be the most essential factors to consider when seeking to understand how social norms begin to impact children's behaviour during development.

1.2.2 Reputation Management and Sensitivity to Observation in Adolescence and Late Childhood

Whilst direct research on sensitivity to normative expectations in young children is extremely limited, there is evidence which suggests adolescents use normative expectations to guide their behaviour. Van Hoorn et al., (2016) examined the influence of peer feedback on giving behaviour in a sample of 12- to 16-year-old adolescents. In an initial baseline condition, participants were asked to decide how much to keep and how much to contribute to the group. In another set of conditions, while they made their decisions, participants were observed online by ten peers (confederate stooges) who provided

feedback on their allocations (which was controlled by experimenters). Across different conditions, peers either signalled a thumbs up for prosocial decisions, a thumbs up for antisocial decisions or gave no feedback at all. Participants' behaviour significantly aligned with what choices their peers approved of, regardless of whether this was antisocial or prosocial. In a later unobserved condition, participant's behaviour returned respectively closer to baseline (i.e. adolescents in the prosocial condition gave less, and adolescents in the antisocial condition gave more). A control condition, in which no feedback was given, did not significantly influence the participants' decisions, even though they were observed for the same number of rounds. This suggests that adolescents are significantly susceptible to what others in their group approve of, and this guides their behaviour, especially when being observed by those group members. These results highlight the important intersection between normative expectations and observation: you should act in line with what you believe is expected of you to gain social approval, especially when others are around to witness compliance or violations.

This kind of direct empirical investigation on sensitivity to normative expectations has not yet been extended to studies with young children. There remains to be a clear picture of how children's behaviour may be related to how sensitive they are to others' evaluations. Though existing research suggests that normative expectations are becoming more influential during the period between early and middle childhood, they do not explicitly detail how the influence of normative expectations interacts with observation, which is essential to revealing the development of a more adult-like understanding of norms. Revealing this interaction would require experiments which manipulate both what others approve of, and whether others observe the child's own behaviour. However, to date no studies have explored both of these variables simultaneously in young children between early and middle childhood. However, there have been some studies with children of this early age which have explored either reputation management or sensitivity to observation.

1.2.3 Reputation Management and Sensitivity to Observation in Early and Middle Childhood

Prior studies have reported that children's earliest prosociality is largely unaffected by observation or reputational concerns. For example, children under age three have been reported to act similarly in helping paradigms when they are unobserved, compared to whether they were observed by a parent (Warneken & Tomasello, 2012) or by another adult (Hepach et al., 2017). Although prosociality between the ages of 1 to 3 years has been widely documented (e.g. Brownwell, 2012), the motivations underlying these early behaviours appear to differ from those observed in older children. During infancy and toddlerhood, prosociality is thought to be driven by preferences for interacting with others, or completing others' goals (Dahl & Paulus, 2019). It is not until later in childhood, around the preschool years, that prosocial behaviours are thought to be driven by an empathic motivation to aid

others' distress, or an ability to normatively evaluate how – or if – one should respond prosocially in a given situation (Dahl & Paulus, 2019; Paulus, 2014). This suggests that prosocial motivations shift throughout development from being intrinsic or based on personal preferences in infancy, toward normatively and reputationally based in childhood. For instance, it is not until the age of 5 years that children seem to start to become aware of observation or reputational concerns. At the age of 5 years, children show some understanding of how reputations are obtained (Hill & Pillow, 2006) and the factors that can influence them (Banerjee, 2002; Gee & Heyman, 2007), but an adult-like understanding of reputation is still developing between age 5-8 years (Aloise-Young, 1993; Bennett & Yeeles, 1990; Engelmann and Rapp, 2018; Hill & Pillow, 2006; Heyman, Fu & Lee 2007, Heyman et al., 2014). One useful distinction, which aims to explain the differences in reputational concerns between early and middle childhood, is between implicit and explicit reputation management (Engelmann & Rapp, 2018). Implicit reputation management, emerging around the age of 5, refers to instances in which children's concern for reputation becomes integrated with whether they are being watched and which behaviour is more likely to create a positive impression. In contrast, explicit reputation management, which emerges around the age of 8, involves more flexible and deliberate attempts to influence others' impressions, in order to strategically manage reputation (Engelmann & Rapp, 2018). This suggests that the onset of a more advanced understanding of reputation is occurring during middle childhood, the time that children are also showing increasing sensitivity to social norms and punishment. Fu et al., (2015), presented 3-, 4- and 5-year-old children with an experimental task where they could cheat (i.e. peek at hidden toys) in a guessing game to improve their chances of receiving a reward. Before playing the game, some children were primed to think about their reputation (by being told other children in the class had reported they were a good child; the 'reputation' condition), while other children were not primed to think about reputation (the control condition). The 5-year-old children cheated less when primed about their reputation, but the 3- and 4-year-olds did not show this pattern. This suggests that to at least some extent, around the age of 5 years, children are becoming increasingly likely to avoid actions which might contradict what others think about them. However, this does not mean that children are changing their actions simply to avoid others' disapproval, evidence for this would come from experimental results showing that the reputation prime only influenced children's behaviour when *others would know that they cheated*. This can be controlled by manipulating whether or not children are being observed.

A number of studies suggest that it is also around 5 years of age that children's behaviour begins to be influenced by the presence of other people, suggesting they may be starting to understand the significance of other people observing their actions (i.e. that this is when their behaviour can be judged negatively by others). Engelmann, Hermann and Tomasello (2012) presented 5-year-old children with a sticker game in which the aim was to place different shaped stickers in their correct

locations on a sheet. Some children were given an extra sticker and told they could donate it to an anonymous peer recipient (the helping condition), while other children weren't given enough stickers – meaning that to complete the task they would have to steal a sticker from the peer recipient (the stealing condition). In both the helping and stealing conditions the study manipulated whether or not children were observed, and as young as age 5 children tended to help more and steal less when being observed, suggesting that they are modifying their behaviour based on whether or not they were being observed. Subsequent studies found that this effect of observation was particularly pronounced when the observer was an ingroup member (Engelmann, Over, Herrmann and Tomasello, 2013). In related studies, 5-year-old children have also been found to be most generous when their identity and actions were revealed to the recipient (Leimgruber et al., 2012), and when the recipient could subsequently reciprocate (Engelmann, Over, Herrman and Tomasello, 2013).

These results suggest that children's behaviour, particularly their prosocial behaviour, is sensitive to observation by relevant third parties from the age of 5. One possibility, which has been implicitly assumed by research thus far, is that these results show that children are beginning to track what others approve of or want them to do (i.e. normative expectations), and that they are responding to observation by acting in accordance with their normative expectations to gain social approval. Yet, an alternative explanation could be that children are simply learning through experience and reinforcement how best to behave when others are around and watching, but they may not understand why others reinforce their behaviours in certain ways. In other words, they may not be carefully tracking the beliefs and approval of other individuals nor strategically adjusting their choices to avoid the disapproval of those specific individuals. This is a subtle distinction which cannot be made using data from existing studies, but it is crucial to distinguish these possibilities if we are to identify the emergence of a truly adult-like psychology of social norms. Although these different accounts would likely manifest behaviourally in the same way in many experimental situations, one way to test whether children are perceiving and adapting to normative expectations in an adult-like fashion would be to explicitly manipulate both their expectations (i.e. their beliefs about others' approval) while also manipulating the presence of observation. This would allow us to determine how children's early prosociality is shaped by their sensitivity to social norms and expectations, and pave the way to understand which psychological mechanisms are supporting children's understanding of social norms early in development.

1.3 Developmental and Cultural Factors that Influence the Emergence of Normative Expectations

To summarise, previous research suggests that (i) children may be sensitive to social evaluation from infancy, (ii) social norms begin to influence children's behaviour from as young as 3 years of age, and

(iii) children display a sensitivity to observation (and a desire to manage their reputations) from about 5 years of age. One possibility is that from about age 5 children are becoming sensitive to what others approve of or want them to do (e.g., normative expectations), but existing experimental studies do not allow us to evaluate this possibility because these study designs have not varied both normative expectations (i.e., what others approve of) and whether or not children are being observed. Future research is therefore needed if we are to properly describe the link between normative expectations and observation in children, and how this begins to shape children's behaviour (in particular their prosocial behaviour) during early and middle childhood. Furthermore, future research should go beyond simply describing how children's sensitivity to normative expectations changes with age, and begin to explore how the psychology for social norms is built upon earlier cognitive precursors and sociocultural factors. By doing this, it will be possible to describe the development of children's sensitivity to normative expectations more clearly, and to test recent theoretical models of both the nature and origins of the human psychology for social norms, and the ontogenetic roots of human prosociality (Hepach & Warneken, 2018).

Recent theoretical work proposes that our psychology for social norms is built upon the development of other socio-cognitive abilities, such as joint attention (Tomasello, 2019). This idea is supported by empirical findings that joint attention is related to the emergence of mental-state reasoning abilities in childhood (Charman et al., 2000; Kristen et al., 2011; Nelson, Adamson & Bakeman, 2008). Normative expectations necessarily require an individual to consider what others want us to do, which is likely related to our ability to track others' perspectives and mental states (House, 2018). Indeed, children who pass standard Theory of Mind tasks are more likely to behave fairly in a two-player social dilemma, which suggests that children who are capable of thinking about how their behaviour may be received are more likely to act in a way that will be deemed fair by the recipient (Takagishi, et al., 2010). This compliments neurocognitive research, which has suggested that brain areas associated with Theory of Mind are activated during cooperative behaviour in some economic games (Chiu et al., 2008; McCabe et al., 2001). Therefore, it seems plausible that our understanding of normative expectations is related to these mental-state reasoning abilities, and that both are built upon the development during infancy of capacities for joint attention (Tomasello, 2019).

Within Tomasello's (2019) framework, executive functioning is also noted as a more domain-general capacity that is believed to support children's normativity. When confronted with rules of behaviour that may not align with natural tendencies, children need to (i) remember the rule, (ii) inhibit self-motivated interests and (iii) remain flexible as to what is appropriate across constantly changing social contexts; all which require elements of executive functioning. As discussed further in Chapter 3, elements of executive functioning have been strongly linked to Theory of Mind abilities (Carlson &

Moses; 2001; Carlson, Moses & Breton, 2002; German & Hehman, 2006; Leslie et al., 2005; Marcovitch et al., 2015; Powell & Carey, 2017), suggesting that executive functioning may influence children's sensitivity to normative information by supporting the ability to infer mental states of others.

Alternatively, our motivation to conform to normative expectations may be an extension of the imitation abilities which develop during the first year of life (Carey, 2009; Tomasello & Carpenter, 2005). For example, children often engage in "overimitation" by copying all actions by an adult regardless of necessity (Lyons et al., 2007; Whiten et al., 2009). This is potentially relevant for the transmission of cultural information, including social norms, which requires accurate imitation (of adults and older children) by infants and young children (Richerson et al., 2016; Tennie et al., 2009), even if those actions seem arbitrary or unnecessary. Therefore, imitation, and specifically overimitation has been theorised to allow for the transmission of normative information and behaviour between individuals (Legare & Harris, 2016). Despite strong theoretical grounding, little empirical work to date has directly explored the link between children's social and cognitive abilities and the emergence of normativity. Yet, this is a necessary step if we are to truly understand the human psychology for social norms, so exploring the relationship between children's developing normativity and these mechanisms, as demonstrated across Chapters 3 and 4, will allow insights into how social norms and expectations influence behaviour. More importantly, testing this relationship at a time when these skills are just beginning to emerge will hopefully provide enough individual variation to statistically capture predictive relationships.

Two additional factors, culture and gender, are also believed to account for individual variation in (i) prosociality, (ii) socio-cognitive abilities, and potentially (iii) the development of normativity. As argued in Chapter 2, although there is currently a small pool of research to draw from, variation has been demonstrated in aspects of social norm understanding across cultural groups. Most of this research has focused on variation in fairness norms (e.g. Cowell et al., 2017; Paulus, 2015), although recent work has turned to children's understanding of normative information and links to prosociality in middle childhood (House et al., 2020a). Whilst it is currently unknown whether girls or boys respond to normative expectations in similar or different ways, there is substantial research which suggests there may be gender differences in prosociality, along with many of the theorised predictors of normativity. As such, throughout all empirical chapters, both gender and culture will be included in the exploration of children's sensitivity to approval and observation. Details of how they are included can be viewed in the relevant chapters.

In conclusion, normative expectations are believed to play a pivotal role in enforcing conformity to adaptive and prosocial behaviours within societies. Especially when observed, individuals act in

accordance with the group norms, as any violations witnessed can affect reputation within the social group. Although children show an awareness of social norms and sensitivity to observation from a young age, it is currently unknown how these interact in development to give rise to a specific sensitivity to normative expectations. Despite strong theoretical grounding, there have also been extremely limited empirical tests of which social or cognitive mechanisms support the emergence of this sensitivity later in childhood. In my PhD research I aim to address these gaps in the literature through both cross-sectional and longitudinal cross-cultural studies of the development of normative expectations in early to middle childhood. This work will identify when children become sensitive to this important aspect of normative behaviour and will have both theoretical and practical applications for our understanding of human social development.

1.4 Current Thesis

My own research is aimed at addressing existing gaps in the current literature on children's sensitivity to normative expectations. Firstly, I aim to explore when in development children begin to adapt their prosocial behaviour in response to observation, and whether this changes across levels of social expectation. Included in this exploration, I aim to test whether the developmental trajectory of normative expectations varies with gender, or across two distinct cultural groups. Secondly, I aim to explore how sensitivity to normative expectations is predicted both by concurrent cognitive skills and cognitive skills in infancy, to begin to understand the mechanisms which support children's understanding of normative information.

I adopt a combination of longitudinal and cross-sectional designs to explore these aims. The longitudinal sample were recruited when infants were 3 months and research visits at their homes were completed regularly until the COVID-19 pandemic (18-24 months of age) and resumed at 4 to 5 years old. The cross-sectional sample were recruited when children were between 4- and 11-years-old, and only took part in one or two testing sessions. Around half the samples, both longitudinal and cross-sectional, were from in and around the city of York in the UK, and the other half were from rural villages in the Masindi District in Uganda. Recruiting participants from these two distinct cultural groups allowed us to take a cross-cultural approach and compare a WEIRD (Western, Educated, Industrialised, Rich and Democratic; Henrich et al., 2010) and non-WEIRD cultural group. Given the variability in prosociality across cultures, a multi-cultural approach is of great value when attempting to address the link between social norms and prosocial behaviour, as it allows us to determine the similarities and differences in the developmental trajectories of children's sharing across social environments where social norms about sharing are different (House et al., 2020a), which may provide insight that monocultural work never can.

The UK and Uganda are ideal study sites for the present thesis because of differences in attitudes that they possess. For example, prior research with the longitudinal sample of UK and Ugandan participants has highlighted differences in parenting practices and socialisation goals for their children (Holden et al., 2022; further described in Chapter 2). Furthermore, there are country-level differences in these cultures attitudes towards norm compliance, with Uganda tending to be less tolerant of norm deviations in comparison to the UK (Uz, 2015). These differences are likely important when considering how children learn social norms: parental priorities and childrearing practices, along with national views about how strictly one should follow and enforce social norms, could have an impact on the social expectations placed on, and held by, children. Indeed, socialisation practices have been reported to shape early prosocial development (Kärtner et al., 2010), supporting the idea that parental goals or expectations can influence children's development. Previous work exploring predictors of social norm development is scarce, especially that which considers multiple cultural groups. In addition, theoretical groundings are Western-centric and do not make any explicit predictions about cultural variation in the development of normativity. Thus, our exploration of normative development in Uganda and how that compares to the UK is particularly novel. Moreover, research regarding gender and sensitivity to normative expectations is also limited, and as I will argue in Chapter 2, empirical work exploring prosociality in girls and boys is inconclusive. Therefore, an exploration of how normative expectations develops in girls and boys from distinct cultural groups is much needed, and this thesis aims to bridge this gap in the literature. I take an exploratory approach to investigating how sensitivity to normative expectations develops in these populations and whether there is any variation across cultural groups or gender (or whether these might interact). Specifically, I address the following research questions:

1. Is sensitivity to normative expectations evident in 4-year-old children, and does this vary with cultural group or gender? (Chapter 2)
2. Is sensitivity to normative expectations evident in 5- to 11-year-old children, and does this vary with cultural group or gender? (Chapter 2)
3. Is sensitivity to normative expectations in early childhood predicted by concurrent cognitive abilities, like Theory of Mind or aspects of executive functioning? (Chapter 3)
4. Is sensitivity to normative expectations in early childhood predicted by cognitive abilities in infancy, like joint attention or imitation? (Chapter 4)

1.4.1 Methodological Approach

My planned PhD thesis research aims to fill existing gaps in the literature by exploring how children's sensitivity to normative expectations changes with age in early and middle childhood, and how this might vary with culture and gender. I will do this by using a novel sharing paradigm which explores

children's sensitivity to normative expectations, both (i) when observed and (ii) when *not* observed. The within-subjects paradigm comprises four modified conditions of a dictator game, in which participants can choose how much of an endowment to keep for themselves, and how much to give to an unfamiliar recipient. During different manipulations, we vary both the experimenter's approval of sharing generously, and whether the participant is observed whilst sharing. If children share most when they are observed by an experimenter who approves of being generous, this would suggest that their behaviour is being guided by normative expectations. If children share most when observed by the experimenter, but they do not differentiate between the two approval conditions, it suggests that children are sensitive to observation, but are not specifically guided by normative expectations. If children share most when the experimenter approves of being generous, but they do not differentiate between observed and unobserved trials, it suggests that children may be sensitive to normative information but they are not sensitive to observation (specifically, to the principle that non-conformist behaviour may impact their reputation). This would indicate that they do not yet have a nuanced understanding of normative expectations in the way these are thought to impact adult behaviour. This paradigm, which is explained in more detail in Chapter 2, will allow us to understand more about the motivations driving children's prosocial behaviour, in an effort to tease apart whether children's sharing is based on their beliefs about what others approve of, or more simply, reinforced or learnt behaviour.

Testing sessions either took place in participants' homes (for all longitudinal and most 4-year-old cross-sectional participants) or in a laboratory setting at the University of York in the UK, and at Nyabyeya Forestry College in Uganda (for the 5- to 11-year-old cross-sectional participants). These differences in testing location reflect the priorities of the different stages of data collection. During the longitudinal data collection, an effort was made to make the sessions as easy and convenient for parents as possible to prevent attrition, which included alleviating the need for families to travel by conducting home visits. On the contrary, to maximise the sample size of the cross-sectional participants (and increase statistical power), testing sessions were held at the laboratory. This had the advantage of being able to test more participants on a single testing day, in comparison to only two or three home visits on one day. This was important, especially because the 5- to 11-year-old children were attending school in the UK, and so the availability of families was significantly reduced.

1.4.2 Outline of Thesis

In Chapter 2, I explore the developmental trajectory of children's sensitivity to normative expectations, by measuring their sharing behaviour across different manipulations of approval and observation as described above. I do this by testing young children (4 years of age; Part 1) and an older range of children (5 to 11 years of age; Part 2) in the UK and Uganda.

In Chapter 3, I investigate how 4-year-old children's sensitivity to normative expectations is related to concurrent cognitive abilities, such as Theory of Mind and executive functioning. This is motivated by theoretical perspectives which suggest that skills related to the representation and coordination of perspectives contribute to children's later normative self-governance (House, 2018; House & Tomasello, 2018; Tomasello, 2019). Both 4-year-old longitudinal and cross-sectional participants completed tasks to measure Theory of Mind and executive functioning abilities during the 4.5-year testing session, along with the Normative Expectations task. Exploring individual variation across these tasks will allow us to detect elements of cognition which may be supporting children's sensitivity to others approval and observation.

Finally, in Chapter 4, I took a longitudinal approach and explored how 4-year-old children's sensitivity to normative expectations is related to aspects of social cognition measured in infancy, such as joint attention, overimitation and general cognitive development, measured at 15 months, 18 months and 11 months, respectively. Specific theories dictate that joint attention (Tomasello, 2019) and imitation (Carey, 2009; Essler, Christener, Becher & Paulus, 2023; Tomasello & Carpenter, 2005) support aspects of social cognition which the human psychology for social norms is built upon. To ensure the specificity of any relationship between children's sensitivity to normative expectations and joint attention or overimitation, general cognitive ability was also explored as a predictor of children's later normativity.

Chapter 2: Normative Expectations in Early and Middle Childhood: The Influence of Culture and Gender

2.1 Abstract

Social norms are a powerful motivator of human prosociality, yet our understanding of how sensitivity to norms develops in early and middle childhood remains limited, particularly across diverse cultures or gender. We follow social norms by attending to beliefs about what others approve of (i.e. normative expectations), but how does this attention to social approval develop? Across two studies, girls and boys from the UK and Uganda were tested on their willingness to share, using a modified Dictator Game. In four experimental conditions we varied (i) whether the experimenter approved of generous sharing, and (ii) whether the experimenter observed the child's behaviour. Study 1 found a gender difference among 4-year-old children: girls shared more than boys overall, but boys' sharing was more sensitive to gaining approval – mirroring adult-like responses. These differences may reflect emerging variation in cognition or socialisation. In contrast, Study 2 found no evidence of a gender difference in the 5- to 11-year-olds. Children were sensitive to approval cues, but this did not reliably increase with age or interact with being observed. Observation effects were strongest in younger children, but diminished by middle childhood, which may suggest non-linear development of children's responsiveness to normative expectations. Despite differences in cultural attitudes towards norm-following, no cultural differences in children's sensitivity to normative expectations were observed in either study. Taken together, these findings suggest that future research on the development of social norms and prosocial behaviour must account for children's beliefs about what others approve of, and investigate how individual differences in underlying socio-cognitive mechanisms shape this emerging sensitivity.

2.2 Introduction

As discussed in Chapter 1, social norms offer an explanation, at least in part, to the unique and unparalleled prosociality exhibited by our species (Boyd & Richardson, 2009). These shared rules of behaviour, which are accepted and often enforced by members of a society, play a crucial role in maintaining cooperation by shaping expectations about appropriate behaviour (Bicchieri, 2006; 2016; Chudek & Henrich, 2011; Richerson et al., 2016). These expectations (or 'social expectations', as termed by Bicchieri, 2006; 2016) are grounded in an individual's beliefs about other people – specifically, beliefs about what other people are doing (i.e. empirical expectations), or beliefs about what other people approve of (i.e. normative expectations). Normative expectations are thought to be particularly important for driving behaviour in situations where an individual's self-interest conflicts with prosocial norms (Bicchieri, 2016). For example, tipping in a restaurant requires an obvious cost to the self, one which might want to be avoided for self-interested reasons, but if there is a personal belief that other people around you would disapprove of not leaving a tip, you would probably be more likely to do so to avoid that disapproval. Therefore, normative expectations can promote behaviour more aligned with social norms, as people wish to avoid punishment that others might impose on them for self-interested, or non-normative behaviour (Chudek & Henrich, 2011; Richerson et al., 2016).

Whilst there is substantial theoretical basis and empirical evidence which supports the relationship between normative expectations and prosociality in adults (Bicchieri & Muldoon, 2018; Chudek & Henrich; Kallgren, Reno & Cialdini, 2000; Raihini & McAuliffe; Richerson et al., 2016; Smith & Louis, 2008), there is very little research which has directly explored this relationship in children. What research exists suggests that between early and middle childhood, children's sensitivity to normative expectations is beginning to develop and influence their prosocial behaviour. This assumption is, in part, based on the developmental co-emergence of (i) children's increasing sensitivity to normative information (House et al 2020a; House & Tomasello, 2018); children's willingness to punish selfish behaviour, and (iii) societal variation in prosociality and fairness (Amir et al., 2024; Blake et al., 2015; Cowell et al., 2017; House et al., 2013; House, 2017; House, 2018; House et al., 2020a; Huppert et al., 2019; Rochat et al., 2009). Together, these suggest that at some point in early to middle childhood, children are becoming aware of normative expectations and are beginning to act in a way that society deems as acceptable (leading to variation in prosociality across societies). Whilst direct evidence of this relationship is lacking, indirect evidence can be drawn from the work on children's understanding of reputation.

It has been widely documented that adults often maximise their prosocial acts when in the presence of an audience who can witness their actions, both in lab-based economic games (Andreoni & Petrie, 2004; Bohnet & Fret, 1999; Christens, Dannenberg & Sachs, 2019; Reinstein & Reiner, 2012; Rege & Telle, 2004; Samek & Sheremeta, 2014) and naturalistic field studies (Alpizar, Carlsson & Johansson-Stenman, 2008; Alpizar & Martinsson, 2013; Bereczkei, Birkas, & Kerekes, 2007; Dannenberg, Johansson-Stenman & Wetzel, 2022; Izuma, Saito & Sadato, 2010; Lacetara & Macis, 2008; Soetevent, 2005). In circumstances where normative expectations are present, and punishment will be imposed for non-compliance with what others want you to do, it is crucial to pay attention to whether or not you are being observed. In other words, to maximise your reputational status and gain social approval, you will be more likely to act in accordance with your beliefs about what others expect of you (i.e. normative expectations) when you are being observed. On the other hand, individuals may be more likely to act with self-interest when doing so will *not* result in disapproval by others. This could happen (i) when others are not observing your behaviour or (ii) when others don't seem to want you to act in any particular way. Whilst there is limited evidence on the intersection of these two contextual factors, there is research on children's sensitivity to observation and reputation management as separate factors in early childhood.

Children's reputation management is believed to begin developing around the age of 5, and to continue becoming more sophisticated as they age. From 5 years of age, children understand that their behaviour contributes to the formation of a shared opinion with their peers (Hill & Pillow, 2006), and they can promote self-descriptions that are aligned with group preferences or socially desirable behaviours (Banerjee, 2002; Gee & Heyman, 2007). Also from 5 years, children will cheat less during a game if primed to think about how others perceive them (Fu et al., 2015). It is also around this age that children begin to understand the significance of when other people can judge their behaviour – specifically, when others are observing their actions. This has been demonstrated across numerous paradigms in which children tend to act more prosocially (and less antisocially) when observed, in comparison to when they are not observed (Engelmann, Hermann & Tomasello, 2012; Engelman et al., 2013; Leimgruber et al., 2012). These studies strongly suggest that from the age of 5, children are becoming aware of their reputations, and are paying attention to contextual factors which can influence them. However, as no study has directly explored in parallel both children's sensitivity to (i) observation and (ii) others' approval (i.e. normative expectations), it is difficult to form strong conclusions about the reasons underpinning the influence these two factors may be having on behaviour.

As explained in Chapter 1, the current research tells us that children are behaving differently when observed, but it cannot tell us *why* children are behaving differently. Is it because young children are

beginning to track what others approve of and so are acting in a way to avoid punishment or gain social approval, or is it because children have been reinforced to act a certain way when they're being observed, and so are adjusting their behaviour based on past experiences, but without a nuanced understanding of why? Whilst these two explanations present very similarly in terms of behaviour, the underlying psychology is very different: the first is characteristic of an adult-like motivation to comply with social norms, whereas the latter represents a much lower-level explanation related to learned behaviours. If these explanations are to be teased apart, research which explores children's sensitivity to observation across varying beliefs about what others approve of is necessary. Currently, no study has directly manipulated both expectations and observation, so the present research aims to evaluate the possibility that children from around age 5 years are responding to observation manipulations for reasons different to adults. This would provide important new information about the psychological mechanisms which support children's developing understanding of social norms in early childhood and permit clearer tests of theoretical predictions about the early cognitive precursors of social norms and expectations. However, there are good reasons to think that we should consider two additional factors when seeking to understand the development of children's understanding of social norms and prosociality: culture and gender.

2.2.1 Factors Influencing the Development of Normative Expectations about Sharing

2.2.1.1 Culture

The majority of existing developmental studies reported about children's sensitivity to normative information or observation have primarily involved Western populations (with a couple of exceptions; Fu et al., 2015; House et al., 2020a; House et al., 2020b). This reflects a consistent Western sampling bias in developmental psychology, despite calls for sampling diversity and criticisms of conclusions based on monocultural findings (Nielson et al., 2017). However, there is some existing research which shows that there is variation in norms regarding sharing across societies, and sensitivity to these norms seems to be influencing children's sharing behaviour. Children from Western societies show a preference for an equal split from 3- to 4-years-old (Smith, Blake & Harris, 2013), and research suggests that these children begin to share equally from the ages of 5 to 8 years (Benenson, Pascoe & Radmore, 2007; Blake & Rand, 2010; Hamman et al., 2011; Tomasello & Warneken, 2008). However, research so far suggests that egalitarian norms may not be as influential in other samples, compared to Western samples. For example, children in South Africa and Turkey on average shared less than 50% in a Dictator Game, whereas children in the United States, Canada and China shared in a more egalitarian way (Cowell et al., 2017). In another study, Ugandan children from a provincial area in Southern Uganda preferred to share unequally rather than to discard rewards to achieve an equal split, unlike children from the US or South Africa who preferred egalitarian outcomes even at

the cost of discarding rewards (Paulus, 2015). This suggests that Ugandan children do not have a strong preference for an egalitarian split, and in other research they even show a tendency to make hyper-generous offers (Scharpf, Paulus & Wörle, 2017). In this study, children within the Busoga community (living in a large city in Uganda) showed high levels of generosity which persisted irrespective of the social relationship they had with the recipient (e.g. friends or strangers). This lack of ingroup favouritism suggests yet another dimension of cultural variation, as Western children tend to favour in-group members during sharing tasks (Engelmann, Over, Herrman and Tomasello, 2013; Fehr, Bernhard & Rockenbach, 2008; Moore, 2009; Paulus & Moore, 2014). Therefore, what prior research exists suggests that children from non-Western samples may be less egalitarian and sometimes more hyper-generous relative to children from Western samples. These results strongly imply that particular sharing norms vary across different cultural groups and that this impacts children's sharing during childhood, however they don't tell us when during development children's sharing behaviour begins to align with the sharing norms of their society.

The cross-cultural research that currently exists points to a similar developmental trajectory of children's emerging sensitivity to normative expectations. Specifically, third-party punishment for selfish behaviour and cooperative behaviour develop similarly across societies (Amir et al., 2024; House et al., 2020b), which indicates that children from different cultural groups may share a similar trajectory of attending to normative information. Furthermore, a study by House et al., (2013) found that when approaching middle childhood, children's rates of prosociality diverged to be more aligned with behaviour of adults in their own societies. Furthermore, a later study indicated that across eight diverse societies, this divergence develops alongside children's developing responsiveness to social norms (House et al., 2020a), suggesting some commonalities in how children's sharing behaviour is shaped by the norms present in their societies. However, there is other research which might lead us to expect that there might be cultural differences in the development of social norm compliance.

For example, there is cultural variation in adult beliefs that may be related to normative behaviour, such as norm compliance, and the parenting practices and socialisation goals that parents hold for their children. Across a study of 68 countries, Uz (2015) reported differences in opinions on conformity to norms. Specifically, at the country-level, industrialised societies tended to have a looser attitude (i.e. more tolerant of norm violations), whereas more traditional societies tended to have a tighter attitude (indicating that their norms may be more explicit, strict and violations are more harshly punished), which follows prior sociopolitical theories and research about cultural tightness-looseness (Gelfand et al., 2006; Gelfand et al., 2011; Pelto, 1968, Triandis, 1989). These infer a raised pressure for children in traditional societies to attend to normative information from an earlier age, given the emphasis on obedience, and harsher punishments for getting it wrong. Furthermore,

previous research suggests that parenting practices and socialisation goals vary across societies (e.g. Keller et al., 2006), and these can influence children's prosocial behaviour. For example, in some societies, like Delhi in India, punishment is positively related to an increase in prosocial behaviour, whereas in other societies, like Münster in Germany, these punitive practices are negatively related to prosociality in children (Giner Torréns & Kärtner, 2017). Moreover, relational socialisation goals (such as those that prioritise children learning obedience), have been reported to promote prosocial behaviour in children, and some cultures emphasize relational goals more than others (Kärtner et al., 2010; Holden et al., 2022). These studies suggest that there exists cultural variation in parental beliefs that could influence the development of children's prosociality or understanding of social norms, but this has not yet been explicitly tested. On the whole, prior research strongly implies that we do not have strong prior predictions that we ought to observe cultural variation in the development of normative expectations, but that it is still important to investigate this variation as it remains a possibility.

2.2.1.2 Gender

Theoretical models of the psychology of normative expectations do not make clear predictions about gender differences, but there are numerous ways in which gender might play an important role. There is no shortage of research findings about gender differences in prosociality, and while gender effects are often found (though not always e.g. Benenson, Pascoe & Radmore, 2007; Harbaugh & Krause, 2000; Lucas, Wagner & Chow, 2008), the nature of these gender differences is not necessarily consistent. Some research suggests that girls are more prosocial than boys (e.g. Blake & Rand, 2010; Fabes & Eisenberg, 1998; Gummerum et al., 2008; Gummerum et al., 2010; House, Henrich, Sarnecka & Silk, 2013; Leman et al., 2009), while others finds that boys are more prosocial than girls (Zarbatany et al., 1985). Direct explorations into how girls and boys from different cultural groups share are limited, but some research suggests that there may be gender differences in sharing behaviour across different groups. Rao and Stewart (1999) compared sharing behaviour between Chinese and Indian girls and boys, and found that whilst Chinese girls shared more than Chinese boys, there was no difference between Indian girls and boys. While this suggests that cultural variation in gender differences is possible, there isn't strongly compelling evidence for consistent gender effects across a wide range of societies (House, Silk & McAuliffe, 2023). This implies that researchers ought to explore gender differences in their populations that they are studying, but that they should remain agnostic as to the nature of those gender differences until they have data from a specific community.

Through a normative psychology lens, inconsistent evidence for gender differences in prosociality across prior studies could be the result of the specific normative expectations that were present in each study context or task; girls and boys could be reacting to differences in their beliefs about what others approve of in particular situations. Whilst females tend to be rated as more empathetic and altruistic than males (Eisenberg et al., 2006; Fabes et al., 1999; Hine, 2017; Hine & Leman, 2013), both men and women have been argued to act in prosocial ways which reflect their gender-stereotypical roles (Eagly, 2009; Hine, 2017). Furthermore, given that knowledge about gender is believed to be learnt very early in development (Ruble, Martin & Berenbaum, 2007), it remains a possibility that children are forming normative expectations about their gender-specific role in helping across a variety of situations which demand different types of help. Situations in which different types of help are required might demand different characteristics, such as empathy or risk, and as such, girls and boys may feel they're expected to respond differently based on their beliefs about what is socially desirable for their gender. For example, a girl may be more likely to comfort because she believes others expect girls to act with empathy, just like a boy may be more likely to climb a tree to rescue a cat because he believes others expect boys to provide physical assistance (Gardner & Gabriel, 2004; Spence & Buckner, 2000). Moreover, some have suggested that these stereotypes may not be entirely culturally bound (Fabes & Eisenberg, 1998; Whiting & Edwards, 1973), suggesting there may be some commonalities in how different cultures believe prosociality should be differentially expressed for girls and boys. This need not imply complete invariability in the content of gender roles across societies, but rather suggest that some gender differences may be more common across societies (e.g. the expectation that girls to follow rules more than boys). This suggests that it isn't just important to explore gender differences in how children respond to normative expectations, but also to explore gender differences in very different cultural groups.

Research regarding how gender influences sensitivity to others' approval is scarce, especially in children, but there is some evidence that boys may be more sensitive to observation than girls. In a study by Takagishi et al., (2015), older boys (aged 9-10), but not younger boys or girls, were more generous during a Dictator Game when there was a possibility of being observed in comparison to private sharing. However, similar results have also been reported in younger boys: Dreman and Greenbaum (1973) reported that 5- to 6-year-old boys were less generous when anonymous, unlike girls whose prosocial behaviour was not affected by anonymity. Taken together, these studies imply that boys may be more sensitive to their beliefs about what level of generosity is approved of by others and act accordingly when observed, unlike girls who seem to show a blanket level of prosocial behaviour. However, as explicit information about whether sharing was approved of wasn't provided during these studies, it is difficult to form strong conclusions about how sensitive children were to others' approval and how much they were trying to avoid eliciting disapproval by acting differently

when observed. Other studies which have mainly focused on children's sensitivity to normative information (but not observation) have not found gender differences (e.g. House & Tomasello, 2018), so it could also be possible that children's normativity develops along a similar trajectory regardless of gender.

2.2.2 Motivation of the Current Research

Overall, current research tells us that in Western populations children begin to show some awareness of their reputation by about age 5, but also that the psychology for norms has not yet developed into a fully-fledged understanding of social norms. This should lead to a large amount of individual variation in 4- to 5-year-old children's concern with approval and understanding the importance of being observed, making this an important age to target when studying the development of the psychology of normative expectations.

Despite the evidence that both normative expectations and observation play a role in prosocial behaviour in human adults, we have few answers to questions about the developmental trajectory that gives rise to children's beliefs about how others evaluate their behaviour (i.e. normative expectations). In particular: at what point in development do normative expectations begin to influence behaviour? Further research is needed if we are to properly understand how normative expectations and observation begin to work together to shape children's behaviour (particularly prosocial behaviour) during early and middle childhood. If the interaction between observation and expectations is not clear in young children, it is possible that their sensitivity to observation, as seen in previous studies of reputation management, may be driven by alternative factors, such as children learning through reinforcement of specific behaviours how to act when others are around, but not a nuanced understanding of how their behaviour aligns with others' approval.

Across Study 1 and Study 2 we explored normative expectations in childhood. Study 1 explored normative expectations in a sample of children aged 4 years, which will be used in future chapters to investigate the cognitive factors which may shape the development of children's normative expectations in early childhood. Study 2 then extended this to investigate the further development of children's understanding of normative expectations in a wider age range, from 5 to 11 years, to assess how children's sensitivity to normative expectations changes during the period of middle childhood when prior research has found substantial changes in children's sensitivity to social norms.

In order to address the persistent sampling bias in developmental research (Nielson et al., 2017), and to probe the role of culture in the development of sensitivity to prosocial norms, we conducted Study 1 and Study 2 with samples of children from two different cultural contexts: the UK and Uganda. As discussed in Chapter 1, these samples were chosen as they varied on several dimensions of interest.

First, it's likely that children in the UK are more egalitarian than children in Uganda (Paulus, 2015; Scharpf, Paulus & Wörle, 2017), so it will be possible to explore how different sharing norms might interact with children's concern with approval. Second, on a national level, previous research has identified the UK as a looser culture than Uganda, where compliance to norms is more strictly enforced (Uz, 2015). Third, previous research with these specific communities has revealed that Ugandan mothers align more with relational socialisation goals (including obedience and social harmony) than UK mothers, which may impact the development of sensitivity to social norms (Holden et al., 2022).

2.2.2.1 Study 1

The current study employed a novel sharing paradigm which explores sensitivity to normative expectations in 4- to 5-year-old children from the UK and children from Uganda, both (i) when observed and (ii) when not observed. Prior to the experimental manipulations, children were trained on the basic Dictator Game (DG), in which participants could choose how much of an endowment to keep for themselves, and how much to give to an unfamiliar peer recipient. The primary purpose of this Training DG was to ensure children knew how to play the game, but it also provided a measure of children's baseline generosity in our samples from the UK and Uganda (Research Question 1). However, as it was always presented first and not included in the counterbalanced design, we analysed it separately from the subsequent experimental phase (Test Trials) as it would be subject to order effects. After completing the Training DG, the main experimental phase then began. The within-subjects 2-by-2 design comprised four modified conditions of the Dictator Game, across which we manipulated both: (i) the experimenter's approval of sharing generously, and (ii) whether the participant was observed whilst sharing (Research Question 2, Research Question 3). As part of these analyses, we investigated variation in children's normative expectations across culture (Research Question 2) and gender (Research Question 3).

2.2.2.2 Study 2

After exploring children's sensitivity to approval and observation in early childhood, Study 2 aimed to use the same paradigm with a wider age range of children, across middle childhood in the UK and Uganda (i.e. from 5 to 11 years). Baseline generosity in the Training DG was measured across both societies (Research Question 4), along with how children's sensitivity to normative expectations (i.e. the interaction between observation and approval) developed over the course of this age range (Research Question 5; Research Question 6). Again, we investigated variation in children's normative expectations across culture (Research Question 5) and gender (Research Question 6).

2.2.3 Study 1 Research Questions

Research Question 1: Does baseline sharing in the familiarity phase (Training DG) vary with cultural group or gender at 4 years?

As noted above, there are few theoretically motivated predictions about gender differences in the development of normative expectations, and the present study was not designed to test any such predictions. However, as gender differences in prosociality and sensitivity to observation have been reported and given that cultural variation in such gender differences ought to be accounted for, we considered the role of gender in children's behaviour in the current task. To our knowledge, no research has investigated gender differences in prosociality among children in Uganda, so we investigate gender differences but make no specific a priori predictions about the nature of these differences, or their interaction with culture.

Research Question 2: Is sensitivity to normative expectations evident in 4-year-old children, and does it vary across cultural groups, controlling for gender?

If children aged 4-5 years understand normative expectations as do adults, we would predict that they will share more when they are being observed, but this effect of observation ought to be stronger (i.e. more positive) in (i) a condition where the experimenter expresses approval for generosity relative to (ii) a condition where the experimenter is indifferent to being generous. In previous research children's behaviour during observation has been assumed to be grounded in others' expectations, one possibility is that children of this age may only be sensitive to observation and have not yet integrated this with their beliefs about what others approve of. If this is the case, children would share most when being observed, regardless of whether the experimenter expresses approval for being generous. This would suggest that children's generous behaviour during observation may be better explained by something other than a nuanced understanding of what others approve of (e.g. learning through experience and reinforcement how to behave when others are present, but not necessarily understanding the beliefs motivating others' reinforcement). Another possibility is that children of this age are sensitive to normative information, but not to observation, meaning that they would share most when the experimenter approves of being generous but would not differentiate between observed and unobserved trials. This would indicate that children do not yet have a nuanced understanding of normative expectations as they function in adults: they are attentive to what others approve of, but are not motivated to try and avoid causing disapproval in others (i.e. avoid violating others' expectations when they are watching).

As existing research indicates cross-cultural similarities in the developmental trajectory of children's emerging sensitivity to normative expectations (i.e. third-party punishment for selfish behaviour and cooperative behaviours seem to develop similarly across societies; Amir et al., 2024; House et al., 2020b) we did not necessarily predict cultural differences in how the relationship between approval and observation would influence children's prosocial behaviour. However, conducting this study across two cultural groups with different parenting practices (Holden et al., 2022), socialisation goals for children (Holden et al., 2022) and cultural attitudes about compliance with norms (Uz, 2015) would hopefully allow us to detect any cultural differences if they were present. With that said, we did not know of any prior data that would lead us to predict specific differences in the development of normativity across the UK and Ugandan population of study. Our exploratory approach would have, however, still allowed us to detect any substantial cultural differences given the previously reported group differences in goals, attitudes and behaviour.

Research Question 3: Does sensitivity to normative expectations at 4 years interact with gender?

Although not a primary focus of our study, for the reasons noted above it is important to consider how gender might have affected performance on the task, considering the research relating to gender differences in prosociality (Benenson, Pascoe & Radmore, 2007; Blake & Rand, 2010; Fabes & Eisenberg, 1998; Gummerum et al., 2008; Gummerum et al., 2010; Harbaugh & Krause, 2000; House et al., 2013; House, Silk & McAuliffe, 2023; Leman et al., 2009; Lucas, Wagner & Chow, 2008; Zarbatany et al., 1985). To test Research Question 2, we included gender as a control factor in the analysis. As gender appeared to be having an effect on children's sharing behaviour, we decided to run further exploratory analyses to understand whether gender influenced children's sensitivity to normative expectations, leading to Research Question 3.

2.3 Study 1

2.3.1 Methods

2.3.1.1 Study Populations

The data in this study were collected from a longitudinal and cross-sectional sample, both including participants from the UK and Uganda. The two populations sampled were either living in and around the city of York (UK) or in the Nyabyeya Parish, Masindi District (Uganda). Participants for each of the following studies were drawn from these sample populations. More in-depth details regarding participant ages and gender can be seen in the relevant 'Participants' sections.

The UK longitudinal participants were recruited from flyers handed out at Baby Sensory classes, advertisements on a website for parents in York (York Mumbler) and by word of mouth. All mothers and infants lived in or around York for the duration of the research project in permanent structures with plumbing and electricity. Cross-sectional participants were recruited using similar channels; flyers, advertisements on social media, word of mouth etc. Participants from the UK mostly identified as White British (82%). All mothers only spoke English with the child, all had a minimum of secondary school education, and all could read and write in English (see more in depth participant details in Holden et al., 2022).

Ugandan participants either spoke Swahili (57%), Alur (28%) or Lugbara (15%). The Ugandan longitudinal participants were recruited at village meetings, where female research assistants made oral presentations to explain the project to pregnant women. Any interested mothers registered the birth of their child with the research assistants to ensure accurate birth dates. Mothers who then met the eligibility criteria of speaking Swahili, Alur or Lugbara to their infant and who consented to join the project were registered for the study. Further cross-sectional participants were recruited by female research assistants visiting local areas and inviting mothers who had children of the relevant age. Ugandan participants lived in mud or brick houses within a compound of two or more buildings. No houses had plumbing or mains electricity but some had small solar panels. Water sources and latrines were located outside of the houses. Ugandan mothers spoke a median of 2 languages with their children (range = 1-5; although mothers and children were tested in their most dominant language), and their education and literacy levels were much more varied than the UK mothers (Holden et al., 2022); 29% could read and write in at least one language, and mothers either had no formal education (18%), primary school education (66%) or secondary school education (16%).

2.3.1.2 Participants

One-hundred-and-forty-one children were tested for Study 1, with a final sample of N = 129 (61 boys, 68 girls). Twelve children were excluded due to either experimenter error (n = 5), the mother observing the child during the task (n = 2), or children being under 4 years (n = 3) or over 5 years (n = 2) of age. Sixty-eight children were from the city of York, UK (36 girls), and 61 from the Masindi District, Uganda (32 girls). Included in the sample were 79 children who were part of a large longitudinal study that collected data with the same children from between the ages of 3 months and 5 years, but only data from the 4.5-year timepoint are used in this study. Within this longitudinal sample, 42 participants were from the UK (mean age = 4.60, SD = 0.11, 20 girls), and 37 were from Uganda (mean age = 4.67, SD = 0.12, 18 girls). The further 50 children were cross-sectional participants (i.e. additional participants who only completed one testing session) whose data were collected between the ages of 4 and 5 years and included 26 participants from the UK (mean age = 4.69, SD = 0.20, girls = 16) and 24 participants from Uganda (mean age = 4.52, SD = 0.30, girls = 14).

2.3.1.3 Ethics

Ethical approval was obtained from the Department of Psychology Ethics Committee at the University of York and the Ugandan Virus Research Institute Regional Ethics Committee, and the study was approved by the Ugandan National Council for Science and Technology. In the UK, written consent was obtained from mothers for their child's participation in the tasks. Due to low levels of literacy in Uganda, information sheets and consent forms were orally presented to mothers (or primary caregivers) in a language the mother was fluent in, by a research assistant fluent in both (i) their local language and (ii) English. If mothers/primary caregivers were comfortable using a pen, they signed the consent form, otherwise they provided a thumb print in lieu of a signature.

2.3.2 Procedure and Materials

2.3.2.1 Research Design and Procedure Overview

The experiment allowed children to share tokens (which could be exchanged for rewards) with four different peers across four distinct test trials, each of which represented a distinct combination of normative expectations (i.e. experimenter approval) and observation (in a 2x2 design). The procedure for Study 1 was conducted by local researchers who were fluent in the participant's dominant language in both the UK and Uganda. Data were collected as part of 2–3-hour home visits during which multiple experimental procedures were conducted. In Uganda these sessions usually took place outside (e.g. under a tree), whereas in the UK they were most commonly conducted inside homes. The procedure comprised several phases: the pre-test phase included the introduction, video

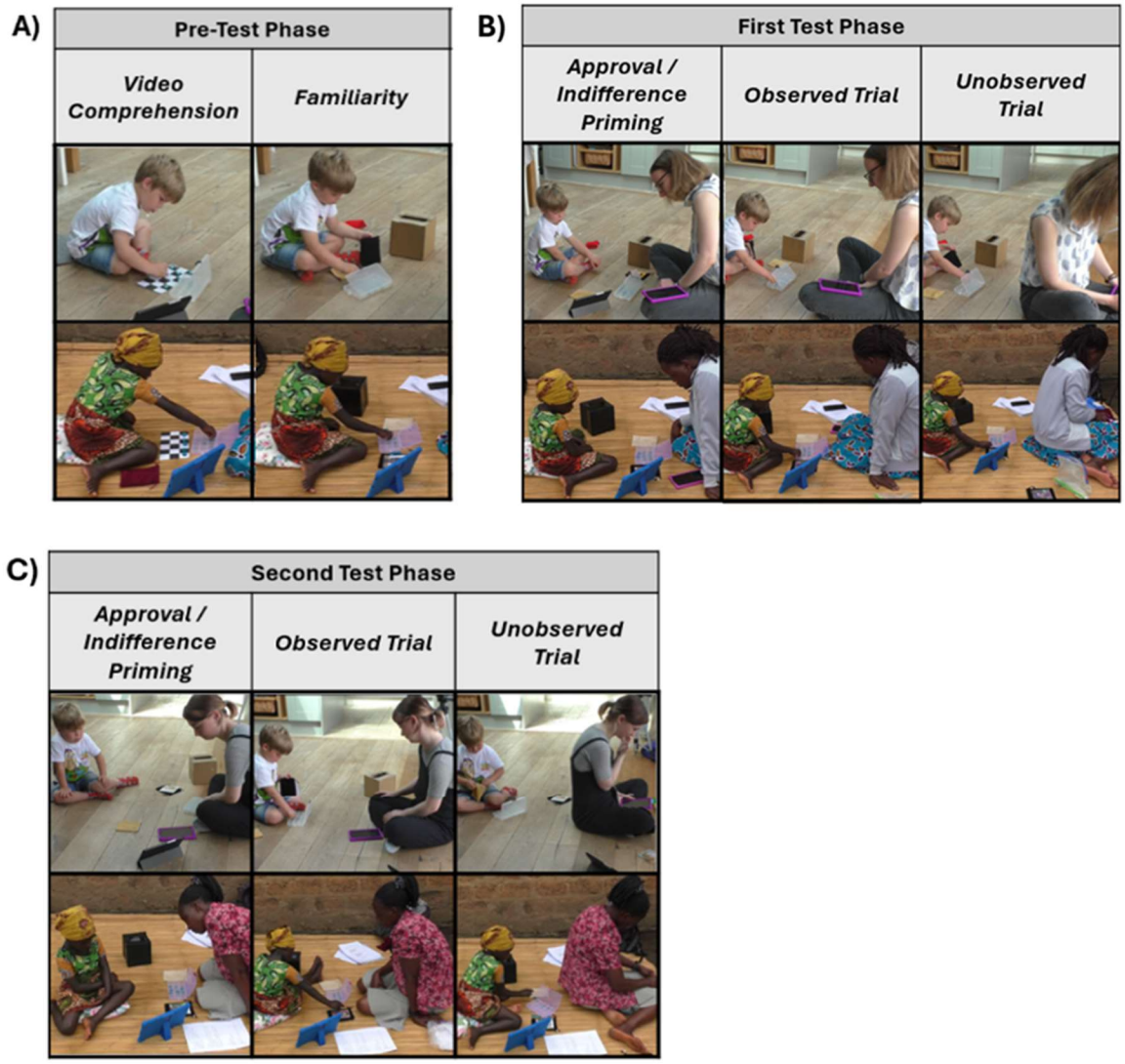
comprehension and familiarity phases, whilst the test phase included the first priming and first testing phases, followed by second priming and second testing phases (see Figure 2.1).

2.3.2.1.1 Introduction Phase

At the start of the full 2–3-hour testing session, children were introduced to tokens (small glass beads) which they could exchange for prizes at the end of the session, and questions were asked to ensure they understood that obtaining more tokens meant that they could obtain more prizes. Children were also shown photos of children to identify a set of photos of same-gender children they did not recognise, that would be used as the unfamiliar recipients of sharing behaviour in the test trials. For the UK sample, twenty-four photos of children (12 boys; 12 girls) were collected via free online photo sources (Freepik and Pexels, n.d.). For the Ugandan sample, the 24 photos (12 boys; 12 girls) were either (i) taken with approval by a colleague who was working at a school in a different district of Uganda or (ii) collected from Flickr, from a photo album from the World Bank labelled as being taken in Uganda and listed as having creative commons licences for re-use (World Bank Photo Collection, n.d.). It was confirmed by local research assistants that the photos used were plausibly children who lived around our study site in Uganda and were under the age of 12. Each participant initially viewed 5 same-gender photos of other children, and the photos used were counterbalanced across participants. Children were asked if they knew the children in the 5 photos to try and get a sample of photos which were unfamiliar to the child. If children said they recognised the children in any of the photos, they were shown the other photos in the set to find 5 photos of children which they reported they were not familiar with. If children said they recognised all 12 photos, the original set of 5 photos was used.

Figure 2.1

Example of each Experimental Phase of the Normative Expectations Task: A) Pre-Test Phase, B) First Test Phase, C) Second Test Phase. The top row shows a UK example, whilst the bottom row shows a Ugandan example.



2.3.2.1.2 Video Comprehension Phase

The video comprehension phase was designed to help us understand whether children could acquire information from videos that would be similar to those which we would show them in the test trials. In this phase, we presented children with two different videos which each expressed either approval or disapproval of how to engage in a task (i.e. approving of tokens placed in one location and disapproving of them being placed in a second location). Children were then asked to engage in the task themselves by imitating the behaviour of the model in the video. This indicated whether children

could extract from the video both (i) statements of approval and disapproval, and (ii) relevant behaviours associated with those statements.

During this phase, children were presented with a laminated checkerboard (containing 9 black squares and 9 white squares) and 9 tokens in separate compartments of a small transparent plastic box. On a tablet computer participants were shown two videos (the 'Like' and 'Don't Like' videos): in both videos the same adult model gave a statement of approval or disapproval before placing tokens from the plastic box in specific locations on the checkerboard. In the 'Like' video, the model said, "I like putting the tokens here", and then proceeded to place the tokens on the black squares. In the 'Don't Like' video, the model said, "I don't like putting the tokens here", and proceeded to place the tokens on the white squares (i.e. in the opposite location from the 'Like' video). In Uganda, these statements were translated into each of the three languages (Swahili, Alur and Lugbara) and then back-translated into English to check for accuracy, then for each language a set of videos was recorded using a model who was fluent in that language. The two videos were shown back-to-back (total duration = 36-46s depending upon the language), with the order being counterbalanced across participants. After watching the videos, the experimenter asked the child to put the tokens on the checkerboard in the same way that the model liked to place the tokens. If children refused to place the tokens, or placed the tokens incorrectly, the tokens were placed back in the transparent box and the child was asked to watch the video again. After watching the video for the second time, the experimenter pointed to a white square and asked, "Does my friend like it when a token is here?" This question was repeated whilst the experimenter pointed to a black square, another white square and another black square. Children's answers were recorded, and the trial proceeded regardless of the child's answers.

2.3.2.1.3 Familiarity Phase: Training Dictator Game

In the familiarity phase participants played a version of the Dictator Game (DG), wherein children allocated 9 tokens to themselves or to an unfamiliar same-gender child. This allowed them to become familiar with the format of the game before the test trials, which were based on a nearly identical token allocation procedure. In the familiarity task, participants were presented with 9 tokens placed in the same plastic box from the Video Comprehension Phase. The decision to use 9 tokens was based on trying to find the right balance between: (i) increasing individual variation in allocations (i.e. in case children were predisposed to give an equal split between themselves and the recipient), (ii) giving children the opportunity to allocate a substantial number of tokens (to increase sensitivity of the measure), but also (iii) avoiding the task taking too long. The plastic box was used because it placed each token in an individual compartment, which made it easier to track how children allocated each token in the set as a separate decision.

The experimenter explained to the children (i) that the tokens were all theirs, (ii) that another unfamiliar peer child (depicted in a photo) did not have any tokens, and (iii) that they could keep tokens (put them in their own bag) or give tokens to the other child (put them in a different bag with the unfamiliar peer's photo on it). Participants were told that they did not have to give any tokens, but if they wanted to give tokens, they could give as many as they wanted to. If participants could not correctly identify which bag was theirs and which bag was for the other child, the explanation was repeated and comprehension was checked again, until the participant could correctly identify the bags. Anonymity was established during the trial by making the number of tokens distributed by the child "secret" and by the child not being observed when allocating the tokens (i.e. Unobserved). After participants had completed distributing the tokens, they were asked to place the other child's bag in a cardboard box (referred to as the "secret box") that the experimenter said they did not look inside. The box measured 6"x6"x6" and had a rectangular slit in the top which allowed participants to place the bag inside.

After the child had been instructed on how the task worked, the experimenter turned around with their back to the child and also instructed anyone else in the vicinity (i.e. mothers) to do the same. This continued until the child finished distributing the tokens. If the child showed signs of confusion, or there was a prolonged period of hesitation after the experimenter had turned around, the experimenter asked the child if they would like to hear the instructions again. In these cases, the experimenter turned back around and repeated the instructions. The trial finished when participants had allocated the tokens, closed both their own bag and the peer recipient's bag, and placed the recipient's bag in the "secret box".

2.3.2.1.4 Experimental Test Phases

This point in the procedure marks the beginning of the testing phase, and the introduction of the experimental manipulations. The design consists of two conditions, each with two trials (so four trials in total). These trials include a 2x2 cross of experimenter approval (experimenter either [i] approved of generosity or [ii] was indifferent to generosity), and experimenter observation (experimenter either [i] observed the child allocating the tokens or [ii] did not observe). These manipulations are explained in more detail below.

<u>Test Block 1:</u>
<u>Generosity Priming</u>

In the first Priming Phase, children and the experimenter watched what we will call the "Generous Prime": a video of an adult model acting very generously in the DG (giving away 8 tokens and keeping just 1). The experimenter then expressed (i) approval or (ii) indifference about the way the

model had played the game (i.e. what we will call the "Approval Prime", see next section for details). The model in the Generous Prime video was the same model in the Video Comprehension task. The set-up of the task was very similar to the DG the child had played in the Familiarity Phase. After the model had given 8 tokens to the other child, and 1 to herself, the video ended. Due to pandemic travel restrictions delaying piloting in Uganda, the Ugandan protocol needed to be adjusted after testing had already begun in the UK, this resulted in some differences between the protocols. Children in Uganda were asked comprehension questions immediately after watching the Generous Prime video to ensure that they knew what the model had done in the video. However, children in the UK received these comprehension questions at the end of the task, and only if they failed the checkerboard comprehension (which occurred rarely; explanation for this difference in the protocol between sites is further explained in Appendix 2.1.1.1). Participants were included in the analysis regardless of their answers to the questions, as the questions were mostly used to determine whether children needed further instruction of the Generous Prime during the task in Uganda, rather than as exclusion criteria. This provided a way to ensure that children in Uganda were able to obtain information from the Generous Prime videos, given that their responses suggested that they often found the Video Comprehension Phase to be a challenge (see below).

Approval Priming

After watching the Generous Prime video, The experimenter then presented the participants with the Approval Prime: the experimenter said to the child either (i) that they approved of how the model played the game in the Generous Prime video ('Approval' condition), or (ii) that they were indifferent to how the model played the game in the Generous Prime video ('Indifference' condition). Importantly, unlike other studies in which participants receive information about what other people do (i.e. descriptive norms; e.g. Messer et al., 2021), keeping the behaviour of the model consistent across participants and instead manipulating the experimenter's approval of said behaviour allowed us to test specifically how children's beliefs about others' approval (i.e. normative expectations) influenced sharing behaviour. In the 'Approval' condition, the experimenter stated, "I like the way that person played the game", whereas in the 'Indifference' condition, the experimenter stated "I don't care if people play like that person. I like it when people play however they want to". These statements were carefully constructed based on piloting data in the UK to retain information about approval, whilst encouraging substantial individual variation for analyses. Whether the 'Approval' or 'Indifference' condition came first (i.e. during the First or Second Priming phase) was counterbalanced across participants.

Observation Manipulation: Test Trials 1 & 2

After hearing the Approval Prime, participants were presented with two test trials in which they played the DG again, but with novel peer partners in each trial. The observation manipulation was then implemented; one of these trials was unobserved (akin to the first DG), and the other one was observed by the experimenter (i.e. the experimenter did not turn around and instead watched the child distribute the tokens). The order of these trials was counterbalanced across participants. Upon completing these test phases, the first part of the task was finished, and children moved on to play different games within the testing session.

Test Block 2:
Approval Priming

After playing a series of other intervening games, the second Test Block was begun. Children watched the Generous Prime video again and a different experimenter expressed the alternative Approval Prime in response to the way the model had played (in comparison to Test Block 1; i.e. if the first experimenter said they approved of the Generous Prime in Test Block 1, a second experimenter said that they were indifferent to the Generous Prime in Test Block 2; or vice versa). In order to make it plausible that in this phase the experimenter had a different attitude about the Generous Prime video, we used a second experimenter who had not previously expressed an attitude about the Generous Prime video, while the first experimenter was busy with other paperwork. The second experimenter stated the same task instructions as the first (e.g. they never look inside the 'secret box'), and asked the participants to watch the Generous Prime video with them. The experimenter then gave the Approval Prime statement that was not given in the first priming phase (i.e. if the first experimenter said, "I like the way that person played the game", the second experimenter said, "I don't care if people play like that person. I like it when people play however they want to.")

Observation Manipulation: Test Trials 3 & 4

After hearing the Approval Prime, participants were presented with the final two rounds of the DG; one observed by the second experimenter, and one unobserved. The procedure was designed so that observed and unobserved trials were the same within children but counterbalanced across children.

2.3.2.1.5 Token Counting

Upon returning to the lab, the tokens in each of the recipient bags were counted and documented on the 'post-visit count' sheet.

2.3.2.2 Video Coding

2.3.2.2.1 Trial Validity

As described in the participant's section above, two children were excluded due to the mother observing during the task, which was problematic for two reasons: (i) the children's behaviour could have been influenced by maternal expectations and (ii) the unobserved trials were observed. Additionally, 5 children were excluded due to experimenter error, such as errors regarding the bags that were used during the trial (n = 1) and the experimenter giving the wrong Approval Prime (n = 4), which caused both conditions to be identical (i.e. either both approval or both indifference). Other minor deviations from the protocol that did not result in exclusion of data are detailed in Appendix 2.1.1.2. Finally, subsequent checking of children's ages also revealed that some children were under 4 (n = 3) or over 5 (n = 2) at the time of test, and so these were also excluded, meaning 129 children were included in the analyses.

2.3.2.2.2 Coded Variables

We coded, from the video recordings of the trials, whether children attended to the videos they were shown during the Video Comprehension Phase (i.e. the checkerboard video) and the Test Trials (i.e. the Generous Prime video), which was a descriptive check to ensure that children generally attended to the important information presented to them (Table 2.1). Children's comprehension for the checkerboard video in the Video Comprehension Phase was then assessed by firstly, coding whether they put the tokens on all the black squares of the checkerboard. Second, if they placed any tokens on white squares, they were shown the checkerboard video a second time and their answers to additional forced-choice questions were coded. Children's answers to questions about the Generous Prime video were also coded. The coding of children's comprehension across the Checkerboard and Generous Prime videos were also included as descriptive checks for whether children seemed to understand aspects of the task. Importantly, the accuracy of the Approval Prime was also coded in order to exclude any trials in which the approval or indifference statements were not delivered correctly to the children, and finally, the number of tokens kept and given across the familiarity phase and experimental phases were coded (for full details for how each variable was coded, please see Table 2.1). Coding was completed via video observation and entering information into a Microsoft Excel spreadsheet (Version 2301) by the lead author.

Although the coding of children's answers to questions about their comprehension of the Generous Prime was not intended to be used as exclusion criteria, we re-ran the most comprehensive model (testing Research Question 3) after excluding participants who did not pass the comprehension of the Generous Prime on the first attempt in both the 'Approval' and 'Indifference' conditions. This was to

ensure that the main pattern of results was not unduly influenced by children who were least likely to have successfully gained normative information from the Generous Prime video (as explained in the Statistical Analysis section and Appendix 2.1.3).

Table 2.1

Video coding criteria for the Normative Expectations task

Variables Coded	Definition
Attention to Checkerboard Video	Binary coded (yes/no) whether the participants attended to the checkerboard video. 'No' was coded if the child diverted their attention away from the video for more than 3 seconds at a time, or up to 3 seconds more than 5 times. If participants did not reach this limit, 'Yes' was coded.
Placement of Checkerboard Tokens	Binary coded (yes/no) whether participants successfully imitated the way the model liked placing the tokens on the checkerboard (all tokens on the black squares instead of the white squares)
Checkerboard Comprehension	If participants failed to place the tokens correctly, they watched the checkerboard again and were asked comprehension questions (e.g. "Does my friend like it when a token is here?" whilst pointing to two white squares and two black squares in succession). This variable was binary coded (yes/no) whether participants answered all 4 questions correctly.
Attention to Generous Prime video	Binary coded (yes/no) whether the participants attended to the Generous Prime video. 'No' was coded if children diverted their attention away from the video for more than 3 seconds at a time, or up to 3 seconds more than 5 times. If participants did not reach this limit, 'Yes' was coded.
Generous Prime Comprehension	Coded according to how participants answered the comprehension questions about the Generous Prime. Responses in Swahili, Alur or Lugbara were translated to English. Participants were first asked, "What happened in the video?" If participants successfully stated in the first free response question that the model gave more to the other person, this was coded as 'Correct – Free Response'. If participants did not give this answer, they were asked, "Did she give more to herself or more to the other person?". If participants correctly stated

Variables Coded	Definition
	the model gave more to the other person, this was coded as 'Correct – Binary Response'. All incorrect answers were coded as 'Incorrect' and all non-responses were coded as 'No Response'.
Accuracy of Approval Prime	Binary coded (yes/no) as to whether the experimenter delivered the Approval Prime correctly. Statements in Uganda were translated into English to check for any deviations from the intended meanings.
Tokens Given/Tokens Kept	The number of tokens given and tokens kept were coded by observing the videos. Any instances of uncertainty (e.g. poor visibility) were marked, and compared against the post-visit counting of the tokens in the bags.

Interobserver reliability: After training, a research assistant coded trials from 28 children (21.71% of the total sample), and interobserver reliability between the two video coders was calculated (using Cohen's Kappa; Cohen, 1960). Interobserver reliability was generally good. The coding of the number of tokens the children gave and kept showed excellent interobserver reliability between the two video coders (ICC(1, 1) = .94, 95% CI [.92, .96], F(138, 130) = 34.9, $p < .001$), based on a one-way random effects model with absolute agreement and single measure. Kappas for all but two categorical variables ranged between 0.74 and 1, indicating good inter-rater reliability. For the remaining two variables, the distributions of the coding were heavily skewed (for both coders), so even though the coders agreed on most cases the Kappas were very sensitive to a small number of disagreements (i.e. 1 - 2). One of these variables ("Attention to Checkerboard Video") captured whether children attended to the first checkerboard video they were shown, and the coders agreed on 26/28 cases (93.10% agreement). The second variable ("Accuracy of Approval Prime") captured whether the experimenter gave the Indifference statement correctly, in which the coders agreed on 27/28 cases (96.43% agreement), and there were zero cases in which the coders both coded the Indifference statement as being given incorrectly. This indicates a high level of agreement between the coders.

Reliability of post-visit token counts: For 15/129 children, the video provided poor visibility of the child distributing tokens in one or more of their 5 trials, meaning that the number of tokens given could not be accurately coded from the video and the number of tokens was taken from the post-visit

count (22/645 trials). To check the reliability of the post-visit counts for these children, we examined the correlation between the video coding and the post-visit count for these children's remaining trials where clear video was available for these children (N = 53 trials). The real-time coding and video coding were in perfect agreement (ICC(1, 1) = 1.00), based on a one-way random effects model with absolute agreement and single measures. Confidence intervals could not be estimated because the two methods produced identical values for all participants.

2.3.2.3 Statistical Approach

The main analyses were pre-registered on AsPredicted (Wharton Credibility Lab, n.d.) for this chapter, unless otherwise stated. Generalised Linear Mixed Models (GLMMs) were used for most analyses throughout this thesis, using the following statistical approach. Each model consisted of an outcome variable, which was always the proportion of tokens given and kept by the participant. In R, such an analysis of proportions is possible by using a two-columns matrix with the number of tokens kept, and number of tokens given for each trial per individual (Baayen, 2008). Multiple predictor variables and usually, the interactions between the predictor variables were included in the model, alongside random effects. The models were fitted in R version 4.2.3 (R Core Team, 2023), using the function 'glmer' of the package 'lme4' (Bates et al., 2015) with the 'bobyqa' optimiser.

Models that included multiple datapoints from the same participants included 'participant identity' as a random intercept to control for repeated observations of participants (i.e. pseudoreplication). Other included random intercepts were the identity of the experimenter on a given trial, and the order of conditions. Random slopes were included if deemed to be theoretically identifiable (i.e. there was enough variation within each level of the grouping factor), using a function written by Roger Mundry ('fe.re.tab'; personal communication). By including all possible random slopes, we maximise the precision of fixed effect estimates and keep the type 1 error rate at 5% (Schielzeth and Forstmeier, 2009). The random slope parameters were manually dummy coded and centred. Initially these were maximal models (Barr et al., 2013), including parameters for the correlations among random intercepts and slopes. However, if the correlations between the random slopes were estimated to have coefficients of exactly 1, which indicated that it was likely not identifiable (Matuschek et al., 2017), we removed the correlation parameters of the random slopes from the model which only ever led minor reductions in model fit. Normal distribution of the Best Linear Unbiased Predictors (deviations of each level of the grouping factors from the intercept) was tested using a function provided by Roger Mundry, and all were normally distributed.

We determined Variance Inflation Factors (VIFs) using the function 'vif' of the package 'car' (Fox & Weisberg, 2019), with models that excluded the interaction terms (as interactions involving factors

can lead to large VIFs). Collinearity was not an issue for any variables in our models (Quinn and Keough 2002). We either assessed model stability with dfbeta-values (Field, 2005) for Generalized Linear Models (GLMs), or using a function kindly provided by Roger Mundry ('glmm.model.stab.'; personal communication) for GLMMs, which estimates the stability of each estimated coefficient when excluding each level of the grouping factors one at a time (Nieuwenhuis et al., 2012). Model stability for the main predictor variables was acceptable for all models. Each model was checked for overdispersion using a function provided by Roger Mundry ('overdisp.test'; personal communication). If there were issues with overdispersion (i.e. the dispersion parameter was over 1.5), beta binomial models were used instead (Douma & Weedon, 2019). These models were fit in R Version 4.2.3 (R Core Team, 2023), using the function 'glmmTMB' from the package 'glmmTMB' (Brooks et al., 2017). The beta binomial models were checked for overdispersion using the functions 'resid', 'var' and 'df.residual' from the package 'stats' (R Core Team, 2023) and the function 'mean' and 'sum' from the package 'base' (R Core Team, 2023). If models were still overdispersed, an Observation Level Random Effect (Gelman & Hill, 2007) was included to account for unexplained variance. This approach usually results in underdispersed models, with more conservative tests.
To test the significance of the predictor variables and avoid cryptic multiple testing, the fit of the full model was compared to the fit of a null model, which comprised only the control variables (if applicable) and the random effects, using a likelihood ratio test (Dobson, 2002). An increasing number of predictor variables elevates the Type 1 error rate (Forstmeier & Schielzeth, 2011) because in the absence of any effect, each predictor has a 5% chance of revealing significance. Therefore, this approach protects against cryptic multiple testing by testing the predictor variables as a whole (which accounts for their number), against a model which does not contain the predictor variables. In the case that the full-null model is non-significant, this indicates that none of the coefficients are having a significant influence on the outcome variable, and as such, will not be further interpreted by examining the coefficients.

If the full-null model comparison is significant, it indicates that one or more of the predictor variables in the full model (which have been omitted from the null model) are having a significant impact on the outcome variable. In this case, the coefficients were then interpreted, beginning with the highest order interaction term. If the highest order interaction term was not significant, a reduced model was constructed, which excluded this interaction but retained the constituent lower-order interactions and/or main effects. For example, a non-significant 4-way interaction would be replaced with four 3-way constituent interactions, and if none of these were significant, these would be replaced with the constituent 2-way interactions, and so on and so forth until the model was reduced to (i) significant interactions and/or (ii) main effects. The structures of the final models are reported in the Statistical Analysis section, and details of the process followed to obtain each of these final models can be found

in Appendix 2.1.2. Confidence intervals of the model estimates were retrieved by using the function 'bootMer' of the package 'lme4' (1000 boots) for GLMMs, or using the function 'predict.glm' of the package 'stats' for GLMs. Any models not outlined in the 'Statistical Analysis' section, but that were included in the preregistered analysis plan, can be viewed in Appendix 2.1.4.

2.3.2.4 Statistical Analysis

Research Question 1: Model 1

Does baseline sharing behaviour in the familiarity phase (Training DG) vary with cultural group or gender at 4 years?

Whilst this analysis was not included in the pre-registered report, it is included for two reasons: (1) to give an overview of baseline generosity, adding context to the experimental manipulation; and (2) to indicate what kinds of effects of cultural group and gender (and interactions between these) might exist within these samples. Details of the model construction can be viewed in Table 2.2 (Model 1). See Appendix 2.1.2.1 for an outline of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 129 proportions of tokens given from 129 participants.

Research Question 2: Model 2

Is sensitivity to normative expectations evident in 4-year-old children, and does it vary across cultural groups, controlling for gender?

A description of this pre-registered model can be viewed in Table 2.2 (Model 2). See Appendix 2.1.2.2 for details of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 516 proportions of tokens given, from 129 participants over 4 trials.

Research Question 3: Model 3

Does sensitivity to normative expectations at 4 years interact with gender?

This analysis was not described in the pre-registered analysis report, but was included because of the unexpected and consistent effect of gender in the previous models. Theory makes no predictions about gender differences in the development of normative expectations, but given that gender differences are frequently observed and may vary between cultures in ways that are difficult to anticipate, we allowed gender to interact with all parameters from Model 2. A description of this model can be viewed in Table 2.2 (Model 2). See Appendix 2.1.2.3 for details of the process followed to reach the final model and stability parameters for the final model. The sample consisted of 516 proportions of tokens given, from 129 participants over 4 trials.

As described in the ‘Coded Variables’ section in the Methods, in order to test whether the results from this most comprehensive model were being unduly influenced by children who may not have understood the normative information in the Generous Prime video, we applied the same statistical analysis to a reduced data set, consisting only of participants who passed the Generous Prime comprehension on their first attempt in both the ‘Approval’ and ‘Indifference’ conditions (see Appendix 2.1.3 for details).

Table 2.2

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final models exploring the relationship between baseline sharing (Model 1), and sensitivity to normative expectations controlling for gender (Model 2) and exploring gender (Model 3).

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
1	Binomial GLMM	Tokens given|kept in Training DG	Cultural Group *Gender	NA	Observation ID	NA	0.31	Cultural Group = 1.00; Gender = 1.00
2	Binomial GLMM	Tokens given|kept in test trials	Observe *Approval *Cultural Group	Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[2] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[2]	1.27	Observe = 1.00; Approval = 1.00; Cultural Group = 1.00; Gender = 1.00
3	Binomial GLMM	Tokens given|kept in test trials	Observe *Approval *Cultural Group *Gender	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[2] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[2]	1.20	Observe = 1.01; Approval = 1.01; Cultural Group = 1.00; Gender = 1.00

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts were unidentifiable and therefore excluded from the model

2.3.3 Results

2.3.3.1 Descriptive Statistics: Mean Tokens Given

Although the data on tokens distributed in the training dictator game (in the familiarity phase) cannot be directly included in our main analyses (for the reasons described in the Introduction), the Training DG provides a sort of baseline measure of children's sharing in this kind of task without the additional features of our experiment (i.e. the Generous Prime, Approval Prime, and Observation by experimenters). For this reason, it is useful to examine descriptively how children shared in the Familiarity Phase vs. the Experimental Phase (although this was not included in the pre-registered report), which we do by providing the mean number of tokens given in Table 2.3. For the Ugandan boys, and all children in the UK, the average numbers of tokens given was higher in the Experimental Phase than in the Training Phase, though this was not true for the Ugandan girls. Within the Experimental Phase, the mean number of tokens given was highest across the approval conditions (relative to the indifference conditions), again for the Ugandan boys and all children in the UK, but not for the Ugandan girls.

Table 2.3
The mean number of tokens given across the Training DG (familiarity phase) and the experimental test trials.

	Mean Tokens Given			
Condition	UK		Uganda	
	Girls	Boys	Girls	Boys
Familiarity Phase				
Training DG	2.78	2.96	5.06	3.86
Experimental Phases				
Approval Observed	4.16	3.94	5.50	5.36
Approval Unobserved	4.27	3.59	5.00	5.39
Indifference Observed	3.84	3.36	5.88	4.25
Indifference Unobserved	3.57	3.12	5.75	5.14

2.3.3.2 Attention to Videos and Comprehension Questions

Dividing these descriptive statistics by gender was not part of the pre-registration, but we did this here because girls in Uganda seemed to be the only sub-group whose sharing behaviour did not increase from the Training DG (familiarity phase) to the Test Trials (Experimental Phase), which is the pattern that we would generally expect given the additional information provided in the Generous Prime. This could be due to an order effect for these participants, but it could also be that these girls were less aware of the information in the Generous Prime video. To investigate this possibility, we include information about Attention to Videos and Comprehension Questions by gender, to explore whether Ugandan girls in particular did not attend to, or understand, the task.

The children in the UK attended to the checkerboard videos in the Video Comprehension Phase (85-100%; Table 2.4) and passed the comprehension questions with relative ease (86% of the girls and 75% of the boys; see Table 2.4). Even though a similar proportion of Ugandan and UK children attended to the videos, far fewer Ugandan children passed the explicit comprehension questions. The checkerboard video depicts an activity that is unusual in general, but particularly so for Ugandan children who have likely had limited experience with games involving boards and counters or tokens. The Generous Prime videos were probably also generally unusual for children in Uganda, but at this point in the procedure they had already been introduced to the task materials (i.e. the tokens and the bags) in the Checkerboard Video and had used the materials themselves in the Training DG, so we would expect them to perform somewhat better on the questions. When we examine the additional comprehension questions that we added for Ugandan participants to assess their understanding of the Generous Prime video, these children indeed did much better. Children in both the UK and Uganda attended to the Generous Prime videos (Table 2.5), and Ugandan children passed the questions about the content of the Generous Prime at relatively high rates (Table 2.6). Boys and girls also answered the questions similarly (across the first and second presentations, 83% of girls and 89% of boys passed the questions in the Approval condition, and 80% of girls and 86% of boys passed the questions in the Indifference condition; Table 2.6). In the UK, Generous Prime comprehension was not checked immediately due to the differences in the protocol described in Appendix 2.1.1.1, however, when Generous Prime comprehension was checked at the end of testing with the children who failed the checkerboard video comprehension, 100% of girls (n = 5) and 83% of boys (n = 6) answered the Generous Prime questions correctly, which were similar to the rates of Ugandan children. This suggests that Ugandan children struggled specifically with the “checkerboard” activity, but not with extracting relevant information from videos.

Table 2.4

Summary statistics of whether children attended to and comprehended the checkerboard videos shown in the Video Comprehension phase of the Normative Expectation task, broken down by cultural group and gender.

Task	Group	Gender	N	N Successful (%)
Attended to the checkerboard video (first time shown)	UK	Girls	36	36 (100%)
		Boys	32	30 (93.75%)
	Uganda	Girls	32	29 (90.63%)
		Boys	29	28 (96.55%)
Placed the tokens on the checkerboard the way the model liked (all tokens on black squares)	UK	Girls	36	26 (72.22%)
		Boys	32	18 (56.25%)
	Uganda	Girls	32	3 (9.38%)
		Boys	29	2 (6.90%)
Attended to the checkerboard video (second time shown, if necessary)	UK	Girls	10	10 (100%)
		Boys	14	12 (85.71%)
	Uganda	Girls	28	25 (89.29%)
		Boys	28	22 (78.57%)
Answered questions correctly about how the model liked to place the tokens on the checkerboard (if necessary)	UK	Girls	10	5 (50.00%)
		Boys	14	6 (42.86%)
	Uganda	Girls	28	0
		Boys	27	0

Table 2.5

Descriptive statistics of how many 4-year-old children attended to the Generous Prime video prior to each test trial condition in the Normative Expectations task, broken down by cultural group and gender.

Group/ Gender	Percentage of Children Who Attended to the Generous Prime Video Prior to Each Test Trial Condition (%)	
	Approval Condition	Indifference Condition
UK		
Girls	91.67	86.11
Boys	93.75	87.50
Uganda		
Girls[1]	87.50	93.75
Boys[1]	93.10	86.21

[1]If there were two presentations of the Generous Prime video, child was coded as attended if they attended in either presentation

Table 2.6

Summary statistics of how 4-year-old Ugandan children answered questions about the Generous Prime in the Normative Expectations task.

Presentation of Video	Gender	Percentage of Children Who Correctly Answered Questions About the Generous Prime Video Prior to Each Test Trial Condition (%)	
		Approval Condition	Indifference Condition
After first Presentation	Girls	65.63	68.75
	Boys	65.52	65.52
After first or second presentation (combined)	Girls	83.87	80.65
	Boys	92.59	86.20

2.3.3.3 Research Question 1

Model 1: Does baseline sharing behaviour (in the Training DG in the familiarity phase) vary with cultural group or gender at 4 years?

Overall, the full model explained significantly more variation in the proportion of tokens given to the other child in the Training DG than the null model (likelihood ratio test: $\chi 2 = 13.91$, df = 3, $p = .003$). Although the model showed a trend towards an interaction between cultural group and gender (estimate = 1.09, SE = 0.65, $p = .091$; see Appendix 2.1.2.1), the final model revealed a significant main effect of cultural group (Table 2.7). This suggested that Ugandan children overall were more generous than children in the UK (Figure 2.2).

Table 2.7

Results of final Model 1 exploring 4-year-old children's baseline sharing behaviour in the Training DG (estimates and standard errors, together with confidence limits, and significance tests).

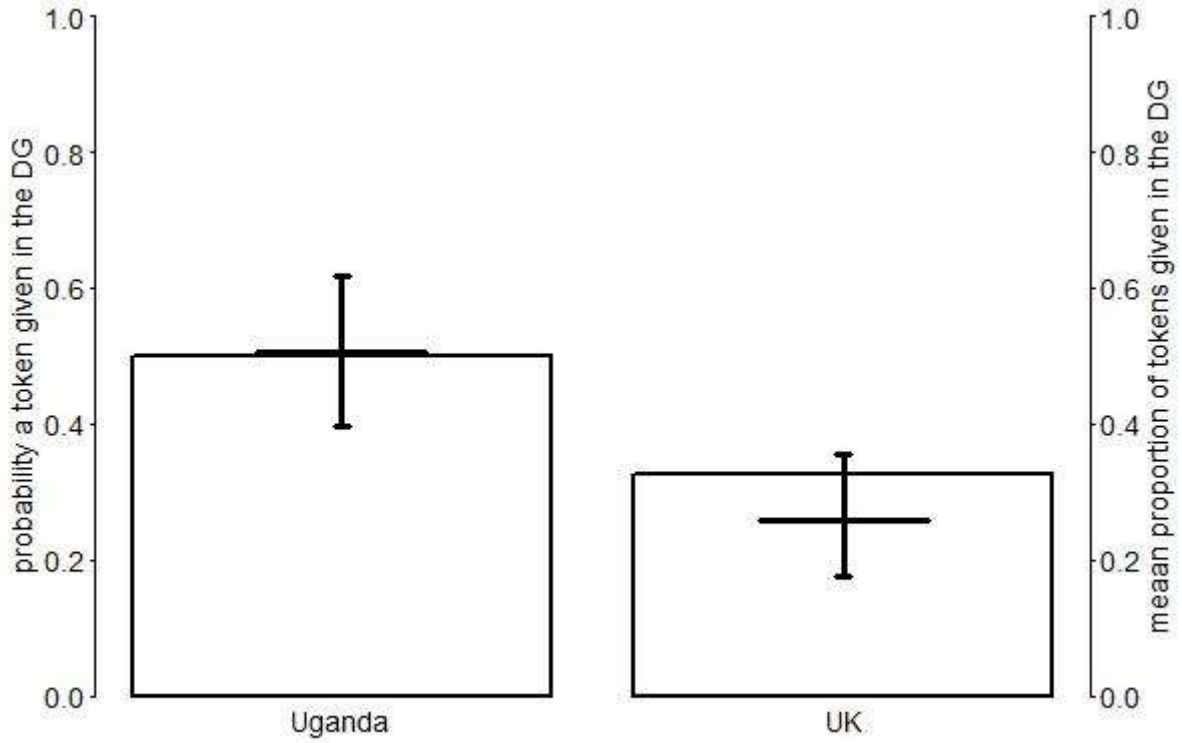
Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.14	0.28	-0.44	0.70	0.48	.631
Cultural Group[(1)]	-1.09	0.33	-1.74	-0.46	-3.31	.001
Gender[(2)]	-0.23	0.33	-0.87	0.40	-0.70	.483

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

Figure 2.2

Probability that a token is given in the Training DG for per level of Cultural Group (UK/Uganda). Bar plots show the raw data of the mean proportion of tokens given. Horizontal lines with error bars depict fitted model and its 95% confidence interval.



2.3.3.4 Research Question 2

Model 2: Is sensitivity to normative expectations evident in 4-year-old children, and does it vary across cultural groups, controlling for gender?

Overall, the full model explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi 2 = 18.41$, df = 7, $p = .010$). The model did not show a significant three-way interaction between 'approval', 'observe' and 'cultural group', so a reduced model was run without this interaction (see Appendix 2.1.2.2 for details). The reduced Model 2 results suggested there was a trend towards the interaction of 'observe' and 'approval' (estimate = 0.26, SE = 0.15, $p = .084$), and visualisation of the data (collapsing across the UK and Uganda) indicates a trend towards the pattern of normative expectations as previously defined (see Figure 2.3). With that said, given that this interaction was not significant, a final reduced model was run with only the main effects. There was a significant main effect of 'cultural group' (Table 2.8), suggesting that overall, Ugandan children were more generous than children in the UK. There was also a trend towards a main effect of 'gender' (Table 2.8), suggesting that girls were more generous than boys.

Table 2.8

Results of final Model 2 exploring 4-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.84	0.36	0.17	1.59	2.31	.021
Observe[(1)]	0.11	0.14	-0.16	0.39	0.78	.434
Approval[(2)]	0.17	0.19	-0.21	0.55	0.90	.370
Cultural Group[(3)]	-1.41	0.39	-2.17	-0.60	-3.59	<.001
Gender[(4)]	-0.75	0.39	-1.55	0.05	-1.93	.053

[(1)] dummy coded with the Unobserved condition being the reference level

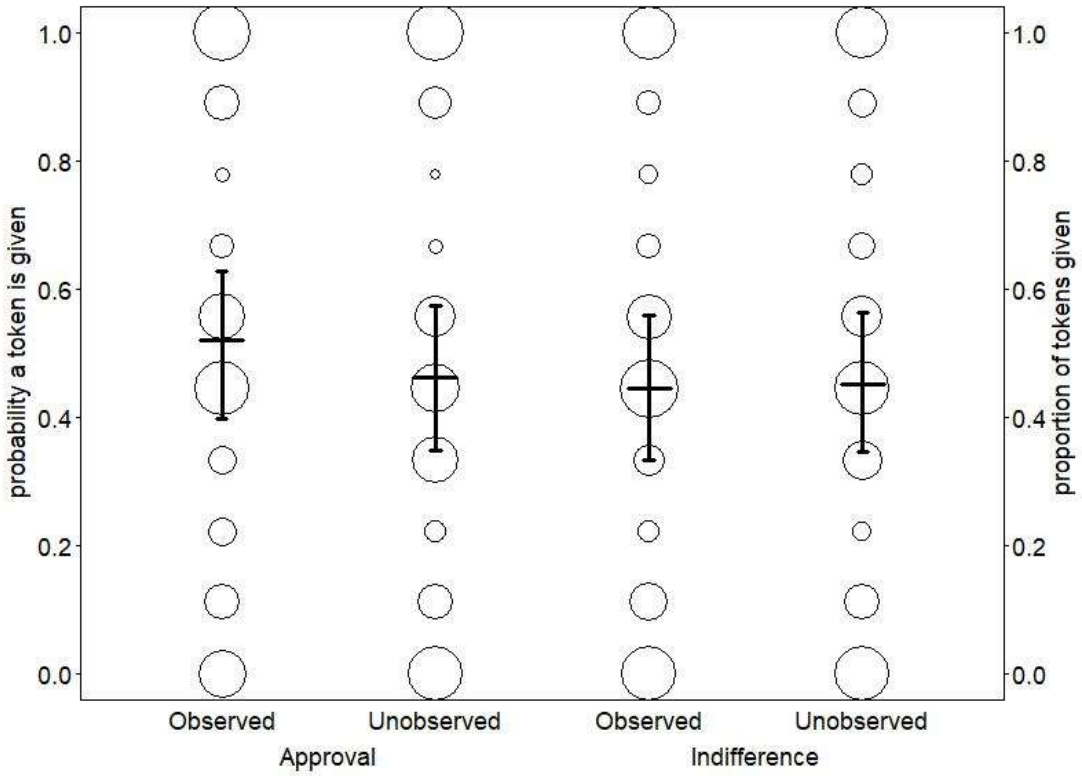
[(2)] dummy coded with the Indifference condition being the reference level

[(3)] dummy coded with the Ugandan sample being the reference level

[(4)] dummy coded with girls being the reference level

Figure 2.3

Probability that a token is given for per level of Observation (Observed/Unobserved) and Approval (Approval/Indifference) with gender centred. Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 1 to 28). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



2.3.3.5 Research Question 3

Model 3: Does sensitivity to normative expectations at 4 years interact with gender?

Overall, the full model explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi 2 = 64.30$, df = 15, $p < .001$). The model did not show a significant 4-way interaction between 'observe', 'approval', 'gender' and 'cultural group' (see Appendix 2.1.2.3), so a reduced model was run without this interaction. There was a significant interaction between 'observe', 'approval' and 'gender' (estimate = 0.82, SE = 0.31, $p = .009$; Table 2.9), and visualisation of the data suggested that the expected pattern of normative expectations (i.e. the observation and condition interaction) was stronger in boys than girls (Figure 2.4). As can be seen in Figure 2.4, the probability a token is given remains consistent for the girls across all four conditions, and the boys only match this level of generosity when observed by an experimenter who

approves of generosity. In the remaining four conditions (i.e. when children are unobserved, or in the indifference conditions), the boys have a lower probability of giving a token than the girls.

Table 2.9

Results of final Model 3 examining 4-year-old children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	1.12	0.43	0.25	2.06	2.61	.009
Observe(1)	0.13	0.29	-0.46	0.73	0.45	.652
Approval(2)	-0.51	0.29	-1.12	0.04	-1.75	.079
Cultural Group(3)	-1.76	0.57	-2.92	-0.65	-3.09	.002
Gender(4)	-0.74	0.62	-2.04	0.42	-1.20	.232
Observe*Approval	0.21	0.23	-0.31	0.78	0.82	.414
Observe*Cultural Group	0.07	0.35	-0.67	0.83	0.20	.845
Approval*Cultural Group	0.91	0.31	0.26	1.54	2.93	.003
Observe*Gender	-1.15	0.29	-1.72	-0.60	-3.91	<.001
Cultural Group *Gender	-0.10	0.83	-1.69	1.56	-0.12	.903
Approval*Gender	0.80	0.30	0.22	1.36	2.66	.008
Observe*Approval*Cultural Group	-0.47	0.30	-1.11	0.15	-1.55	.122
Observe*Cultural Group*Gender	1.04	0.32	0.45	1.70	3.29	.001
Approval*Cultural Group*Gender	-1.12	0.32	-1.72	-0.45	-3.47	.001

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Observe*Approval*Gender	0.82	0.31	0.19	1.51	2.62	.009

(1) dummy coded with the Unobserved condition being the reference level

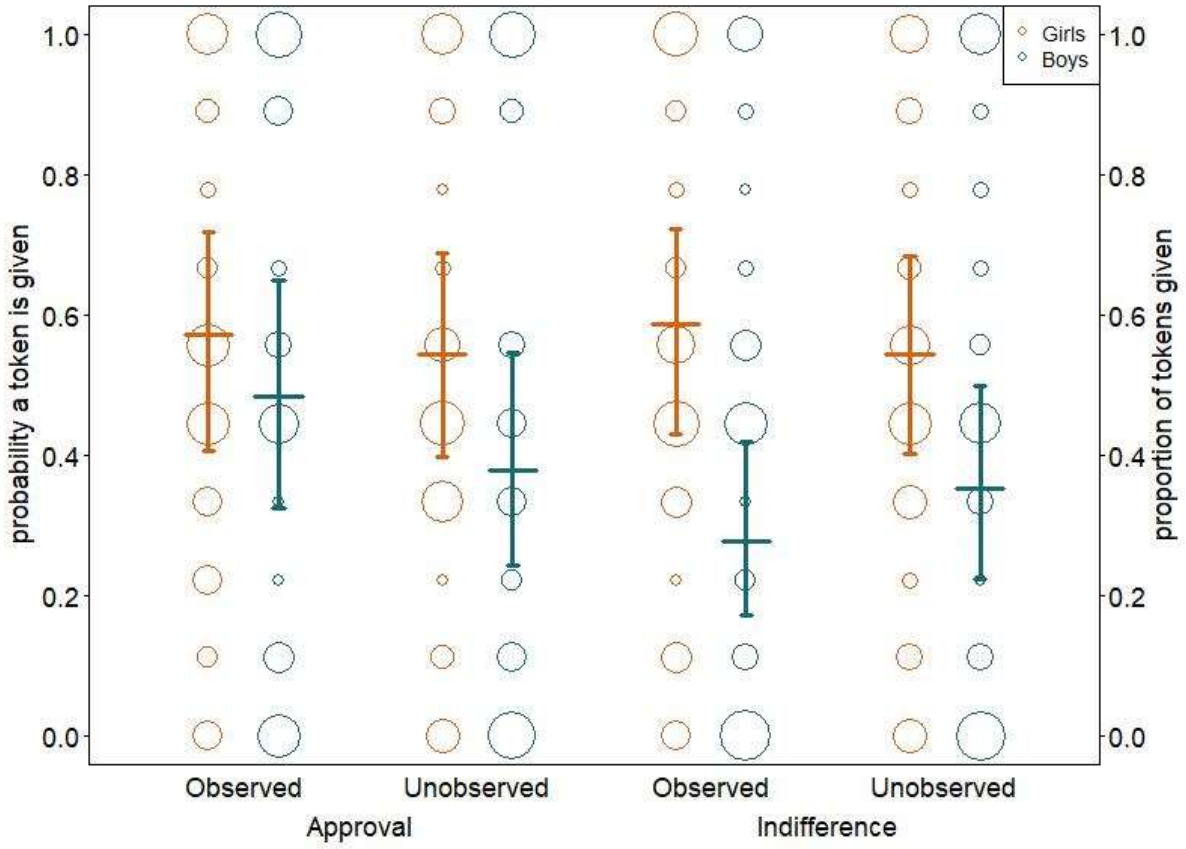
(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

Figure 2.4

Probability that a token is given for per level of Observation (Observed/Unobserved) and Approval (Approval/Indifference) for girls (orange) and boys (blue). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 1 to 18). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.

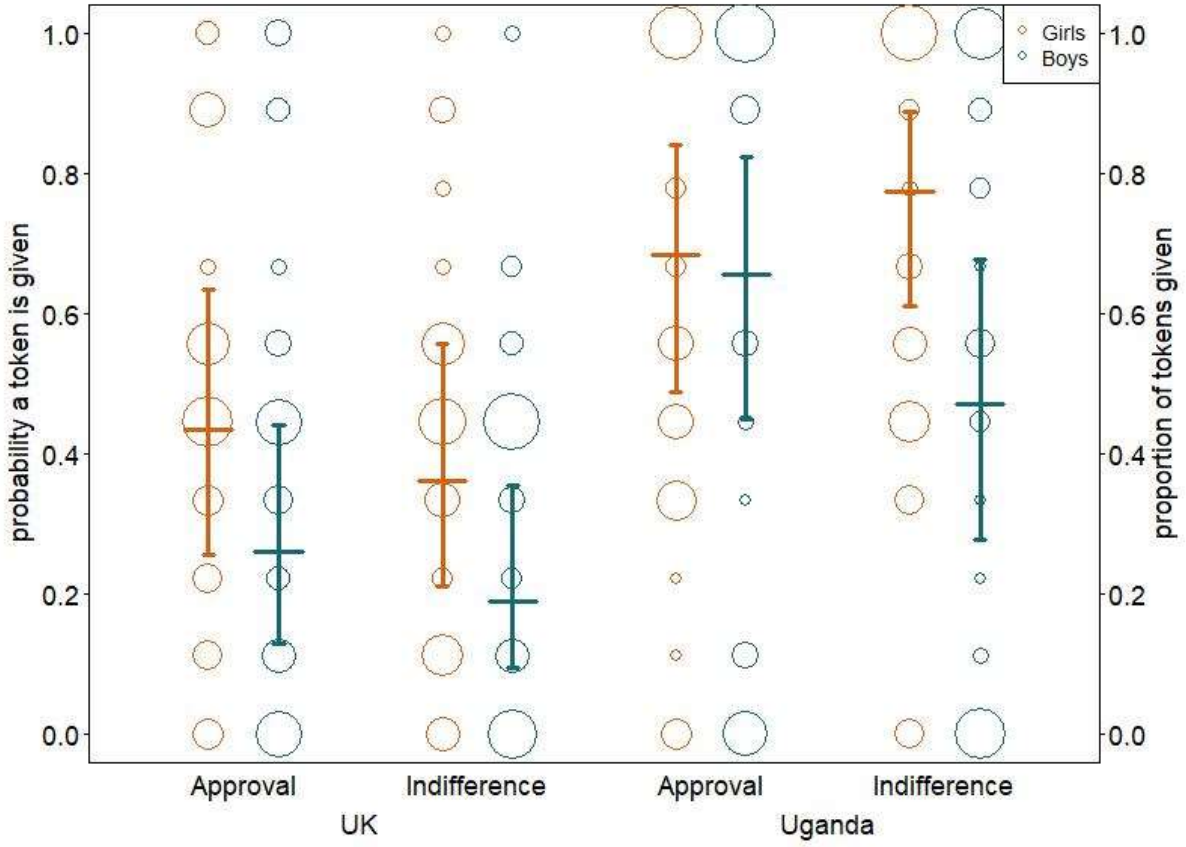


There was also a significant interaction between 'approval', 'cultural group' and 'gender' (estimate = -1.12, SE = 0.32, p = .001; Table 2.9). Visualisation of the data suggests that whilst most children (boys and girls in the UK, and boys in Uganda) showed a slight tendency to be more generous when

this was approved of by the experimenter (the expected pattern), girls in Uganda showed the opposite: they were somewhat less generous when generosity was approved of by the experimenter compared to when the experimenter was indifferent to generosity (Figure 2.5). Visualising this result doesn't present a clearly interpretable pattern of cultural variation in gender differences, but it's inclusion in the model means that we can have more confidence that the overall gender difference in normative expectations is not being driven by a culture-specific gender difference in the effect of Approval.

Figure 2.5

Probability that a token is given for per level of Approval (Approval/Indifference) and Cultural Group (UK/Uganda) for girls (orange) and boys (blue). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 1 to 25). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.

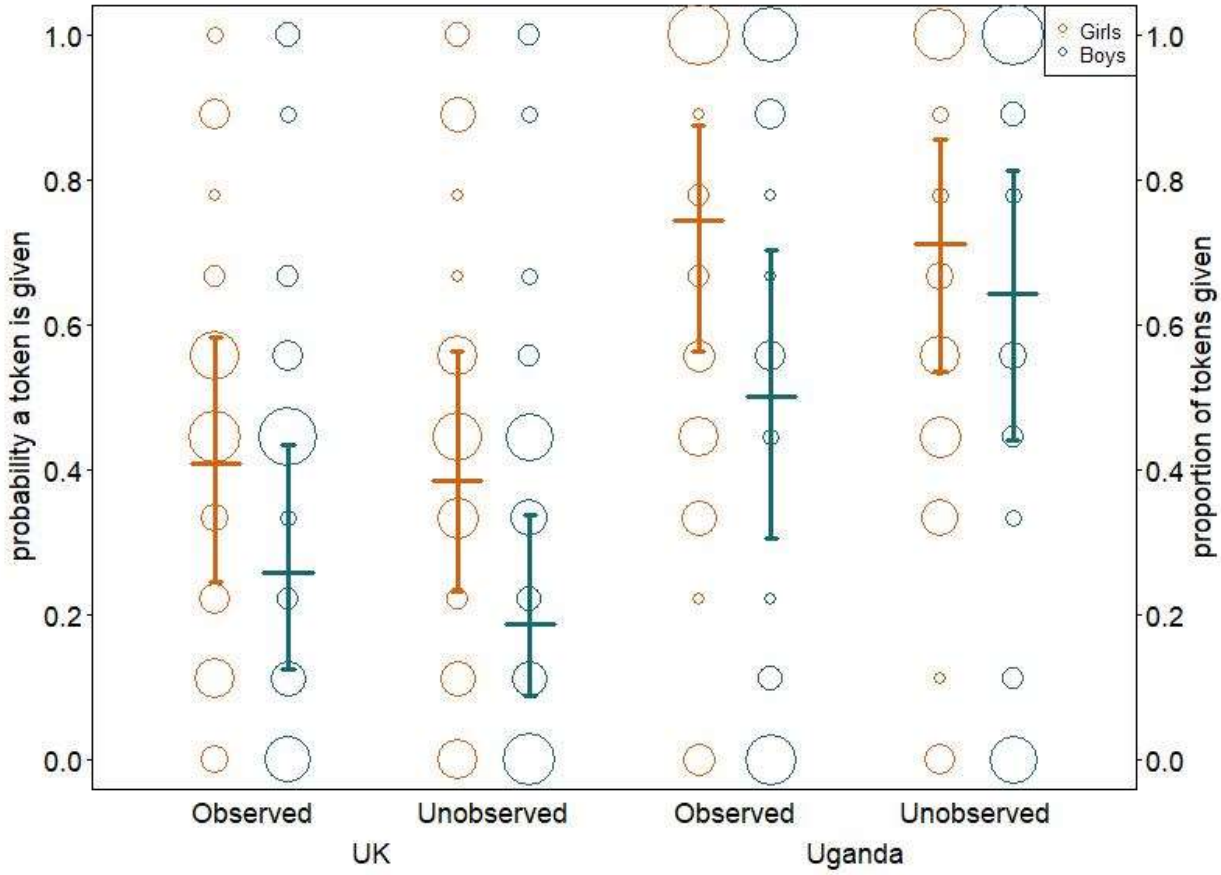


Finally, there was a significant interaction between 'observe', 'culture' and 'gender' (p = .001; Table 2.9). Visualisation of the data suggests that whilst most children (boys and girls in the UK, and girls in Uganda) showed a slight tendency to be more generous when observed by an experimenter (the expected pattern), Ugandan boys showed the opposite pattern, in that they were more generous when not observed, in comparison to when they were observed by an experimenter (Figure 2.6). Again,

visualising this result doesn't present a clearly discernible pattern of cultural variation in gender differences, but it's inclusion in the model means that we can have more confidence that the overall gender difference in normative expectations is not being driven by a culture-specific gender difference in the effect of observation.

Figure 2.6

Probability that a token is given for per level of Observation (Observed/Unobserved) and Cultural Group (UK/Uganda) for girls (orange) and boys (blue). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range:1 to 24). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



We also re-ran Model 3 with a reduced data set comprising of only the children who passed the Generous Prime comprehension questions on the first attempt in both Approval and Indifferent conditions (as explained in the 'Coded Variables' and 'Statistical Analysis' sections). The pattern of results did not change for any of the significant interactions reported for Model 3 (see Appendix

2.1.3.2), giving us confidence that the main interactions of interest were not being by certain groups of participants (e.g. Ugandan girls) not understanding the Approval statements.

2.3.4 Discussion

This study provides the first direct experimental study of young children's sensitivity to normative expectations in early childhood, by investigating children's sensitivity to others' approval when they are both observed and not observed. Results indicated that while a sensitivity to normative expectations was not strongly present in the overall sample of 4- to 5-year-old children (despite a trend in that direction), there is evidence for a gender difference in normative expectations. Boys showed a pattern of generosity consistent with normative expectations (as defined here), by being most generous when both (i) generosity was approved of by an experimenter and (ii) they were observed by that experimenter. In contrast, girls were more generous overall across conditions. Neither children's sensitivity to normative expectations, nor the gender difference in this sensitivity, reliably varied across our cultural groups (we will discuss this further in the Chapter Discussion). Our results do suggest that there may be some further cultural differences in how boys and girls react to both Observation and Approval as independent variables, but these are not likely to drive the overall gender difference in normative expectations that we observed.

The behaviour of boys in our task is broadly consistent with the existing literature on reputation in early childhood, which concludes that 5-year-old children adapt their behaviour (including prosocial behaviour) to align with others' approval (i.e. they "manage their reputation"). This stands in contrast to alternative explanations for children's behaviour, such as the idea that children are simply learning from experience to behave in specific ways (e.g. to share) when others are present and/or observing their behaviour. For instance, children's behaviour may have been reinforced in particular ways when others are present, and under similar conditions children simply repeat that behaviour without a nuanced understanding of *why* others are reinforcing them in certain ways (i.e. without a representation of what others approve of). This study adds additional weight to the claim that from the age of 4 years, boys at least are (i) forming expectations about what others approve of, and (ii) adjusting their behaviour to conform to those expectations specifically when others are watching (i.e. when deviating from the expectation would risk others' disapproval). We did not predict that this sensitivity to the intersection of Approval and Observation would be stronger in boys than girls, though this is in line with the limited amount of previous research that exists on this subject. This research suggests that boys are more sensitive to observation in prosocial paradigms than girls (Dreman and Greenbaum, 1973; Takagishi et al., 2015), and that boys engage in more strategic behaviour when building their reputations in an effort to maximise profits (Garbarini et al., 2014). This further strengthens the claim that in the present study boys were concerned about others' approval and responded to this strategically (i.e. when they were being observed). It also reduces the concern that boys of this age might respond to observation simply because of behavioural reinforcement. With that said, there are a number of alternative reasons which may explain why girls

overall did not vary their sharing behaviour according to others' approval and observation in a similar way to the boys, which will be discussed in the Chapter Discussion.

Whilst culture did not seem to reliably influence children's sensitivity to normative expectations (i.e. the primary focus of the research in question), culture did interact with effects of gender, observation and approval in unexpected ways. Both boys and girls from the UK tended to react to the manipulations in the expected directions: giving more when experimenters approved of generosity, and giving more when they were being observed. However, the results for Ugandan children are more paradoxical. Ugandan boys showed the expected pattern of response to the approval manipulation, meaning that they were more likely to be generous when an experimenter approved of generosity (regardless of observation), while the Ugandan girls did not. In contrast, Ugandan girls showed the expected pattern of responses to the observation manipulation (i.e. giving more when observed), but the Ugandan boys did not. This suggests that there may be some cultural variation in what girls and boys attend to when considering how to behave: Ugandan boys pay more attention to what other people say is important, whereas Ugandan girls attend more to whether people are watching them. One possible interpretation of these gender differences could relate to how girls and boys are socialised in Uganda, for example, the degree to which they are encouraged to focus on copying how others do tasks (i.e. making them more susceptible to observing and being observed) or to the degree they act in line with what is generally expected of them or approved of (i.e. making them more motivated to gain others' approval). These results should be replicated before strong conclusions are drawn from them, but future studies could also explore this cultural variation further by testing specific predictions based on cultural differences in children's socialisation. Regardless of how we might interpret these interactions, their inclusion in our statistical model controls for this variation and indicates that the overall gender difference in sensitivity to normative expectations is independent of these effects.

In line with our predictions about baseline generosity, before the manipulations were introduced, there were cultural differences in how children shared tokens. However, baseline sharing in the UK was lower than predicted based on previous research with Western children which suggests a preference for an equal split (Benenson, Pascoe & Radmore, 2007; Blake & Rand, 2010; Hamman et al., 2011; Smith, Blake & Harris, 2013; Tomasello & Warneken, 2008), as children tended to share fewer than 4/9 tokens on average. Other previous research has suggested that an equal split might be an upper limit to Western children's sharing behaviour, which is more in line with what was observed for UK children's sharing throughout the paradigm. McAuliffe, Raihani and Dunham (2017) found that 8- to 9-year-old American children were more influenced by a generous norm than a selfish norm in a dictator game, but even then, sharing behaviour was capped at 50%. Similar results have also been

reported in 4- to 8-year-old children from the UK (Messer et al., 2021). As predicted, there were cultural differences in baseline sharing behaviour, with Ugandan children sharing more overall than UK children. This may reflect an absence of particularly strong egalitarian norms in Uganda, which is consistent with previous research suggesting that Ugandan children can be hyper-generous in their offers (Scharpf, Paulus & Worle, 2017). Further support for the claim that egalitarian sharing norms may not be common in Uganda, comes from a recent study of the Ik people of Uganda (Townsend et al., 2020), which found that adult members of the group tended to be selfish in a control condition of the dictator game where no other information was provided (20/30 participants gave 25% or less, and 11/30 participants gave nothing). This suggests that different sharing norms may be present within Uganda as a whole, as this study was with adults who had a different ethnic background to those in our study, and were from the poorest region of Uganda (Townsend et al., 2020). Overall, our findings suggest that whilst Western norms of sharing are relatively consistent, there may be considerable diversity in sharing norms across Ugandan regions and ethnic backgrounds.

In conclusion, this study presents a snapshot of how sensitivity to normative expectations is evident in boys from as early as 4 years, in both the UK and Uganda, and whilst girls were more generous than boys, they did not respond to observation or statements of approval in the same way as boys. This novel task was created to explore emerging normativity at a timepoint in which it is theoretically *beginning* to shape prosocial behaviour, but using this task with children who are transitioning into and through middle childhood will allow us to understand how children's normativity becomes more *established* in terms of shaping prosocial behaviour. More research is needed to fully understand the developmental trajectory of how sensitivity to normative expectations develops across this wider range of childhood, and how these differences observed in girls and boys may vary with age.

2.4 Study 2

2.4.1 Introduction

Study 1 was able to give us a snapshot into children's understanding of normative expectations at about 4-5 years of age, a particularly important period of social development (e.g. for reputation and social norms). But, how does this understanding change over the course of childhood? There is evidence that children's psychology for norms is still changing from early to middle childhood. From the ages of 6- to 11-years, children's behaviour becomes increasingly guided by information about what is right to do (House & Tomasello, 2018; House et al., 2020a), which overlaps with the age range in which prosocial choices vary across societies (Blake et al., 2015; Cowell et al., 2017; House, 2017), and begin to mimic patterns of adult behaviour (House et al., 2013; House et al., 2020a). Together, these findings point to the continued development of prosocial normativity as children leave early childhood and enter middle childhood, on their way towards acquiring an adult-like sensitivity to social norms about prosocial behaviour. More specifically, these findings suggest that the signatures of normative expectations that we observed in Study 1 among 4-year-old children (particularly boys) might increase in strength at older ages as children approach adolescence and adulthood. Therefore, this study will examine older children's understanding of normative expectations, using a similar protocol to that described in Study 1, but with children aged 5-11 years in the UK and Uganda, leading to our first research question:

Research Question 4: Does baseline sharing in the familiarity phase (Training DG) vary with cultural group, gender or age between 5 to 11 years?

Based on the results from Study 1, and due to previous research suggesting that children with a Western background show a preference for an even split (Benenson, Pascoe & Radmore, 2007; Blake & Rand, 2010; Hamman et al., 2011; Smith, Blake & Harris, 2013; Tomasello & Warneken, 2008), but maintain a sharing cap at 50% (McAuliffe, Raihani & Dunham, 2017; Messer et al., 2021), we predict children in the UK will overall share fewer than half the tokens, but sharing levels will not be drastically low. Furthermore, based on the results from Study 1, we predict that overall sharing in Uganda will not have as much as an egalitarian focus as in the UK, with children sharing more generously or more selfishly than an equal split, based on prior research previously described (e.g. Paulus, 2015; Scharpf, Paulus and Wörle, 2017; Townsend, Aktipis, Balliet & Cronk, 2020). Previous research also suggests that children's generosity increases with age, both in Western cultures (e.g. Silk & House, 2012) and in comparisons of diverse cultural groups (House et al., 2013), so we predict both cultural groups will share more generously as age increases. As previously stated, this Training Dictator Game in the familiarity phase was not included in the experimental manipulation, so the

research questions pertaining to children's sensitivity to approval and observation (and the intersection between them) are separate and are as follows:

Research Question 5: Is sensitivity to normative expectations evident in 5- to 11-year-olds and does it vary across cultural group and ages, controlling for gender?

Given the reported universality of children's tendency to attend to normative prosocial information during middle childhood, despite the exact age being somewhat variable (House et al., 2020a), we predict that children will become more sensitive to the experimenter's approval, combined with observation, as they age in both cultures. In practical terms, we expect to see an interaction between the Approval Condition (i.e. whether the experimenter expresses approval of generosity), Observation (i.e. whether the experimenter is watching) and age.

Although we do not expect culture to influence this relationship, particularly given the results from Study 1, we will include the relevant analysis in order to detect cultural differences in case they are present, given that this specific relationship between approval and observation has not yet been examined cross-culturally across development. Furthermore, due to the gender differences found in Study 1, we will also include an exploratory analysis to determine whether the trajectory of children's prosocial normativity is different for girls and boys. This generates our next research question for Study 2:

Research Question 6: Does the development of children's sensitivity to normative expectations between 5- to 11-years of age interact with gender?

We do not have strong predictions about gender differences in the development of normative expectations, but prior research has found that prosociality can vary with age and gender (Eisenberg & Fabes, 1998; Silk & House, 2012). Previous research has not found a gender difference in how children's prosocial behaviour is guided by normative information over this age range (House et al., 2020a), so our prediction is that the gender difference we observed in Study 1 will diminish across this older age range.

2.4.2 Methods

2.4.2.1 Participants

Two-hundred-and-twenty-six children between the ages of 5- and 11-years were tested for Study 2, with a final sample of N = 211 (103 boy, 108 girl). Fifteen children were excluded due to either experimenter error (n = 3), the mother or another individual observing the child during the task (n = 5), or the child's age being unknown, or not between the ages of 5- to 11-years (n = 7). Ninety-eight children were from the city of York, UK (51 girls; mean age = 8.01, SD = 1.78), and 113 from the Masindi District, Uganda (57 girls; mean age = 8.27, SD = 1.88). Participants were from the same area as those described in the Study Population section. All Ugandan participants were fluent in Swahili, which is the language the task was conducted in.

2.4.2.2 Procedures and Materials

2.4.2.2.1 Procedure Overview

Data were collected during 2-hour testing sessions in which participants visited a research facility (both in the UK and Uganda) and participated in tasks in a quiet room. The procedure was similar to Study 1 (see Table 2.10 for summary); participants completed the same Normative Expectations task employed a within-subject design that comprised several phases: introduction, video comprehension, familiarity, first priming and first testing, followed by second priming and second testing (see Figure 2.1). The only notable change to the protocol was both samples (i.e. in the UK and Uganda) received the same comprehension questions about the Generous Prime video, meaning there was no longer any need for the comprehension questions at the end of the procedure for children that failed the Video Comprehension phase (i.e. with the checkerboard) in the UK sample. This change was implemented to avoid any concern that the additional Generous Prime comprehension questions in Uganda could have drawn more attention to the normative information, so these questions were also included in the UK procedure for a more balanced design. Implementing this in the design could have an impact on children's generosity (i.e. drawing more attention to how the model behaves could encourage children to give more overall), however, we expected that this change would influence the four conditions equally. Data were collected by local researchers who were fluent in the participant's language.

Table 2.10
Phases of the Normative Expectations task with descriptions

Phase	Description and purpose of phase
Introduction	Children were introduced to the tokens to ensure they understood they could exchange them for prizes at the end of the session, and were also shown photos of children to identify a set of same-gender photos they did not recognise.
Video Comprehension Phase	Two videos were presented which expressed either approval or disapproval of a model placing tokens on a checkerboard (once on the black and once on the white, respectively), and the children were asked to imitate the behaviour approved of by the model in the video (i.e. place the tokens on the black squares), to understand whether they can extract information regarding approval from videos.
Familiarity Phase	Participants played the Dictator Game (DG), where they allocated 9 tokens to themselves or an unfamiliar same-gender child, which allowed them to become familiar with the format of the game before the manipulations began.
First Priming Phase	Participants received comprehension questions to check children's understanding after viewing the Generous Prime video in both the Approval and Indifference conditions. Participants were asked, "What happened in the video?" If they did not explicitly state that the model gave more tokens to the other person, they were asked a binary choice question, "Did she give more to herself, or more to the other person?" If participants got this question incorrect, they were asked to watch the video again and the same questions were repeated. An experimenter then expressed approval of or indifference to the model's behaviour.
First Testing Phase	After hearing the approval or indifference statement about the generous behaviour displayed by the model in the Generous Prime video, participants were presented with another two trials in which they played the DG again, one unobserved and one observed.
Second Priming Phase	Similarly to the first priming phase, the comprehension questions were implemented after the child watched the Generous Prime video. A second experimenter then expressed approval of or indifference to the model's behaviour. Five children in Uganda were not shown the video for the second time after getting the Generous Prime wrong initially due to experimenter error, but their data were still included.

Phase	Description and purpose of phase
Second Testing Phase	After hearing the statement, participants were presented with the final two rounds of the DG; one observed by the experimenter, and one unobserved.
Token Counting	Post-session, the tokens in each of the recipient bags were counted and documented.

2.4.2.3 Video Coding

2.4.2.3.1 Trial Validity

Validity coding revealed that some children needed to be excluded from the analyses, as described in the participant's section above. Five children were excluded due to the mother or another individual observing during the task. This was problematic for two reasons: (i) the children's behaviour could have been influenced by maternal/other expectations, and (ii) the 'unobserved' trials would involve some observation. Additionally, 3 children were excluded due to experimenter error, such as: multiple bags being given to the child to divide the tokens into (instead of just one each for the child and recipient during the trial; n = 1), the experimenter forgetting to refill the tokens between the trials (n = 1) and the experimenter giving the wrong Approval Prime (n = 1) which caused both conditions to be identical (i.e. either expressing both approval or both indifference). Subsequent checking of children's ages also revealed that some children were under 5 (n = 4) or over 11 (n = 2) at the time of test, and one child's age in Uganda was unknown, and so these were also excluded, meaning 211 children were included in the analysis. Other minor deviations from the protocol that did not result in exclusion of data are detailed in Appendix 2.2.1.1.

2.4.2.3.2 Coded Variables

All video coding was completed by a research assistant. Coding was completed via video observation and entering information into a Microsoft Excel spreadsheet (Version 2301). The same variables were coded as in Study 1 (Table 2.1).

As in Study 1, to determine robustness of the results, the most comprehensive model (testing Research Question 6) was rerun excluding participants who did not pass the comprehension of the Generous Prime on the first attempt in both the 'Approval' and 'Indifference' conditions. This was to ensure that the results were not unduly influenced by children who may have struggled to gain normative information from the Generous Prime video (as explained in the Statistical Analysis section).

Interobserver reliability: To assess reliability, the lead author coded trials from 45 children (20.74% of the total sample), and interobserver reliability between the two video coders was calculated (using Cohen's Kappa; Cohen, 1960). Interobserver reliability was good for all categorical variables, with kappa's ranging between 0.89 and 1. The coding of the number of tokens the children gave and kept showed excellent interobserver reliability between the two video coders (ICC(1, 1) = .95, 95% CI [.94, .96], F(223, 220) = 41.1, $p < .001$), based on a one-way random effects model with absolute agreement and single measures.

Reliability of post-visit token counts: The videos for 19/211 children provided poor visibility in one or more of their 5 trials, meaning that the number of tokens given could not be coded from the video and the number of tokens was taken from the post-visit count (21/1055 trials). To check the reliability of the post-visit counts for these children, we examined the correlation between the video coding and the post-visit count for the remaining trials where clear video was available for these children (N = 74) The real-time coding and video coding showed excellent agreement (ICC(1, 1) = .98, 95% CI [.96, .98], F(73, 74) = 78.9, $p < .001$) based on a one-way random effects model with absolute agreement and single measures.

2.4.2.4 Statistical Approach

The same statistical approach described in Study 1 was applied to analyses in Study 2. The structures of the final models are reported in the Statistical Analysis section, and details of the process followed to obtain each of these final models can be found in Appendix 2.2.2. Any models not outlined in the 'Statistical Analysis' section, but that were included in the preregistered analysis plan, can be viewed in Appendix 2.2.4.

2.4.2.5 Statistical Analysis

Research Question 4: Model 4

Does baseline sharing behaviour in the familiarity phase (Training DG) vary with cultural group, gender or age between 5 to 11 years?

Whilst this analysis was not included in the pre-registered report, it is included to give an overview of baseline generosity, adding context to the experimental manipulation. Details of the model construction can be viewed in Table 2.11 (Model 4). See Appendix 2.2.2.1 for an outline of the process followed to reach the final model and model stability parameters. The sample consisted of 211 proportions of tokens given from 211 participants.

Research Question 5: Model 5

Is sensitivity to normative expectations evident in 5- to 11-year-olds and does it vary with age across cultural groups, controlling for gender?

A description of this pre-registered model can be viewed in Table 2.11 (Model 5). See Appendix 2.2.2.2 for details of the process followed to reach the final model and model stability parameters. The sample consisted of 844 proportions of tokens given, from 211 participants over 4 trials.

Research Question 6: Model 6

Does sensitivity to normative expectations between ages 5 and 11 years interact with gender?

This analysis was not described in the pre-registered analysis report, but was included because of the differences in young girls and boys sensitivity to normative expectations reported in Study 1. A description of this model can be viewed in Table 2.11 (Model 6). See Appendix 2.2.2.3 for details of the process followed to reach the final model and the model stability parameters. The sample consisted of 844 proportions of tokens given, from 211 participants over 4 trials.

As described in the 'Coded Variables' section in the Methods, in order to test whether the results from this most comprehensive model (Model 6) were being unduly influenced by children who may not have understood the normative information in the Generous Prime video, we applied the same statistical analysis to a reduced data set, consisting only of participants who passed the Generous Prime comprehension on their first attempt in both the 'Approval' and 'Indifference' conditions (see Appendix 2.2.3 for details).

Table 2.11

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final models exploring the relationship between baseline sharing (Model 1), and sensitivity to normative expectations controlling for gender (Model 2) and exploring gender (Model 3) for the 5- to 11-year-old children.

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
4	Binomial GLM	Tokens given|kept in Training DG	Cultural Group *Gender*Age	NA	NA	NA	1.38	Cultural Group = 1.01; Gender = 1.00; Age = 1.01
5	Binomial GLMM	Tokens given|kept in test trials	Observe *Approval *Cultural Group*Age	Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)|Experimenter on Trial (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[2]	0.80	Observe = 1.06; Approval = 1.09; Cultural Group = 1.02; Age = 1.03; Gender = 1.02
6	Binomial GLMM	Tokens given|kept in test trials	Observe *Approval *Cultural Group* Age*Gender	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)||Experimenter on Trial[2] (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[2]	0.80	Observe = 1.06; Approval = 1.09; Cultural Group = 1.02; Age = 1.03; Gender = 1.02

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

2.4.3 Results

2.4.3.1 Descriptive Statistics: Mean Tokens Given

Although the distributed tokens in the Training Dictator Game (in the familiarity phase) cannot be directly compared to the experimental conditions as it was not part of the counterbalanced design, it is useful to examine how children shared in this baseline condition, along with the experimental trials, to understand whether the manipulations worked (although this was not included in the pre-registered report). Therefore, the mean number of tokens given to the other child is shown in Table 2.12. In each group, the tokens given across the experimental manipulations increased from the baseline DG, particularly so for children in Uganda. The mean number of tokens given was highest across the approval observed condition for all groups, suggesting the paradigm broadly worked as expected.

Table 2.12

The mean number of tokens given across the DG (familiarity phase) and the experimental test conditions for 5–11-year-old children.

Condition	Mean Tokens Given			
	UK		Uganda	
	Girls	Boys	Girls	Boys
Familiarity Phase				
Training DG	3.55	3.47	3.75	3.59
Experimental Phases				
Approval Observed	4.63	4.68	7.11	5.56
Approval Unobserved	4.12	4.17	6.56	5.26
Indifference Observed	4.10	4.66	6.60	5.25
Indifference Unobserved	3.67	3.85	6.58	4.97

2.4.3.2 Attention to Videos and Comprehension Questions

Descriptive statistics were not pre-registered, but are included to give an idea of how children attended to the videos and responded to the comprehension questions, similarly to Study 1. The

children in the UK attended to the checkerboard videos and passed the comprehension questions with relative ease (91-92%; see Table 2.13). Though a similar proportion of Ugandan and UK children attended to the videos, far fewer Ugandan children passed the comprehension questions about the checkerboard, similarly to Study 1. This may further suggest that this particular activity (placing tokens on specific squares) might be a very unfamiliar task to children in Uganda (seemingly regardless of age), in comparison to children in the UK, who often anecdotally remarked that the set-up was similar to chess or checkers. As the protocol had been altered for this sample to include the Generous Prime questions for both cultures, we were able to compare the comprehension of the Ugandan and UK children. Indeed, when we look at the comprehension questions for the Generous Prime questions in Uganda, children did much better and were more comparable to the UK. Children in both the UK and Uganda attended to the videos (Table 2.14), and passed the comprehension questions at relatively high rates (100% of children in the UK across both conditions, 100% of Ugandan girls and 98% of Ugandan boys in the Approval condition; 98% of Ugandan girls and 92% of Ugandan boys in the Indifference condition; Table 2.15), suggesting that children in both groups had equally attended to and understood the normative information presented to them.

Table 2.13

Summary statistics of whether 5- to 11-year-old children attended to and comprehended the checkerboard videos shown in the Video Comprehension phase of the Normative Expectation task, broken down by cultural group and gender.

Task	Group	Gender	N	N Successful (%)
Attended to the checkerboard video (first time shown)	UK	Girls	51	50 (98.04%)
		Boys	47	47 (100%)
	Uganda	Girls	57	55 (96.49%)
		Boys	56	54 (96.43%)
Placed the tokens on the checkerboard the way the model liked (all tokens on black squares)	UK	Girls	51	47 (92.15%)
		Boys	47	43 (91.49%)
	Uganda	Girls	57	20 (35.09%)
		Boys	56	14 (25.00%)
Attended to the checkerboard video (second time shown, if necessary)	UK	Girls	4	3 (75.00%)
		Boys	4	3 (75.00%)
	Uganda	Girls	37	36 (97.30%)
		Boys	42	39 (92.86%)
Answered questions correctly about how the model liked to place the tokens on the checkerboard (if necessary)	UK	Girls	4	2 (50.00%)
		Boys	4	4 (100%)
	Uganda	Girls	37	5 (13.51%)
		Boys	42	7 (16.67%)

Table 2.14

Descriptive statistics of how many children attended to the Generous Prime video prior to each test trial condition in the Normative Expectations task, broken down by cultural group and gender (5-11-year-olds).

Cultural Group	Gender	Percentage of Children Who Attended to the Generous Prime Video Prior to Each Test Trial Condition (%)	
		Approval Condition	Indifference Condition
UK	Girls	94.12	92.16[1]
	Boys	87.23	87.23
Uganda	Girls	98.25[1]	98.25[1]
	Boys	98.21[1]	98.21[1]

[1]Children could receive up to two presentations of the Generous Prime video, and if they received two presentations of the video then they were coded as having "attended" if they attended to at least one of those two presentations.

Table 2.15

Summary statistics of how 5- to 11-year-old children answered questions about the Generous Prime in the Normative Expectations task.

Cultural Group	Presentation of Video	Gender	Percentage of Children Who Correctly Answered Questions About the Generous Prime Video Prior to Each Test Trial Condition (%)	
			Approval Condition	Indifference Condition
UK	After first presentation	Girls	100	92.16
		Boys	100	100
	After first or second presentation (combined)	Girls	NA	100
		Boys	NA	NA
Uganda	After first presentation	Girls	92.98	89.47
		Boys	98.21	83.93
	After first or second presentation (combined)	Girls	100	98.25
		Boys	98.21	92.86

2.4.3.3 Research Question 4

Model 4: Does baseline sharing behaviour in the familiarity phase (Training DG) vary with cultural group, gender or age between 5 and 11 years?

There was no significant 3-way interaction between 'cultural group', 'age' and 'gender' (Appendix 2.2.2.1), so a reduced model was run without this interaction. There was a significant interaction between cultural group and age (Table 2.16), whereby generosity in Uganda remained somewhat constant between the ages of 5- to 11-years, but generosity increased in the UK over this age range (Figure 2.7). Gender was not a significant predictor of baseline sharing in the familiarity phase (Table 2.16).

Table 2.16

Results of final Model 4 exploring 5- to 11-year-old children's baseline sharing in the Training DG in the UK and Uganda (estimates and standard errors, together with confidence limits, and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	-0.33	0.08	-0.49	-0.18	-4.22	<.001
Cultural Group(1)	-0.07	0.09	-0.25	0.12	-0.70	.482
Gender(2)	-0.04	0.06	-0.16	0.08	-0.63	.530
Age(3)	-0.07	0.09	-0.26	0.11	-0.76	.449
Cultural Group*Age	0.29	0.10	0.10	0.48	3.01	.003

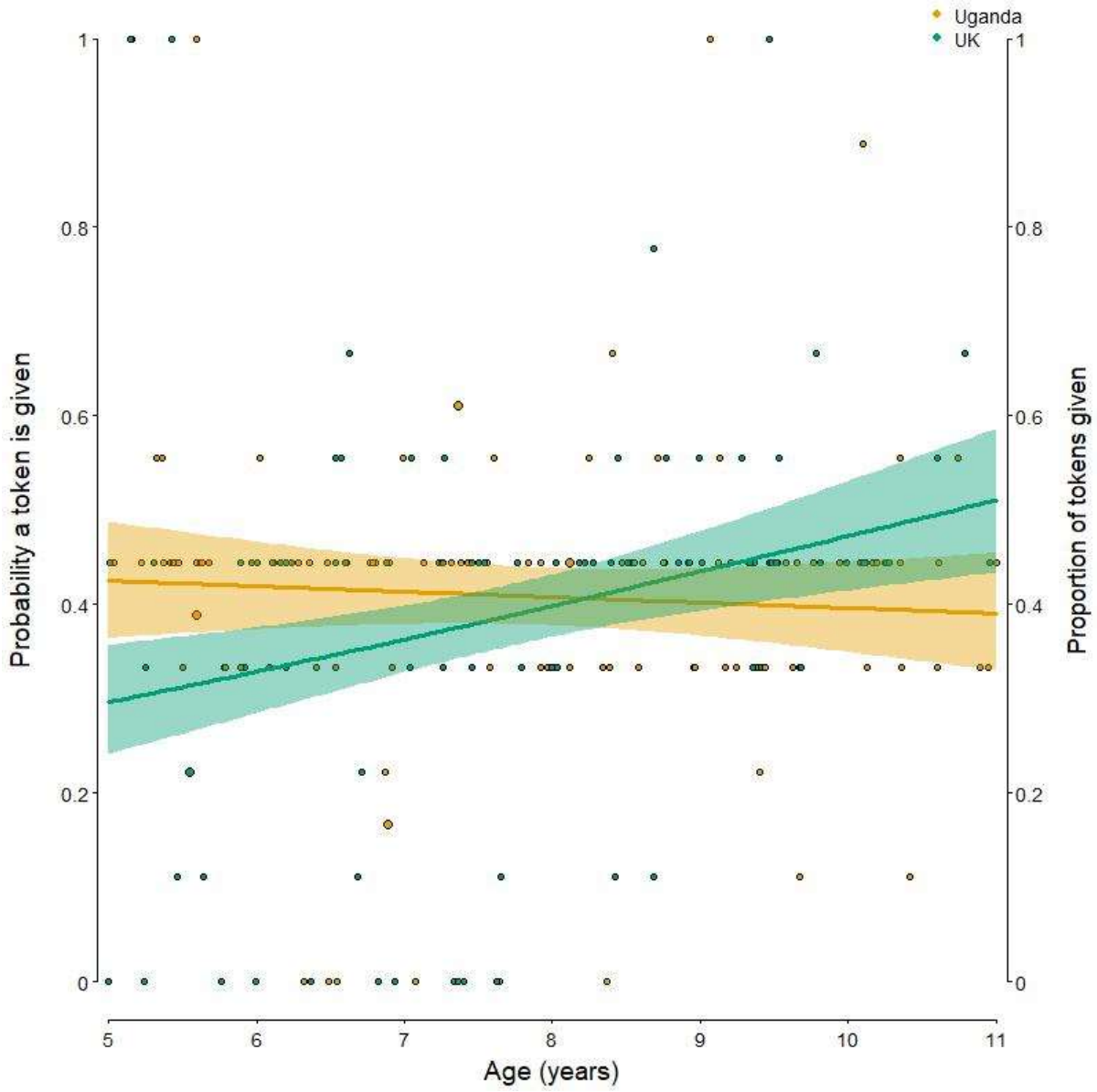
(1) dummy coded with the Ugandan sample being the reference level

(2) dummy coded with girls being the reference level

(3) measured in days; z-transformed

Figure 2.7

Probability that a token is given for per level of Cultural Group (UK/Uganda) across 5-11 years in the Training DG. Points depict the raw data (proportion of tokens given), whereby point area is proportionate to the number of children with the same probability a token is given (range:1 to 9). Lines with coloured areas depict the fitted model and its bootstrapped 95% confidence limits.



2.4.3.4 Research Question 5

Model 5: Is sensitivity to normative expectations common in 5- to 11-year-olds, and does this vary with cultural group and age, controlling for gender?

Overall, Model 5 explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi 2 = 39.343$, df = 15, $p = .001$). In the process of generating Model 5 (i.e. dropping non-significant interaction terms from the initial model structure), the key interaction between 'observe' and 'approval' was not significant (estimate = 0.02, SE = 0.11, $p = .864$) so it was not included in Model 5 (see Appendix 2.2.2.2). Indeed, no interaction terms containing both 'observe' and 'approval' were significant. There was a significant interaction between 'observe' and 'age' (Table 2.17) suggesting that children become less sensitive to observation as they get older (Figure 2.8). There was also a main effect of 'approval' (Table 2.17) which suggests that all children were more likely to give a token when the experimenter approved of generosity (Figure 2.9), and a main effect of 'cultural group' (Table 2.17), with Ugandan children being more generous overall than children in the UK (Figure 2.10).

Table 2.17

Results of final Model 5 exploring 5- to 11-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.75	0.22	0.30	1.21	3.46	.001
Observe[(1)]	0.23	0.20	-0.15	0.61	1.13	.259
Approval[(2)]	0.23	0.08	0.07	0.39	2.87	.004
Cultural Group[(3)]	-1.04	0.19	-1.38	-0.67	-5.50	<.001
Age[(4)]	0.36	0.09	0.20	0.53	4.24	<.001
Gender[(5)]	-0.15	0.17	-0.47	0.17	-0.90	.368
Observe*Age	-0.12	0.05	-0.22	-0.01	-2.16	.031

[(1)] dummy coded with the Unobserved condition being the reference level

[(2)] dummy coded with the Indifference condition being the reference level

[(3)] dummy coded with the Ugandan sample being the reference level

[(4)] measured in days; z-transformed

[(5)] dummy coded with girls being the reference level

Figure 2.8

Probability that a token is given per level of Observation (Observed/Unobserved) across ages. Points depict the raw data (proportion of tokens given), whereby point area is proportionate to the number of children with the same probability a token is given (range: 1 to 9). Lines with greyed-out areas depict the fitted model and its bootstrapped 95% confidence limits.

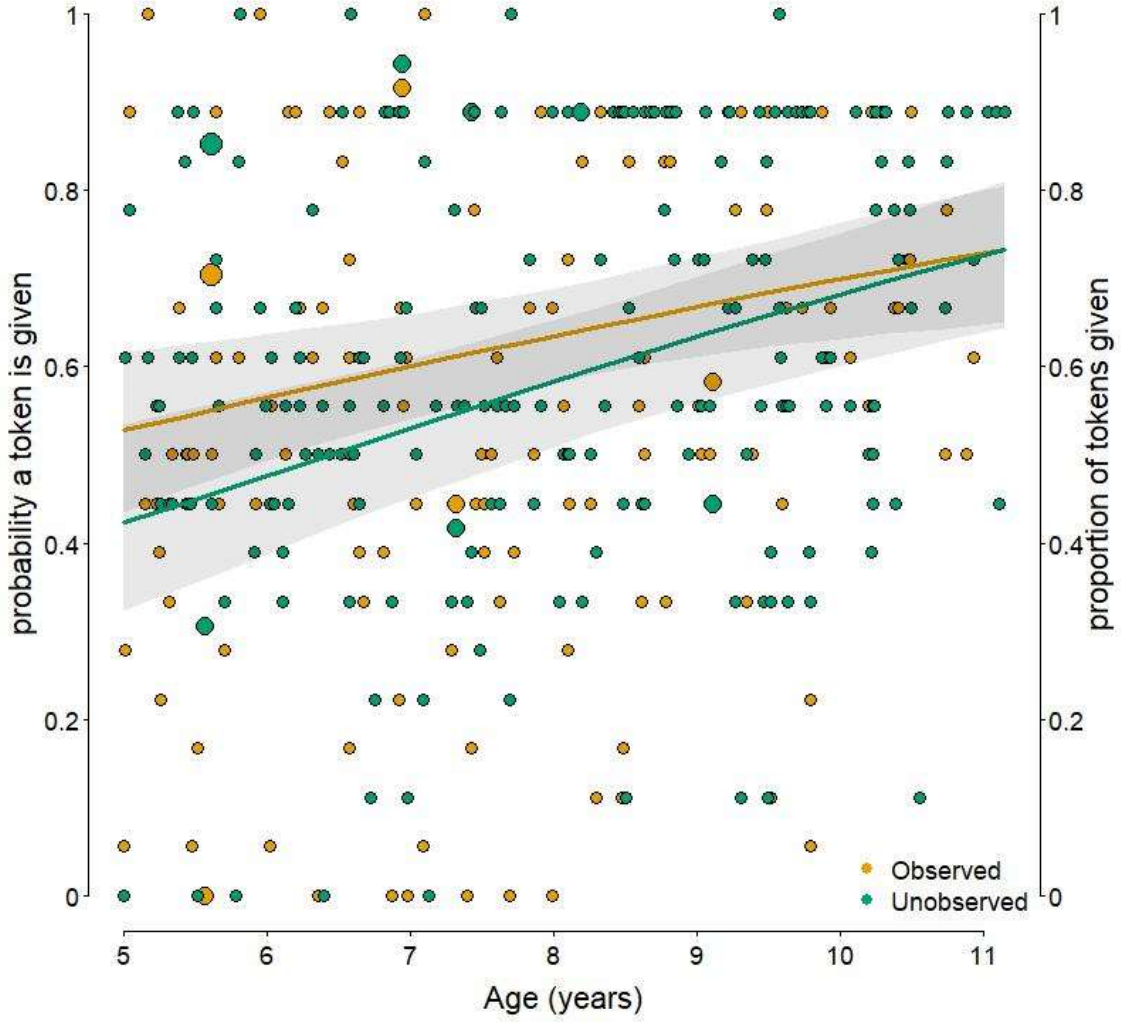


Figure 2.9

Probability that a token is given per level of Approval (Approval/Indifference). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children who gave the same proportion of tokens (range: 5 to 130). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.

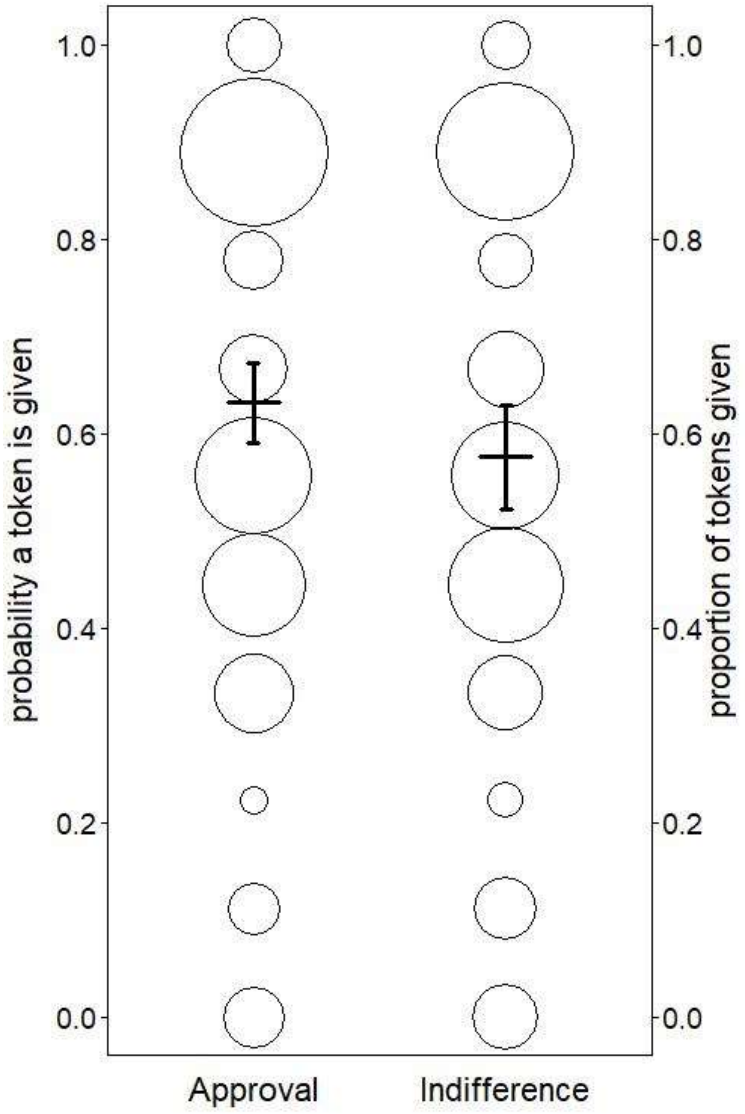
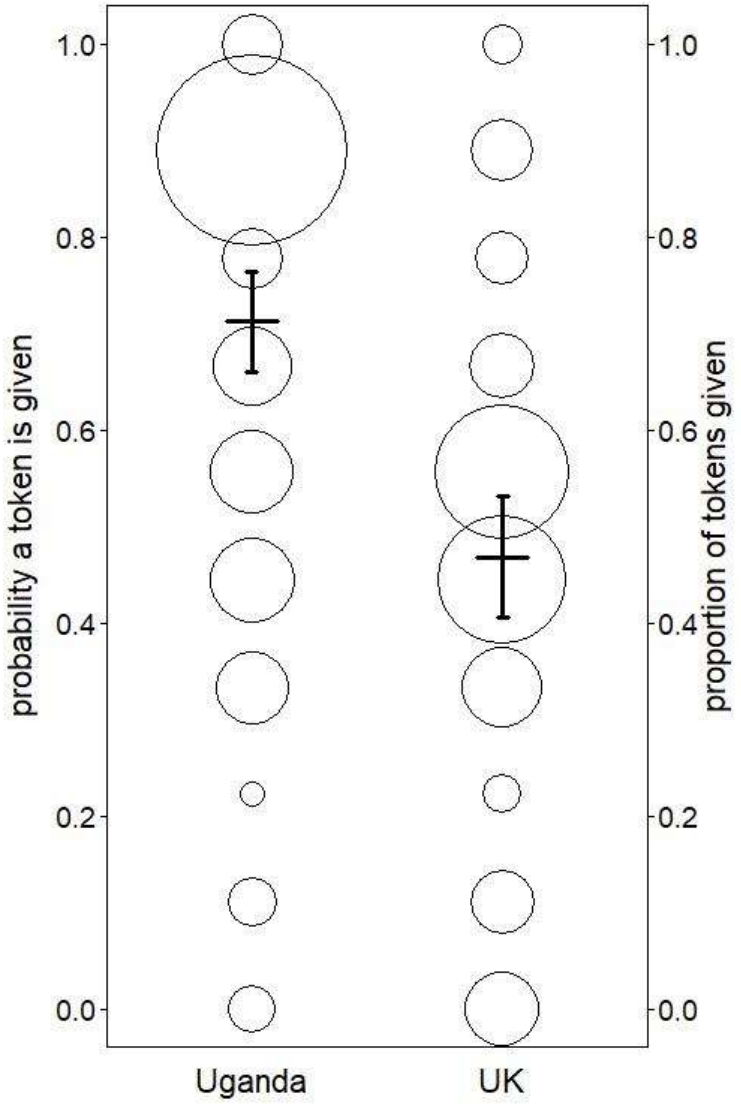


Figure 2.10

Probability that a token is given per level of Cultural Group (Uganda/UK). Bubbles depict the raw data (proportion of tokens given), whereby bubble area is proportionate to the number of children with the same probability a token is given (range: 4 to 220). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



2.4.3.5 Research Question 6

Model 6: Does sensitivity to normative expectations between ages 5 and 11 years interact with gender?

Overall, Model 6 explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi 2 = 63.80$, df = 31, $p < .001$). In the process of generating Model 6 (i.e. dropping non-significant interaction terms from the initial model structure), the key interaction between 'observe' and 'approval' was not significant (estimate = 0.02, SE = 0.11, $p = .874$) so was not included in Model 6 (see Appendix 2.2.2.3).There was a significant 4-way interaction between 'observe', 'cultural group', 'age' and 'gender' (Table 2.18), which upon visualisation (seen in Figure 2.11), suggests that the effect of observation on the probability of giving a token decreases with age, and although this might be more pronounced across different genders within cultural groups (e.g. UK boys and Ugandan girls), the overall pattern remains consistent. There was also a main effect of 'approval' (Table 2.18) which, as before, suggests that children were more likely to give a token when the experimenter approved of generosity, which does not appear to interact with observation, age or cultural group.

We also re-ran Model 6 with a reduced data set comprising of only the children who passed the Generous Prime comprehension questions on the first attempt in both Approval and Indifference conditions (as explained in the 'Coded Variables' and 'Statistical Analysis' sections). The pattern of results did not change for any of the significant interactions reported for Model 6 (see Appendix 2.2.3.2), giving us confidence that the main interactions of interest were not being by participants not understanding the Approval statements.

Table 2.18

Results of final Model 6 examining 5- to 11-year-olds children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.95	0.23	0.50	1.42	4.07	<.001
Observe(1)	0.29	0.31	-0.17	0.75	1.25	.210
Approval(2)	0.23	0.10	0.06	0.40	2.43	.015
Cultural Group(3)	-1.40	0.25	-1.89	-0.91	-5.62	<.001
Age(4)	0.24	0.16	-0.07	0.54	1.54	.124
Gender(5)	-0.29	0.24	-0.75	0.18	1.21	.225
Observe*Cultural Group	-0.03	0.18	-0.39	0.35	-0.15	.878
Observe*Age	-0.27	0.11	-0.49	-0.06	-2.38	.017
Cultural Group*Age	0.19	0.22	-0.22	0.61	0.86	.391
Observe*Gender	-0.37	0.16	-0.65	-0.04	-2.34	.019
Cultural Group*Gender	0.33	0.33	-0.32	0.96	0.98	.328
Age*Gender	0.01	0.22	-0.40	0.43	0.05	.960
Observe*Cultural Group*Age	0.19	0.15	-0.12	0.47	1.25	.212
Observe*Cultural Group*Gender	0.50	0.22	0.07	0.93	2.28	.023
Observe*Age*Gender	0.33	0.15	0.04	0.64	2.11	.035
Cultural Group*Age*Gender	0.11	0.32	-0.55	0.70	0.33	.739

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Observe*Cultural Group*Age*Gender	-0.50	0.22	-0.95	-0.11	-2.29	.022

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

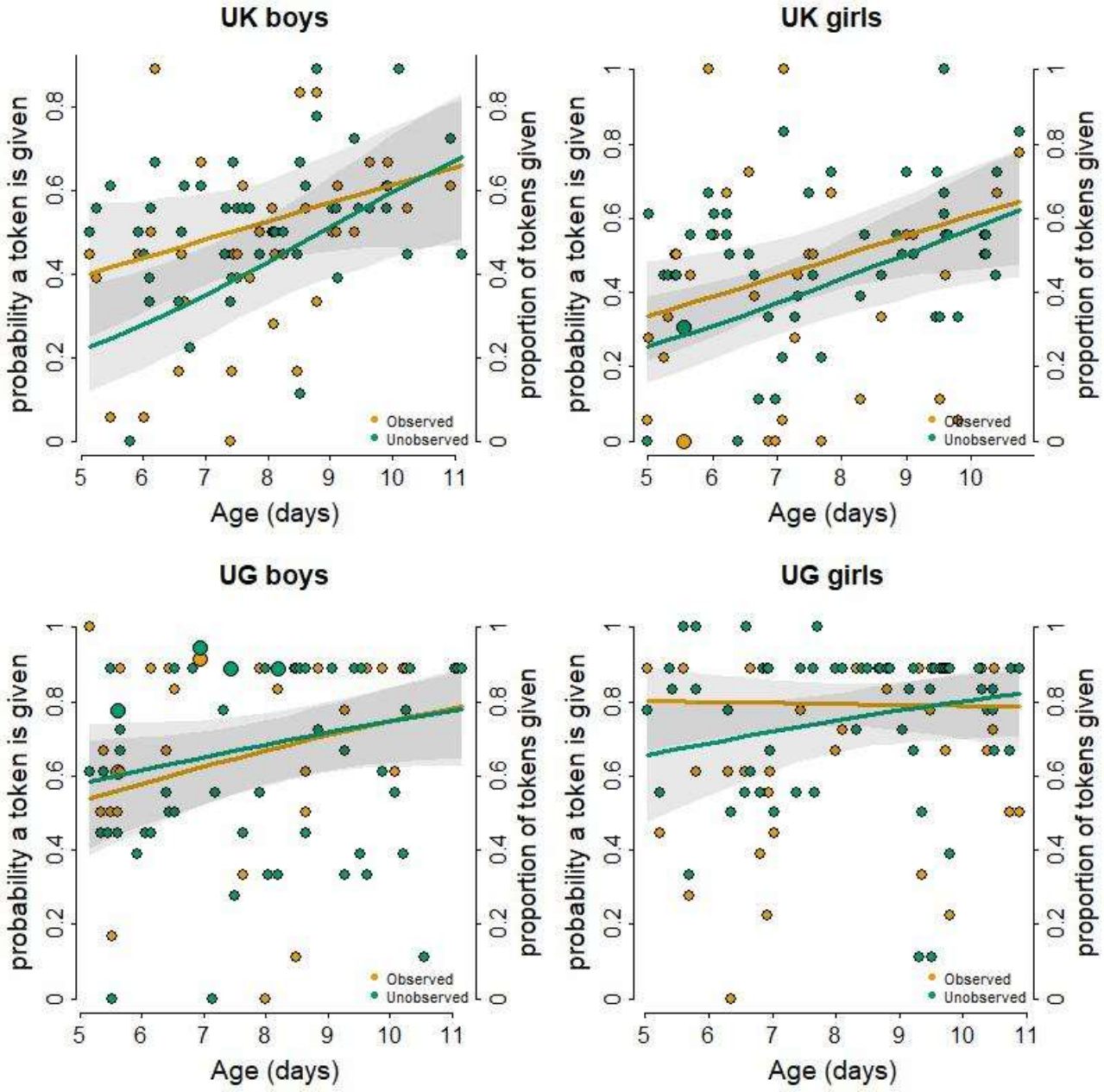
(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

Figure 2.11

Probability that a token is given for levels of Observation (Observed/Unobserved) across the ages of 5- to 11-years, separately for A) UK boys B) UK girls, C) Ugandan boys and D) Ugandan girls. Points depict the raw data (proportion of tokens given), whereby point area is proportionate to the number of children with the same probability a token is given (range: 1 to 9). Lines with greyed-out areas depict the fitted model and its bootstrapped 95% confidence limits



2.4.4 Discussion

Study 2 provides the first direct experimental study of how children's sensitivity to normative expectations develops over the course of early to middle childhood, by investigating children's sensitivity to others' approval when they are both observed and not observed. We found that children did show the predicted sensitivity to the experimenter's approval of generosity (i.e. the Approval Prime), though their sensitivity to approval wasn't reliably greater when they were observed. This indicates that, despite our predictions, in this task children did not become more sensitive to normative expectations as they aged in the way that we expected. One possibility is that this has to do with the fact that the effect of observation on generosity (i.e. children being more generous when being observed) diminished over childhood. This appeared to be true in both the UK and Uganda, though the effect seemed to affect boys and girls differently in the two samples: UK boys and Ugandan girls were most sensitive to observation at younger ages, but this again lessened over time, and as children approached middle childhood (between the ages of 7 and 8), there was no longer an effect of observation on generosity. The reduction in sensitivity to observation is likely important for explaining why we failed to find the expected interaction between observation and approval which would have indicated children were sensitive to normative expectations. Although there were broad similarities in how British and Ugandan children responded to approval and observation in this task, there are signs that they responded differently to the normative information presented about generosity: Ugandan children tended to be much more generous in the Test Trials than were British children. In the Familiarity Trial Training DG the Ugandan children weren't more generous than the British children (though British children did become somewhat more generous with age). However, after observing the Generous Prime (i.e. giving 8 and keeping 1) a large fraction of the Ugandan children gave hyper-generous offers while the British children did not (most giving about 50%).

This study was motivated by the predication that children's generosity would increase with age at the intersection of approval and observation (i.e. normative expectations). Prior research suggests that from 6 years of age, children's behaviour begins to mimic prosocial choices of adults within their society (House et al., 2013; House & Tomasello, 2018; House et al., 2020a), implying that children are being increasingly influenced by normative information from group members on the 'right' amount to share. This suggests that children's psychology for norms is still developing well into middle childhood, but prior research doesn't give much insight into what aspect or aspects of children's understanding of normativity is changing during this period. As argued throughout, we predicted that if children are developing an adult-like understanding of social norms through middle childhood then their sensitivity to others' approval ought to be strongest when their own sharing is being witnessed (and so, any deviations from what was approved of would be noticed and evaluated). Therefore, we would expect that from the ages of 6 that children are becoming even more sensitive to

what the experimenter was saying and share even more generously, especially when that experimenter was watching them. Based on the lack of a significant interaction between approval and observation, it is possible that children's understanding of social norms still differs substantially from that of adults through the period between early and middle childhood. However, it is also important to consider how limitations of the methodology ought to inform our conclusions. For example, due to situational constraints (e.g. testing in bare rooms in Uganda) it was not possible to hide the cameras during testing. This may have hindered children's perception of not being observed during the Unobserved trials, especially as children become more familiar with technology as they age. Also, the Approval Prime was very explicit in its expression of approval, which was deemed appropriate for 4 year-old children in Study 1 but might have been too strong to elicit individual variation in the older children in Study 2. These design choices may have influenced our ability to detect sensitivity to normative expectations in the older sample, although these will be further discussed in the Chapter Discussion.

Finally, it is important to consider how power might have limited the analyses. Our design aimed to increase the statistical power of our analyses by ensuring that the measure was sensitive (i.e. each child provided 9 datapoints for each of the 4 conditions, as each token is analysed as a discrete binary choice of whether it was given or not). However, it is of course possible that additional statistical power would allow us to detect a 5-way interaction between approval, observation, age, cultural group and gender. With that said, the statistical analyses were powerful enough to detect 4-way interactions, so it is reasonable to assume that there should have been sufficient statistical power to detect a 2-way interaction between approval and observation (or a 3-way interaction between 'approval' and 'observe' and 'age'), the interaction we were most concerned about, if it was present.

In conclusion, this study suggests that 5- to 11-year-old children from both the UK and Uganda attend to others' approval of generosity and modify their sharing behaviour accordingly. Surprisingly, this sensitivity to approval did not interact with observation, or age, as expected, and as such the results could suggest that children's sensitivity to normative expectations may not increase linearly between early and middle childhood. However, because younger children showed greater sensitivity to observation relative to older children, it's also possible that constraints of testing and the nature of the task may have made it less likely to elicit meaningful behavioural variation among children in middle childhood.

2.5 Chapter Discussion

Study 1 revealed a trend towards 4-year-old children displaying patterns of generosity motivated by normative expectations, though this was driven largely by the 4-year-old boys, as 4-year-old girls showed a consistent level of generosity regardless of others' approval or observation. This broadly demonstrates the importance of investigating the development of children's understanding of social norms by explicitly studying their sensitivity to what others approve of. Prior research has largely sought to study young children's tendency to do what they're "supposed" to do by manipulating whether or not children are being observed, but not what those children believe others approve of. When extending our view to older children aged 5 to11 years, we again saw that children were sensitive to approval, but children's sensitivity to observation diminished with age and we did not observe the same evidence for gender differences. Overall, this translated into less evidence for normative expectations among older children, rather than more evidence, as we would have expected. Though we did not see evidence for large cross-cultural differences in the development of children's normative expectations, we did see interesting differences in how older children from these groups responded to the normative information presented in this study. Though older British and Ugandan children initially showed relatively similar levels of generosity in the baseline sharing task (i.e. in the Familiarity Phase), after seeing an adult model's hyper-generosity, the Ugandan children were themselves predisposed towards hyper-generous sharing, while the British children largely preferred more egalitarian sharing (i.e. equal splits).

Sensitivity to normative expectations did not continue developing as anticipated between the ages of 4 and 11 years. Whilst in Study 1 the 4-year-old boys showed a pattern of behaviour that was consistent with their sharing being motivated by normative expectations (i.e. an interaction between our experimental manipulations of observation and approval), this pattern of behaviour was not present in Study 2 at older ages (i.e. 5- to 11-year-old children). However, it is important to note that across both studies, children generally showed a sensitivity to the experimenter's approval of generosity (with the exception of 4-year-old girls). This indicates that even from early childhood, children are attuned to feedback from adults about what the right thing to do is, similarly to what has been reported in previous studies (House et al., 2013; House & Tomasello, 2018; House et al., 2020a; McAuliffe, Raihani & Dunham, 2017; Messer et al., 2021; Misch & Dunham, 2021; Raihani & McAuliffe, 2014). However, Study 2 revealed that as children's age increases their sharing was influenced less by being observed: although children's generosity generally increased from 5 to 11 years during the experimental manipulations in our task, they distinguished less between the observed and unobserved conditions. Taken together, these findings could suggest that as children develop through middle childhood the necessity of observation to motivate conformity to social norms may decrease. This does not necessarily fit with how we believe social norm following happens in adults in real-life

scenarios (i.e. adults being more prosocial and acting in line with social norms when observed; e.g. Alpizar, Carlsson & Johansson-Stenman, 2008; Andreoni & Petrie, 2004; Christens, Dannenberg & Sachs, 2019; Izuma, Saito & Sadato, 2010; Lacetara & Macis, 2008; Samek & Sheremeta, 2014; Soetevent, 2005; Rege & Telle, 2004). With that said, it's a possibility that normativity, and in particular, responses to observation, do not develop in a linear fashion from childhood to adulthood. Adults' generosity may be motivated partly by conformity to social norms when being observed, but also by the internalisation of social norms which would motivate generosity even in the absence of observation. It may be that the normative generosity of young children is still largely motivated by external factors (such as being observed by others), and that internalisation of social norms increases through middle childhood, but we need not assume that this developmental process is linear and gradual. Once children begin to internalise what is 'right', they could come to do so promiscuously such that their behaviour becomes heavily influenced by internalised beliefs about what is 'right', and rely less on the presence of other people. Indeed, some research suggests that teens are more generous than adults when putting forward offers in the dictator game (Keller et al., 2013). Both adults and young children struggle to predict their sharing behaviour, often overestimating how generous or selfless they will be unlike participants in middle childhood (Epley & Dunning, 2000; Smith, Blake & Harris, 2013), supporting the idea that there may be nonlinear developmental processes at work in children's emerging prosocial normativity. An alternative framing of this same phenomenon, based on Social Domain Theory, would be that children's reduced reliance on being observed when sharing could reflect the moralisation of prosocial norms. As children approach middle childhood, they may increasingly view sharing not as a context-dependent social expectation (i.e. a conventional norm), but as a moral obligation that applies regardless of who is watching (i.e. a moral norm). If this is the case, then the diminished effect of observation on generosity among older children would not indicate weaker role for normativity, but rather a developmental shift from externally motivated conventional norms to internally motivated moral norms). This is an interesting idea that deserves further study, and our results are perhaps broadly consistent with it, but we caution against drawing firm conclusions as this is based largely on our finding that older children were less influenced by the observation manipulation in our task.

Why might we have observed a gender difference in children's sensitivity to normative expectations at age 4 years? Whilst it remains a possibility that boys begin to strategically manage their reputation earlier than girls, there are a number of alternative reasons which may explain why young girls overall did not vary their sharing behaviour according to others' approval or observation. One reason may be due to the task materials: only female models and female experimenters were used during the study. Given that 4-year-old girls in both cultures displayed a consistent level of generosity across the manipulations, it could be possible that they were copying members of their ingroup (i.e. the female

models), and the ingroup bias was simply weaker among boys (as a group) given that the models were all female. Previous research suggests that young children selectively copy the actions of ingroup members more than outgroup members across various realms, such as accents (Kinzler, Corriveau, & Harris, 2011), language (Howard et al., 2014) and more pertinently, gender (Grace, David & Ryan, 2008; Perloff, 1982). Therefore, the young girls could be simply more influenced by the hyper-generous behaviour of the female model for more social and affiliative reasons (Over & Carpenter, 2012; 2013), instead of reasons purely based on normative expectations. Although, prior research has indicated that the effect of observation increases when children are observed by an ingroup member (Engelmann, Over, Hermann & Tomasello, 2013), so it might also be expected that girls would have been likely to display a greater sensitivity to our observation manipulation and show elevated levels of generosity when observed by a same-gender experimenter, but that was not the case – generosity was consistent across all conditions. On balance the ingroup-bias argument does not necessarily provide a clear explanation for the gender differences observed here.

Alternatively, it could be that young girls are attending more to empirical expectations of the situation, rather than the normative expectations. Although beliefs about what others approve of has been reported to influence people in a number of different studies (Jachimowicz et al., 2018; Raihani & McAuliffe, 2014; Smith & Louis, 2008), beliefs about what other people are doing (i.e. empirical expectations) also motivate people's behaviour, and research suggests that when these are pitted against each other, sometimes the empirical expectations win out (Bicchieri & Xiao, 2009). Whilst young boys might have been focusing on their beliefs about what other people *think they should do*, the 4-year-old girls could have instead relied on their beliefs about what other people *do*, especially if they did identify more with the model in the generous prime. In this case, the boys would have attended more to what the experimenter said they approved of (which varied across experimental conditions), but the girls would have attended more to what the model actually shared (which was the same across all experimental conditions).

The difference in normative prosociality observed between young girls and boys could also be a result of differences in the development of underlying mechanisms, including Theory of Mind and inhibitory control. For example, mental-state reasoning has been theorised to be important in children's ability to attend to normative expectations (House, 2018); for children to attend to beliefs about what others approve of, they need to be able to represent those beliefs in their own mind (and track when they differ from the child's own beliefs). However, previous research indicates that 6- to 8-year-old girls perform better than boys on Theory of Mind tasks (Calero et al., 2013), indicating that this is unlikely to explain the greater sensitivity boys showed to normative expectations than girls in our task. With that said, it is important to consider the relationship between concurrent Theory of

Mind abilities and sensitivity to normative expectations in our specific sample of children. Examining individual variation in mentalising abilities and performance on this task will be more informative than previous group-level differences that have been observed; something that will be further explored in Chapter 3.

Beyond mentalising ability, it has been suggested that children's psychology for prosocial norms may involve inhibitory control, as children are faced with inhibiting self-motivated responses, and instead act in a way more aligned with present social norms (House & Tomasello, 2018). Children's inhibitory control has been found to be related to cooperative behaviour (Ciairano, Visu-Petra & Settanni, 2007) and altruism (Thompson, Barresi & Moore, 1997), further supporting the relationship between this executive function and prosociality. Interestingly, it has also been suggested that inhibitory control supports the process of the internalisation of norms (i.e. when social norms are embedded into your own personal values system, and as such become less dependent on other people; Gintis, 2004; Kochanska et al., 1996). Whilst the relationship between gender and inhibitory control performance is still under debate, most research suggests that girls outperform boys in early childhood (Kochanska et al., 1996; Silverman, 2021; although see Macdonald et al., 2014; who argue this may be due to differences in processing speed). Therefore, one possibility is that the girls in our study could have already internalised these prosocial norms of generosity, and as such, behaving consistently generous could be more of a moral norm for them. This may also explain why young girls tended to share more in the baseline measure of generosity in comparison to the boys. Given that girls are more likely to be perceived as being prosocial than boys (Fabes & Eisenberg, 1998), this could be the result of socialisation; girls' behaviour could be more heavily moralised than boys, and they could be more expected to act in line with social norms. Therefore, our relatively weak indifference statement may not have been enough to pull them away from their already-internalised belief about how they ought to share. Our results are therefore more consistent with the idea that inhibitory control plays an important role in the early development of children's sensitivity to normative expectations. Overall, it is clear that future research is needed to understand these gender differences observed in early childhood normativity, and in particular, the supporting psychological mechanisms through which norms influence prosocial behaviour.

It is possible that the lack of gender differences found in the older group in relation to their normativity may be related to this process of internalisation, and the underlying mechanisms which support it. Whilst research suggests that girls may surpass boys in their inhibitory control from the ages of 2 (Kochanska et al., 1996; Silverman et al., 2021), research exploring gender differences in inhibitory control in middle childhood (ages 7 to 12 years) suggest that gender does not correlate with

performance (Sadeghi, Shalani, & Nejati, 2022). Therefore, if inhibitory control does play an underlying role in children's propensity to follow social norms, including through the internalisation of these as previously discussed (Kochanska et al., 1996), it may be possible that girls at 4-years-old internalise norms more than boys, and the balancing out of skills related to executive functioning across the genders by middle childhood might explain the lack of gender differences found in the older children's response to normative expectations when observed. Even though there were some slight gender differences in how children responded to observation over childhood (i.e. UK boys and Ugandan girls being most sensitive to observation earlier in childhood), generally, observation had a diminishing effect on children's sharing behaviour as they aged. Critically though, children in the older age range are still sensitive to the experimenter's approval of generosity. This may be indicative of a decreasing reliance on observation when deciding whether to conform to normative expectations, possibly because children have begun to internalise norms about what is acceptable to share; children are attending to what others say is important to do, but not relying on being observed to do that. However, to fully understand the role that developing executive function, amongst other developing abilities, plays in the emergence of normativity, future research must longitudinally track the co-development of children's social cognition and executive functioning abilities.

In contrast to our predictions, the main effect of approval reported in Study 2 (i.e. 5- to 11-year-old children sharing more when an experimenter approved of generosity) did not significantly interact with age, gender, cultural group or observation. Whilst middle childhood may be a significant point in development where observation is less required for motivation to conform to social norms, we expected the effect of approval to gradually increase with age from early to middle childhood. This was motivated by evidence that (i) adults are motivated to attend to what other people approve of (i.e. normative information) and act accordingly to avoid eliciting disapproval (Bicchieri & Muldoon, 2018; Kallgren, Reno & Cialdini 2000; Raihani & McAuliffe, 2014; Smith & Louis, 2008), (ii) there seems to be a prominent increase in prosocial tendencies across societies as children enter middle-childhood and (iii) this increase reflects group-specific adult levels of prosociality (House et al., 2013). All this evidence points to children attending more to what is approved of in their own society and begin acting accordingly, increasingly so, from middle childhood.

One possible explanation for our failure to detect changes in sensitivity to approval in 5- to 11-year-olds is that we provided children with a strong, unambiguous Approval Prime. Given that the task was originally created for 4-year-old children, the normative information was relatively strong in its presentation, and quite explicit (i.e. saying, "I like it when people play like that"). In everyday life, individuals don't always express what they approve of so transparently, and so identifying this can require making complex assessments of behavioural patterns and implicit attitudes. By making the

Approval Prime very obvious we may have obscured relevant developmental changes in older children's attention to and understanding of others' approval. In middle childhood, children are becoming more adept at understanding and following more subtle, implicit cues -- most likely as a result of improved Theory of Mind abilities (Bosacki, 2021; Smith & LaFreniere, 2009). Whilst we deemed it important to keep the Approval Prime the same for comparability across ages, it could be possible that older children's advanced social skills were simply not necessary given that there wasn't much work needed by the participant child to understand the valence of the approval or indifference statements. If approval was indicated instead by a subtle behavioural signal, such as a nod or smile, this may have revealed more individual variation and in particular elicited a stronger response in the older children and a weaker response in the younger children. Therefore, it is a possibility that the older children may have been at ceiling with their response given the strength of the Approval Prime, coupled with (at least in the UK) a strong tendency towards egalitarian sharing of 50% (McAuliffe, Raihani & Dunham, 2017; Messer et al., 2021). However, it is important to reiterate that children weren't completely at ceiling, demonstrated by the fact that what the experimenter approved of was still important for the children, which suggests that from the ages of 5 to 11 years, children are responding to approval independently from observation. This is a crucial point, given how prior studies have not considered or controlled (either experimentally or statistically) for children's beliefs about approval when seeking to understand the role of observation on children's sharing behaviour, and raises the point that future studies should include this facet into their designs or procedures.

Contrary to what was expected, the effect of observation on sharing behaviour decreased as children aged. Younger children seemed particularly sensitive to the observation manipulation, including the 4-year-old boys in Study 1, and acted more in line with what was predicted; sharing more when observed in comparison to when not observed. In Study 2, a similar effect of observation is also reliable in the younger boys in the UK sample, but there is no indication that children of either gender in either group are sensitive to observation above about 7-8 years of age. Given the multitude of research which suggests that prosociality increases in the presence of observation for both adults (e.g. Alpizar, Carlsson & Johansson-Stenman, 2008; Andreoni & Petrie, 2004; Christens, Dannenberg & Sachs, 2019; Izuma, Saito & Sadato, 2010; Lacetara & Macis, 2008; Samek & Sheremeta, 2014; Soetevent, 2005; Rege & Telle, 2004) and children (Engelmann, Hermann & Tomasello, 2012; Engelmann et al., 2013; Leimgruber et al., 2012), this developmental result is surprising. As discussed above, this could suggest that children's sensitivity to observation might develop in a non-linear way between early childhood and adulthood, but a perhaps more prosaic and plausible explanation is that the observation manipulation was more convincing for younger children than for older children.

Previous research suggests that turning around is an effective way of ensuring children do not feel observed, in comparison to leaving the room and no longer being visible to the child. This has been evidenced by 4- to 11-year-old children over-imitating less when experimenters turn around in comparison to when experimenters were no longer visible (Marsh, Ropar & Hamilton, 2019). When experimenters are no longer present, children may not know whether they are still being observed or not, which can lead them to acting as though they are still being observed. This concept has also been mirrored in studies measuring false-belief in young children: 3-year-old children pass these tasks when the model in question turns around (and so does not witness the location change), but fail when the model leaves and is no longer visible (Rubio-Fernandez & Guerts, 2013). The decision for experimenters to turn around in the unobserved conditions was grounded in this research, as we anticipated it would be most believable for the children.

However, the constraints of testing our 4-year-old participants during home visits, particularly in Uganda, meant that it was not possible to always hide the cameras while children participated in this task. For this reason, we decided to keep cameras consistently visible across the two sites, even for those participants (largely the 5- to 11-year-olds) who were tested in a lab setting. This visibility of the cameras may have been particularly influential on the 5- to 11-year-olds, rather than the 4-years-olds described in Study 1, because as children age, they have more experience with technology such as cameras, and so may have been aware that they were potentially being observed across all trials (instead of just when the experimenter was watching). This explanation might also help to explain why we failed to observe an interaction between approval and observation: older children might be sensitive to the experimenter's approval of generosity, but with increasing age they simply think that they are always being observed (regardless of whether the experimenter is looking). Whilst we cannot be certain this phenomenon influenced the results, future research into how observation influences prosociality, especially with older children or adults, should ensure carefully controlled procedures (e.g. hidden cameras) to avoid any possibility that participants feel observed during unobserved trials.

Building upon previous research which suggests that children begin to show an awareness of reputation by the age of 5 (Engelmann, Hermann and Tomasello, 2012; Engelmann et al., 2013; Fu et al., 2015; Leimgruber et al., 2012), this is the first study to our knowledge that demonstrates evidence of children's sensitivity to normative expectations around age five (at least in boys). At the same time, our results also suggest that this sensitivity may not develop dramatically differently across a WEIRD and non-WEIRD sample. It is perhaps surprising that we did not find larger cultural differences in the development of normative expectations between the samples, given their reported differences in parenting practices, socialisation goals and attitudes towards norm conformity. Both the relatively relational attitudes of Ugandan mothers (Holden et al., 2022) and the previously reported cultural

'tightness' of Uganda in comparison to the UK (Uz, 2015) would seem to suggest it would be particularly beneficial for Ugandan children to attend to normative information from a younger age than children in the UK, and to have a higher level of conformity throughout childhood. This is especially the case when being observed by another group member, because in relational parenting emphasis is placed on obedience and respecting the social hierarchy and cultural norms (Kâğıtçıbaşı, 1996; Keller et al., 2006), so there are likely to be harsher punishments when non-conformist behaviour is witnessed by a member of the group. However, in both Studies 1 and 2, this was not borne out by data. Although 'approval' and 'observation' each independently interacted with cultural group via a three-way interaction with gender, there were no reliable group level differences in children's sensitivity to normative expectations (as defined here as the interaction between Approval and Observation). This may suggest that the development of sensitivity to normative expectations may be based on something other than cultural attitudes towards parenting practices or socialisation goals, at least in these specific samples. This also raises questions about whether cultural tightness-looseness is a strong determinant of when and how children's sensitivity to normative expectations develops, and perhaps suggests this may be more characteristic of adult groups on a larger scale, instead of at the individual level or in childhood (Schwartz, 2014).

Though it is true that our models included many multi-term interactions and that our power to detect statistical relationship between the variables of interest and cultural group would have been increased by a larger sample of participants, we specifically aimed to mitigate this issue with our within-subjects design in which each participant provided substantial data in all four conditions (equivalent to nine binary choices). We were able to detect many (sometimes unexpected) multi-level interactions with our sample, and even with a reduced sample of participants who were able to answer the Generous Prime comprehension questions correctly on the first attempt (see Appendices 2.1.3 and 2.2.3). However, as these studies reflect a novel approach to exploring the development of normative expectations, and include a sample of participants in Uganda which have received very limited research focus, it is a possibility that the effect sizes of cultural differences may have been small and more difficult to detect. Future research should consider a Bayesian approach for more confidence in these null findings.

In conclusion, this chapter presents the first empirical exploration into how normative expectations shape the development of children's prosocial behaviour; specifically, how 4- to 11-year-old British and Ugandan children's sharing behaviour is influenced by others' approval of generosity and observation. Four-year-old boys, but not girls, showed specific sensitivity to normative expectations by sharing most when they were being watched by an experimenter who approved of generosity. Gender differences in cognition or socialisation may underpin the gender differences observed in 4-

year-olds' sensitivity to normative expectations. Our failure to observe the expected pattern of sensitivity to normative expectations in 5- to 11-year-olds may be due to the apparent decrease in children's sensitivity to observation at older ages in both cultural groups. However, we found evidence that children's sensitivity to others' approval was present from as early as 4 years of age (among boys) through 5-11 years of age. As adults do tend to be sensitive to observation, the age-related decrease in children's sensitivity to observation may reflect the non-linear development of children's normativity from early childhood to adulthood, or limitations of task design. However, our results make it clear that future research on the development of social norms and prosocial behaviour must explicitly investigate, or at least account for, children's beliefs about what others approve of. Future research should also turn to exploring how the development of children's understanding of normative expectations is shaped by individual variation in underlying socio-cognitive mechanisms. This will help to move beyond describing the nature of children's understanding of normative expectations in early and middle childhood, and instead provide insight into how the human psychology for social norms is actually constructed.

Chapter 3: Concurrent Predictors of Children's Sensitivity to Normative Expectations

3.1 Abstract

In Chapter 2, we observed variation in how 4-year-old girls and boys from the UK and Uganda shared in response to normative expectations. Whilst girls were consistently generous across different levels of approval and observation, boys showed a tendency to be more generous when observed by an experimenter who approved of generosity. One plausible explanation for these differences could be variation in the underlying mechanisms upon which normativity is built. Here, we investigated whether 4-year-old children's normative sharing is related to their concurrent ability to represent others' mental states (Theory of Mind) or their executive functioning. Theory of Mind was examined using a battery of five items which tapped into different aspects of mental state understanding (Wellman & Liu, 2004). Executive functioning, specifically cognitive flexibility, was assessed with the Dimensional Change Cart Sort (DCCS) task (Zelazo, 2006). Surprisingly, and in contrast to theoretical perspectives, the results suggested that neither (i) Theory of Mind or (ii) cognitive flexibility were related to children's normative sharing. Whilst it remains a possibility that the ecological validity of the tasks, specifically for our Ugandan sample, limited the ability to detect a relationship between these aspects of cognition and normative prosociality, if replicated, these findings would challenge current theoretical models regarding the mechanisms supporting the emergence of children's normativity.

3.2 Introduction

Chapter 2 found that 4-year-old boys showed a specific sensitivity to approval and observation, while girls were overall more generous than boys, regardless of condition. Cultural group did not appear to influence these patterns, despite evidence for cultural differences across our samples in terms of parenting practices, socialisation goals and attitudes towards norms and norm compliance (Holden et al., 2022; Uz, 2015). One possible explanation for the gender differences found in young children's sensitivity to normative expectations may be a result of the development of underlying mechanisms upon which normativity is built. Two of the most discussed mechanisms, which will be addressed in this chapter, are Theory of Mind (Part 1) and executive functioning (Part 2).

3.2.1 Theory of Mind and the Development of Normative Expectations

Theory of Mind (ToM) refers to the ability to represent and coordinate mental states of the self and others, and mental state reasoning has been theorized to be important in children's ability to attend to normative expectations (House, 2018). In order for children to be able to attend to their beliefs about what other people approve of, they first need to be able to represent the beliefs of other people and understand when these beliefs differ from their own perspective. This also allows children to predict how others will react if norms are followed or violated.

Whilst there have not been prior direct empirical tests of the relationship between ToM and normative expectations, there is research which suggests that ToM is involved in prosociality and social competence; concepts which are often tied to normativity. Theoretically, ToM has been defined as a multi-faceted skill that develops with age, and is crucial for navigating social contexts (Rakoczy, 2022). Imuta et al., (2016) conducted a meta-analysis and found a correlation between ToM abilities and prosocial behaviour, which was particularly strong for children over 6 years of age. Furthermore, ToM performance is related to teacher ratings of children's social skill both in early childhood (Liddle & Nettle, 2006) and middle childhood (Devine et al., 2016; Devine & Apperly, 2021), as well as peer-ratings of suitable behaviour in hypothetical social situations (Bosacki & Wilde Astington, 1999). Whilst these studies are not directly exploring the link between ToM and normativity, they suggest an association between the two, as those who are rated as more socially competent must be navigating social situations successfully, which we would imagine relies on attending to normative expectations (i.e. beliefs about what is approved of or expected) and acting accordingly. However, the question remains as to whether ToM is related to prosociality more generally, or whether ToM influences prosociality via children's greater insight into normative expectations and ability to act in line with what is expected of them in given social situations.

Further indirect evidence exploring the link between ToM and normativity has come from economic games. Firstly, ToM relates to decision making in strategic games with adults. Research using fMRI to examine adults' brain activity while participating in the Ultimatum Game (in which one player proposes how to split a sum a money, and the other play can choose to accept or reject the amount) has found that there is elevated activation in a set of brain areas which have been consistently linked to ToM, suggesting that economic decisions within social contexts involve the ability to reason about mental states of the self and others (Rilling et al., 2004). These brain areas were more active during games with a human partner than with a computer partner, suggesting that playing against another person involves greater representations of mental states - a sign that individuals are assessing what the other person believes to be fair or socially expected when making prosocial, or selfish decisions.

Research also suggests that ToM relates to children's decisions during economic games, specifically when engaging in fair behaviour. Children between 5 and 10 years of age with higher ToM abilities are more likely to respond to offers in these games in a way that prioritises fairness, such as rejecting unfair offers (Castelli et al., 2010), rejecting hyper generous offers and accepting more equal distributions (Castelli et al., 2013). In addition to their responses to offers, the amounts children offer themselves during these games seem to be influenced by how advanced their ToM skills are. Preschoolers who passed a false-belief task (considered to be a hallmark test of ToM) proposed higher offers on average than children who failed (Takagishi et al., 2010). Moreover, children's mentalising ability explains greater avoidance of unsuccessful proposals and higher cooperation with a partner (Sally & Hill, 2006). Given that fair behaviour is inherently normative (as what is considered fair is culturally variable, e.g. Henrich et al., 2005), this suggests that children with higher ToM abilities may be better able to represent what is collectively believed of as fair within their society and compare this against the offer they are proposing, or the offer that another member of their society is proposing.

If ToM influences the development of normative expectations in childhood, then factors which influence the development of ToM might also then play an indirect role in the development of normative expectations. Two particularly important factors to consider for the present research programme are (i) gender and (ii) culture.

Although some research has found no gender differences in ToM abilities (Hughes et al., 2011; Ronald et al., 2006), other studies have reported that girls outperform boys in ToM tasks, both in early childhood (i.e. 2 to 6 years; Charman, Ruffman & Clements, 2002; Marcovitch et al., 2015), and continuing into middle childhood (6 to 13 years; Calero et al., 2013; Devine & Hughes, 2012). This gender difference is reported to occur even when measures of verbal ability are controlled for (Baron-Cohen et al., 1999).

There is some evidence that ToM develops somewhat uniformly across cultural groups. For example, false-belief understanding is acquired at a similar age (roughly between 4 and 5 years) across a wide variety of cultural contexts (Avis & Harris, 1991; Callaghan et al., 2005; Kuntoro et al., 2013; Selcuk, Gonultas & Ekerim-Akbulut, 2023). However, there have been some cultural differences reported.

Although studies with adult participants from Western societies and China show similar performance in ToM abilities (Bradford et al., 2018), research suggests that cultural differences are evident during childhood. A meta-analysis conducted by Liu et al. (2008) explored data from samples of children in China and Hong Kong and found that whilst the trajectory of acquiring false-belief understanding paralleled that of Western samples, the samples differed in terms of onset, with cultural groups showing roughly a 2-year difference in acquisition of false-belief. Furthermore, there is reason to believe that the timetable for acquiring false-belief understanding might exceed 2 years, as children from Samoa have been reported to fail false-belief tasks before the age of 8, with many not passing until the age of 13 (Mayer & Träuble, 2013). One line of argument for the differences in onset is based on specific cultures differentially valuing discussions about mental states, relative to other cultures. For instance, according to some scholars, individuals in some cultural groups, like Samoa, believe that the mind is opaque, and cannot be easily represented by other individuals (Duranti, 2008). It has been argued that these cultural groups are reluctant about discussing mental states (Lillard, 1998; Shore, 1982, as cited in Mayer & Riese, 2013). As such, parents in these societies may discuss mental states less with their children, and may be less likely to ask them questions which promote discussion about their own mental states. This limited exposure could have a knock-on effect for mental state understanding (Duranti, 2008; Mayer & Träuble, 2013). Similarly to Samoan children, low scores on ToM tasks have been examined in Cameroonian children (Chasiotis et al., 2006). Both of these cultures display parenting practices which have been reported to prioritise obedience and inhibitory control (Chasiotis et al., 2006; Mageo, 1998; Mayer & Träuble, 2013), and as a result, putting yourself inside the mind of someone else might receive less parental encouragement as there is a parental expectation for children to do what they're told (without needing to reason about whether that is correct). Therefore, less scaffolding for the development of children's reasoning about mental states may lead to lower pass rates in these particular groups of children on tasks measuring ToM. However, the causation of this has been questioned and criticized, as some authors believe that limited discussions about mental state does not necessarily mean that individuals cannot engage in mentalizing (Lillard, 1998; Mayer & Träuble, 2013) – it just might not be culturally appropriate for children to discuss others mental states.

Cultural variation is also evident when exploring sequential components of ToM. Following criticisms that ToM is more than just false-belief understanding (e.g. Astington, 2001; Bloom & German, 2000),

Wellman and Liu (2004) developed a ToM scale, which assesses the progression of ToM abilities. It examines children's developing understanding of diverse desires, diverse beliefs, knowledge access, false-belief and hidden emotions – the common sequence identified in Western or individualistic cultures (Yu & Wellman, 2024). However, the scale's application across cultural groups has suggested there can be variation in the sequence of acquisition. Children from collectivist cultures have been reported to be able to infer other's knowledge from what has been seen (i.e. knowledge access) before they understand people can have varying beliefs about the same thing (Wellman, Fang & Peterson, 2011; Wellman et al., 2006; Yu & Wellman, 2024). More recently, this has been proposed to likely be a result of more emphasis being placed on shared viewpoints and group harmony in collectivist cultures, limiting children's exposure to diverse beliefs (Selcuk, Gonultas & Ekerim-Akbulut, 2023). Overall, this suggests that culture can not only impact false-belief understanding, but the various elements that contribute to ToM as a whole.

Therefore, the theoretical ground for the link between ToM and children's normativity seems strong, given the links between ToM and prosociality, suitable social behaviours and fairness. Furthermore, it seems both gender and culture may be important factors in children's developing mentalising abilities, which leads us to posit whether the link between ToM and sensitivity to normative expectations develops uniformly across different subgroups of children. While domain-specific socio-cognitive processes like ToM might therefore be reasonably predicted to shape the development of children's understanding of normative expectations, there is also good reason to think that more domain-general cognitive capacities like executive functioning might also play an important role.

3.2.2 Executive Functioning and the Development of Normative Expectations

Executive functioning relates to how control (in the form of cognitive processes) is exercised over thought and action, and it is particularly engaged during times in which thoughts and actions oppose our natural tendencies of how to act or think (Doebel, 2020; Zelazo, 2004). In more recent years, executive function is thought to be more of an umbrella term for several related sub-components that work harmoniously to achieve goals, including working memory, inhibition and task switching (Best & Miller, 2010; Doebel, 2020; Miyake et al., 2000).

Executive functioning seems critical to consider when trying to understand how children navigate social norms: there is usually an appropriate way to act in a given social situation (which is agreed upon by members of society), and children must be able to compare their beliefs or actions against what is deemed appropriate by their group (the 'objective' perspective as described by Tomasello, 2019). Children must then inhibit any responses that do not align with the norm, so as to act in accordance with the group rule instead and be able to flexibly switch between using different

standards of appropriateness in different contexts. In theory, this requires many aspects of executive functioning: children need to be able to hold the rule, or acceptable behaviour, in their mind (engaging working memory), inhibit their own self-motivated interests and constantly assess which rule is socially acceptable across everchanging social situations. Whilst executive function continues to develop throughout adolescence (e.g. Davidson et al., 2006; Doebel, 2020; Luciana et al., 2005), rapid developmental changes occur in early childhood (Carlson & Moses, 2001; Doebel, 2020; Zelazo et al., 2003), making it plausible that improvements in the components of EF could facilitate the start of this normative process in children.

Executive function may directly support the emergence of normative behaviour, but it may also support normative behaviour indirectly, by underpinning ToM, which in turn helps children to contrast different mental states to make inferences about what others approve of. Previous research has found executive function to consistently correlate with ToM abilities (Carlson & Moses, 2001; Carlson, Moses & Breton, 2002). Some have suggested this may be a causal relationship, rather than a correlational one, as experimentally depleting executive functioning abilities (or raising executive functioning task demands) results in lower performance on false belief tasks (German & Hehman, 2006; Leslie et al., 2005; Powell & Carey, 2017). Also, executive functioning predicts ToM abilities (both from 3 to 4 years, and 4 to 5 years), but the reverse relationship is not evident (Marcovitch et al., 2015), further supporting the theory that executive functioning skills underpin the emergence of ToM. These studies suggest that elements of cognitive control are necessary for understanding mental states, particularly when those mental states need to be contrasted, as they do when making inferences about what others approve of.

One aspect of executive functioning, cognitive flexibility (i.e. the ability to task-switch according to the environment), has received special attention in recent years for ToM development (e.g. Tomasello, 2019) due to its relevance in attending to and contrasting multiple perspectives. Indeed, research suggests that cognitive flexibility, over other aspects of executive function, supports false-belief understanding. Devine and Hughes (2014) conducted a meta-analysis which concluded that the strongest correlation with false-belief understanding did not come from inhibition, but from performance on the Dimensional Change Cart Sort task (DCCS; a task-switching game; Zelazo, 2006). In other longitudinal studies, performance on the DCCS task had a stronger correlation with false-belief understanding than other measures of executive function, including inhibition (Diaz & Farrar, 2018). Taken together, these studies suggest that being able to flexibly adapt to your environment may play a critical role in understanding others' beliefs. When children are navigating normative expectations, they need to represent their own belief and the beliefs of others, and then accurately judge when these beliefs differ so as to avoid eliciting others' disapproval. As such, any

cognitive capacity which aids children in contrasting varying beliefs may be directly related to attending and conforming to beliefs about what others approve of.

Besides the link to mentalising, cognitive flexibility has also been associated with prosocial and normative behaviours. Flexible cognition is associated with more cooperative behaviours during social interactions, and with fewer non-cooperative behaviours (Ciairano, Bonino, Miceli, 2006). Furthermore, cognitive flexibility specifically has been found to be related to cultural congruence. Bicultural Turkish-German children who identified equally with both cultural identities outperformed bicultural children who preferred one cultural identity over the other on tasks that required cognitive flexibility (but not working memory or inhibition; Spiegler & Leyendecker, 2017). Those who identify with both cultures are expected to switch between cultural mindsets more frequently (Hong et al., 2000), which naturally will encompass varying social norms and normative expectations. One possibility is that bicultural children with higher cognitive flexibility find it easier to navigate the social complexities of two cultures, making it easier to identify with both, though there is as yet no research directly testing whether cognitive flexibility is specifically related to engaging in socially desirable (i.e. normative) behaviours.

Again, if executive function influences the development of normative expectations in childhood, then factors which influence the development of executive function might also play an indirect role in the development of normative expectations. Once more, two particularly important factors to consider here are (i) gender and (ii) culture.

It is consistently reported that girls outperform boys on measures related to executive function in early childhood (Carlson et al., 2004; Carlson & Moses, 2001; Hughes & Ensor, 2007; Kochanska et al., 1996; Marcovitch et al., 2015). However, some research suggests these differences becomes less pronounced as children age (Marcovitch et al., 2015). There is limited research exploring gender effects specifically in cognitive flexibility, and current results do not paint a clear picture. One study found that Russian girls outperform boys (Bayanova et al., 2022), whilst other studies find no difference (Deák & Wiseheart, 2015; Memisevic & Biscevic, 2018). Interestingly, all three study groups are from different geographical locations (i.e. Russia, Bosnia & Herzegovina and the US), so it is possible that culture might play a role in children's developing executive functioning, and this might be interacting with gender.

There is growing evidence to suggest that cognitive flexibility differs between cultural groups, for example it is becoming more commonly reported that East Asian children outperform Western children in measures related to flexible cognition. For example, 3- to 4-year-old Korean children do better than British children in measures of task-switching, despite working memory and false belief

being similar (Oh & Lewis, 2008). Similar results have been reported in 3- to 5-year-old Chinese children, in comparison to children from the US (Sabbagh et al., 2006), and British children (Ding, Miller & Clayton, 2023). There is limited evidence for how this relates to groups that don't clearly fall within this East-West comparison, but current research suggests that children from such groups perform worse on measures of cognitive flexibility. For instance, 3- to 5-year-old US children showed age-related increases in 3DCCS task (a variant of the DCCS task), but Tswana-speaking South African children did not (Legare et al., 2018). To mirror the explanations for cultural differences in ToM, some have suggested that cultural factors related to parenting (i.e. more emphasis on self-control) might influence children's capacity for executive functioning generally, including cognitive flexibility (Schirmbeck, Rao & Maehler, 2020).

Thus, there is good reason to expect that domain-general cognitive capacities like cognitive flexibility might shape the development of children's understanding of normative expectations either directly, or through supporting other domain-specific socio-cognitive processes like ToM. This theoretical rationale is rooted in the empirical record documenting links between cognitive flexibility and (i) the ability to compare beliefs, (ii) prosociality and (iii) cultural competence. Again, given the variable developmental trajectory of executive function with respect to gender and cultural differences, it is not unreasonable to assume that this hypothesised relationship between cognitive flexibility and normativity may develop differently across subgroups of children.

3.2.3 The Current Study

In order to explore the relationship between concurrent cognitive skills and children's emerging sensitivity to normative expectations, a cross-sectional design was employed which captured variation in both children's (i) ToM abilities, (ii) cognitive flexibility and (iii) sensitivity to normative expectations. ToM was assessed with the ToM scale (Wellman & Liu, 2004) due to it being a more holistic measure of ToM (Wellman & Liu, 2004). Since its development, the scale has been used successfully to track ToM development across a range of cultural groups (see Yu & Wellman, 2024 for a recent review), so we selected it over other standard false-belief tasks in the hope that it would (i) capture more individual variation in children's mentalising abilities and (ii) be accessible to children in both of our British and Ugandan samples. Cognitive flexibility was measured using the DCCS task (Zelazo, 2006); a card sorting task in which children first learn one rule, and then need to inhibit that first rule to switch to a second rule. The task is appropriate for children up to age 5 (Zelazo, 2006) and has been used in different cultural contexts, mostly focused on the East or West but also in South Africa (Legare et al., 2018).

In this Chapter, we will operationalise "normative expectations" by specifically focusing on comparing sharing behaviour in the Approval Observe condition of the Normative Expectations Task

described in Chapter 2 (i.e. where we expect a sensitivity to normative expectations to be most evident), as compared to sharing behaviour in the other conditions (where either "approval", "observation", or both were absent). In other words, we will ask whether children share more in the Approval Observe condition relative to the three other conditions, as this pattern is consistent with our definition of normative expectations and reflects the pattern observed in Study 1 (this is explained in more detail in the Methods section). Our analyses also include a focus on gender differences, given that (i) the prior evidence that gender is related to both ToM and executive function, and (ii) in Study 1 boys showed a pattern of responding to the experimenter's approval and observation, whereas girls did not differentiate between any conditions. Our analyses also focus on cultural group, as we also observed in Study 1 that Ugandan children were overall more generous than children in the UK. Although in Study 1 cultural group did not interact with children's sensitivity to approval or observation, the wealth of literature suggesting cultural differences in both ToM and cognitive flexibility clearly motivates exploring whether there were any differences in how these concurrent cognitive skills relate to normativity across the various subgroups of children.

Research Question 1 (Part 1): Do cultural group or gender influence Theory of Mind performance?

This analysis was exploratory, given the limited research conducted with rural Ugandan participants. Whilst exploratory, it may be sensible to assume that Ugandan children will perform worse on the ToM battery than children in the UK. Our particular sample of Ugandan participants has been found to implement a more relational parenting style, in comparison to our participants in the UK who are more autonomous (Holden et al., 2022). This implies that the Ugandan parents in our sample prioritise obedience and social harmony (Holden et al., 2022), mirroring that of parents from Samoa and Cameroon (Chasiotis et al., 2006; Mayer & Träuble, 2013). Samoan and Cameroonian children showed a slower progression of ToM with a later onset in comparison to Western samples. Although girls are reported to outperform boys in measures of ToM in early childhood (Calero et al., 2013; Charman, Ruffman & Clements, 2002; Devine & Hughes, 2012; Marcovitch et al., 2015), it is unknown how this might interact with our specific cultural groups, particularly in Uganda. Despite some of these studies including racially diverse samples (e.g. Marcovitch et al., 2015), the research is still focused on children from Western samples, and as such, it is unknown whether this gender effect pertains to children from non-WEIRD societies. Therefore, whilst we expect girls in the UK to have a higher ToM score than the boys, it is unclear whether this pattern will be observed in our Ugandan participants.

Research Question 2 (Part 1): Does performance on the Theory of Mind task predict children's sensitivity to normative expectations in 4-year-old girls and boys in the UK and Uganda?

Given indirect evidence that ToM may be related to children's ability to attend to others' approval and act accordingly, we expect that children with a higher ToM score (indicating higher mentalizing abilities) will share more in the Approval Observe condition in comparison to the other conditions. Children with greater mentalizing abilities should be able to better represent what the experimenter thinks is appropriate and predict how the experimenter will respond to their behaviour. Therefore, they should share more in the condition when they're being observed by the experimenter who approves of generosity, but less so when (i) the experimenter is not watching and (ii) the experimenter is indifferent to sharing behaviour. Specifically, this will be marked by an interaction between ToM score and condition: children with a higher ToM should show the greatest differentiation in sharing, sharing more in the Approval Observe condition compared to other conditions, but children with lower ToM scores are expected to show more stable sharing across all conditions. Furthermore, although there is evidence that ToM is influenced by gender and culture, it is unclear from the current literature how this will affect the predictive relationship between ToM and sensitivity to normative expectations. Given that this is the first direct empirical test of this relationship, it would be premature to make strong predictions about the influence of gender and cultural group. As such, the influence of cultural group and gender, and how they interact with ToM to support normativity, was exploratory.

Research Question 3 (Part 2): Do cultural group or gender influence DCCS performance?

This analysis was also exploratory, given the limited research conducted with rural Ugandan samples. As the majority of cross-cultural research relates to differences between Eastern and Western samples, with the majority claiming children in the East have greater cognitive flexibility than those in the West (e.g. Ding, Miller & Clayton, 2023; Oh & Lewis, 2008; Sabbagh et al., 2008), there is limited research which relates to flexible cognition across other diverse samples. The only available research with children from an African population (in South Africa) found that those children performed worse on measures similar to the DCCS than their Western counterparts (Legare et al., 2018), and so based on this limited prior research we might predict that our UK group of children will outperform our Ugandan children on this measure. The picture relating to gender differences is more complex, as it seems that whether girls exceed boys in measures related to cognitive flexibility is culturally dependent – some samples show differences (Bayanova et al., 2022), whilst other groups do not (Deák & Wiseheart, 2015; Memisevic & Biscevic, 2018). Therefore, it is unclear how gender will interact with cultural group to affect DCCS performance.

Research Question 4 (Part 2): Does performance on the DCCS task predict children's sensitivity to normative expectations in 4-year-old girls and boys in the UK and Uganda?

Given that cognitive flexibility seems to be both important for (i) understanding other perspectives (as it has consistently been reported to support false-belief understanding) and (ii) the inhibition of certain rules depending upon contextual factors, it seems reasonable to assume that children's performance on the DCCS should predict their sharing behaviour across the various conditions. Children with greater cognitive flexibility should be able to represent the experimenter's preferences for sharing and understand that preference is particularly relevant when that experimenter is observing their sharing behaviour (i.e. the Approval Observe condition), inhibiting any self-interested predispositions to gain the experimenter's approval. Children with greater cognitive flexibility should be able to follow the following logic: *if the experimenter is watching me and approves of generosity, I should share more to gain their approval, but if the experimenter is not watching me or does not care how I share, sharing more will not have any impact on their approval.* Specifically, this means we looked for an interaction between condition and performance on the DCCS: children who pass the DCCS should share more in the Approval Observe condition in comparison to other conditions, whereas we expect sharing to be stable across conditions in children who fail the DCCS. Following the logic presented in Research Question 2, it is unknown how cultural group might affect the relationship between cognitive flexibility and normativity, and whether this will be expressed differently in girls or boys. Therefore, this question is also exploratory; whilst prior research suggests there can be variation in cognitive flexibility across genders and cultural groups, we do not make any strong predictions about how this will influence the predictive relationship of this aspect of executive function and children's emerging sensitivity to normative expectations.

3.3 Part 1: Research Questions 1 and 2

3.3.1 Methods

3.3.1.1 Participants

The sample consisted of participants who had valid data for the Normative Expectations task (described in Chapter 2; Study 1) and ToM task at 4 years. Out of the 129 participants who had valid data for the Normative Expectations task, two UK participants did not have valid ToM task data, and so the final sample consisted of 127 participants (68 girls). More details regarding the sample of participants can be found in Table 3.1. Further information about the participants and recruitment can be found in the 'Study Populations' and 'Participants' sections of Chapter 2, Study 1.

Table 3.1

The N and mean age of the participants with valid data for the Normative Expectations Task and the ToM task in each cultural group.

Task	Uganda							UK						
	N	Girls	Boys	Long[1]	X-Sec[2]	Mean Age[3]	SD[4]	N	Girls	Boys	Long[1]	X-Sec[2]	Mean Age[3]	SD[4]
Theory of Mind	61	32	29	37	24	4.63	0.27	66	36	30	41	25	4.63	0.18

[1] Number of longitudinal participants in the sample

[2] Number of cross-sectional participants in the sample

[3] Age measured in years

[4] Standard deviation in years

3.3.1.2 Procedure and Materials

The ToM task was modelled on the protocol from Wellman & Liu (2004), which initially comprised 7 items that tapped into aspects of mental state understanding. Each item is described in Table 3.2 and were presented in the same order to the children: 1) Diverse Desires, 2) Diverse Beliefs, 3) Knowledge Access, 4) Contents False Belief, 5) Explicit False Belief, 6) Belief Emotion and 7) Real-Apparent Emotion. To make the scale more practical and effective, many studies have reduced the scale to 5 items (advised as a possibility by Wellman & Liu, 2004), as the 7-item scale included items that were overlapping in terms of which aspect of ToM they were measuring (e.g. Contents False Belief and Explicit False Belief), and the 5-item scale was thought to be more appropriately matched for children's developmental trajectory. The 5-item scale typically contains: 1) Diverse Desires, 2) Diverse Beliefs, 3) Knowledge Access, 4) False Belief and 5) Real-Apparent Emotion. Previous cross-cultural research has tended to use the 5-item scale (Yu & Wellman, 2024), but due to limited evidence regarding the sensitivity of this scale for our rural Ugandan participants, we opted to conduct the full 7-item scale during testing, which gave us the option to reduce to the 5-item scale if appropriate.

The materials in Uganda were designed and constructed with our local research assistants to ensure they were culturally appropriate (see Table 3.2 for an overview of materials used in each sample). Small toy figurines (of animals and people, which are henceforth referred to as puppets) were used along with props (both in picture form and physical objects, like containers) to tell the children the stories. Questions were included as prompts (Prompt Questions) during some of the items to encourage children's attention. For example, during the Knowledge Access story, children were shown a small cardboard box and asked what they thought was inside the box (open question). The children were then shown what was inside the box (either a small dog figurine in the UK, or a small goat figurine in Uganda). The box was closed again, and children were asked again what was inside the box (i.e. Prompt Question). If children failed to answer correctly, the presentation of the figurine was repeated (up to 3 times). Prompt Questions were not used as exclusion criteria as their sole purpose was to encourage the children to engage with the content of the stories.

Additionally, there were two key questions included at the end of each item which were used to assess children's ToM ability; all tasks included a Target Question about the puppet's mental state or behaviour, and a Contrast/Reality Question about reality or the child's own mental state (which contrasted the Target Question). Children could either select one of two images depicted on the A4 laminated sheet to answer (described in Table 3.2) or answer verbally. In the example described above of the Knowledge Access story, children were then presented with a puppet alongside the closed cardboard box and were told by the experimenter that the puppet had never seen inside the box.

Children were then asked (i) if the puppet knew what was inside the box (Target Question) and (ii) if the puppet had seen inside the box (Contrast/Reality Question). To pass each story, the child had to answer the Target and Contrast/Reality Questions correctly. For example, to pass Knowledge Access, the child had to say that the puppet did not know what was inside the box (Target Question), and that the puppet had never seen inside the box (Contrast/Reality Question).

The protocol was translated into the three local languages (Swahili, Alur and Lugbara) for testing in Uganda via a process of translation/backtranslation in English to ensure accuracy. See Appendix 3.1.1 for details of the UK and Ugandan protocol for each item.

Table 3.2

Description of the 7 Stories Used to Assess Theory of Mind (Wellman & Liu, 2004) and the Materials Used in the UK and Uganda

Task	What the Task Assesses	UK Materials	Ugandan Materials
Diverse Desires	Whether a child successfully judges that two individuals (the child vs. a puppet) have different desires about the same items of food.	A puppet and a laminated A4 sheet depicting two separate pictures of (1) a biscuit and (2) a carrot.	A puppet and a laminated A4 sheet depicting (1) a banana and (2) a mandazi (local donut)
Diverse Beliefs	Whether a child successfully judges that two individuals (the child vs. a puppet) have different beliefs about where to find an object, when the child does not know which belief is true.	A puppet and a laminated A4 sheet depicting a garage and a bush.	A puppet and a laminated A4 sheet depicting maize and a structure built for goats to live in (a "goathouse").
Knowledge Access	Whether a child successfully judges (yes-no) whether a puppet knows what is inside the box (when the child has seen inside the box but the puppet has not).	A puppet and a small opaque cardboard box with a dog figurine inside.	A puppet and a small opaque cardboard box with a goat figurine inside.
Contents False Belief	Whether a child successfully judges that a puppet has a false belief about the contents of a distinctive (e.g. labelled) container (when child knows what is in the container, but the puppet did not see inside the container).	A puppet and a raisins box or smarties tube with a small pig figurine inside.	A puppet and a tub of baby junior (petroleum jelly used as a body balm) with a small chicken figurine inside.

Task	What the Task Assesses	UK Materials	Ugandan Materials
Explicit False Belief	Whether a child successfully judges that someone will search for an item in the wrong location, given that individual's false belief about where that item is.	A puppet and a laminated A4 sheet depicting a cupboard and a backpack.	A puppet and a laminated A4 sheet depicting a carrier bag and a large cardboard box.
Belief Emotion	Whether a child successfully judges how a person will feel, when they hold a false belief that a distinctive container has their favourite food inside, and after it is revealed to them that the container actually only has stones in it.	A puppet and a cereal box familiar to the child (confirmed by the mother; e.g. Cheerios, Cocopops, Rice Krispies or Shreddies), which had small rocks inside.	A puppet and a plastic dish which food is served in (confirmed to be typical by the Ugandan research assistants), which had small rocks inside.
Real-Apparent Emotion	Whether a child successfully judges that a person can feel one emotion but also display a different emotion (e.g. after a mean joke was made about them).	A laminated A4 sheet depicting cartoon happy, sad and okay faces, and a laminated A4 sheet depicting a cartoon boy from the back.	A laminated A4 sheet depicting cartoon happy, sad and okay faces, and a laminated A4 sheet depicting a cartoon boy from the back.

3.3.1.3 Measure Extraction

Despite initially employing the 7-item scale as outlined by Wellman & Liu (2004), during analysis, we observed that the Contents False Belief and Belief-Emotion tasks exhibited notably low pass rates among Ugandan children (5/61 and 4/61, respectively). Both tasks are sensitive to cultural and contextual factors, as they require children to be able to make assumptions about familiar containers. Although local research assistants collaborated in identifying containers which would be recognisable and familiar to Ugandan children, and the Prompt Questions aimed to scaffold children's understanding of what the contents of the container should be, it could be that the containers used did not elicit a strong enough belief in the children about what the containers *should* have contained, and as such, may not have reliably served their purpose during these items. Given these limitations, we opted to exclude these two items and instead focus on the 5-item scale, which has been extensively validated for cross-cultural contexts (Wellman & Liu, 2004). This includes Diverse Desires, Diverse Beliefs, Knowledge Access, False Belief and Real-Apparent Emotion. As the False Belief task is not specified, we opted to use the Explicit False Belief measure instead of the Contents False Belief.

However, the final item in the scale (Real-Apparent Emotion) also did not work well with the Ugandan children, as the majority failed to identify the emotions expressed by the cartoon figures (i.e. a precondition of the item), so this item was not completed with most Ugandan participants (N = 49/61), and as such, was also excluded from analysis. Therefore, we ultimately decided to assess ToM using the total number of correct items (0-4) out of the remaining 4 items.

3.3.1.4 Trial Validity

Children were included in the analysis if they had valid data from the Normative Expectation task (described in Chapter 2) and valid data for all four stories of the ToM task. One boy in the UK had missing data for Diverse Desires, Diverse Beliefs and Explicit False belief, and so was excluded from further analysis. There were other minor deviations from the protocol that did not result in exclusion of data, such as Prompt Questions being omitted by the experimenter. A summary of children's responses to the Prompt Questions (including any errors by the experimenter) can be viewed in Appendix 3.1.2.

3.3.1.5 Coded Variables

We coded, from the video recordings of the trials, children's responses to the Prompt Questions (see Appendix 3.1.2 for details) and children's responses to the Target Question and Contrast/Reality Question for each item. To pass, children needed to correctly answer both the Target Question and Contrast/Reality Question correctly for each item. The number of passes was then totalled, and children received a score from 0-4.

Interobserver reliability: The lead author was the anchor coder and coded all videos. After being trained, a research assistant independently coded trials from 28 children (22.05% of the total sample before exclusions). Interobserver reliability between the research assistant and the lead author was calculated by judging agreement on children's responses to each item's Target Question and Contrast/Reality Question (using Cohen's Kappa; Cohen, 1960). All Cohen's kappa's were between 0.91 and 1, indicating excellent interobserver reliability (see Appendix 3.1.3).

3.3.1.6 Statistical Approach

In Chapter 2, it was theoretically motivated to operationalise normative expectations by including Approval and Observation as separate variables in the analyses, as this allowed for separate main effects of Approval and Observation to be explored, despite the ultimate focus being on the interaction of these two variables. Initially, this was important, given that this was the first exploration of young children's sensitivity to normative expectations (conceptualised as an interaction between Approval and Observation), and it remained a possibility that only one of these variables could be driving young children's sharing behaviour. However, the results of Chapter 2, Study 1 suggested that there were not independent influences of Approval and Observation, but rather that (at least for 4-year-old boys) the Approval Observed condition was different from other three conditions, in comparison to the girls. This meant that it was suitable to simplify how we operationalised normative expectations for the following chapters by isolating the condition of most interest: the Approval Observe condition. Therefore, a new variable was created by dummy coding the Approval Observe condition as "1", and the Approval Unobserve, Indifference Observe and Indifference Unobserve conditions as "0". This had the benefit of changing an interaction term to a single variable, and subsequently increasing the statistical power of the analyses. We also considered an alternative approach which would simply use the raw number of tokens in the Approval Observe condition as the response variable, however we rejected this because it would fail to account for individual variation in children's generosity across conditions. Our theoretical model predicts that a signature of normative expectations is that children are more generous when they are observed by an adult who approves of them being generous (rather than when there is indifference to generosity, or sharing behaviour is not being observed). This requires a response variable that allows us to estimate the relative difference

between children's generosity in (i) the Approval Observe condition and (ii) the other conditions. The relative difference between the conditions allows us to tease apart whether children are demonstrating sensitivity to normative expectations, or whether they are merely generally generous across all conditions. Besides this reconceptualization, the same statistical approach as outlined in 'Statistical Approach' under Chapter 2, Study 1 was adopted.

3.3.1.7 Statistical Analyses

Research Question 1: Model 1

Do cultural group or gender influence Theory of Mind performance?

This analysis, along with descriptive statistics, were included to explore ToM performance in our subgroups of children, and specifically to examine whether cultural group, gender, or the interaction between them influenced how children performed on the ToM task. Details of model construction can be viewed in Table 3.3 (Model 1). See Appendix 3.1.4.1 for an outline of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 127 ToM scores from 127 participants.

Research Question 2: Model 2

Does performance on the Theory of Mind task predict children's sensitivity to normative expectations in 4-year-old girls and boys in the UK and Uganda?

This question was theoretically motivated given the significant interaction between approval, observation and gender found in the 4-year-old children, and the wealth of literature which suggests ToM development can be impacted by culture. A description of this model can be viewed in Table 3.3 (Model 2). See Appendix 3.1.4.2 for details of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 508 proportions of tokens given from 127 participants over 4 trials.

Research Question 2: Model 3

Does performance on the Theory of Mind task predict children's sensitivity to normative expectations, controlling for gender and cultural group (rather than exploring interactions with these variables)?

This question was included as a check, to ensure that statistical power was not limiting our ability to detect a significant interaction between the Approval Observe condition and ToM performance, in a complex model including interactions with gender and cultural group. Therefore, cultural group and gender were included as fixed control factors (rather than as interaction terms), and were also included in the null model for the full-null comparison. A description of this model can be viewed in Table 3.3

(Model 3). See Appendix 3.1.4.3 for details of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 508 proportions of tokens given from 127 participants over 4 trials.

Table 3.3

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between performance on the Theory of Mind task and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 and 3).

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
1	Binomial GLMM	ToM correct|ToM incorrect	Gender*Cultural Group	NA	Participant ID	NA	0.86	Gender = 1.00; Cultural Group = 1.00
2	Binomial GLMM	Tokens given|kept in test trials	AO[3] condition*ToM *Gender *Cultural Group	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3] + ToM)||Experimenter on Trial[2] (Gender + Cultural Group + AO[3] + ToM)||Order of Conditions[2]	1.34	AO[3] = 1.00; ToM = 1.37; Gender = 1.00; Cultural Group = 1.37
3	Binomial GLMM	Tokens given|kept in test trials	AO[3] condition*ToM	Gender + Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3] + ToM)||Experimenter on Trial[2] (Gender + Cultural Group + AO[3] + ToM)||Order of Conditions[2]	1.36	AO[3] = 1.00; ToM = 1.37; Gender = 1.00; Cultural Group = 1.37

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

[3]AO = Approval Observe condition

3.3.2 Results

3.3.2.1 Descriptive Statistics: Theory of Mind Items

The percentage of children who passed each task, split by gender and cultural group can be viewed in Table 3.4. Despite a similar pass rate for the Diverse Desires and Explicit False Belief items for UK and Ugandan girls and boys, the Ugandan children had a much lower pass rate on Diverse Beliefs and Knowledge Access, with girls performing worse than boys on the Diverse Beliefs item.

Table 3.4

Number and percentage of children who passed each Theory of Mind item, split by cultural group and gender

ToM Item	Uganda		UK	
	Girls	Boys	Girls	Boys
Diverse Desires	26 (81.25%)	27 (93.10%)	36 (100%)	26 (86.67%)
Diverse Beliefs	8 (25.00%)	13 (44.83%)	33 (91.67%)	26 (86.67%)
Knowledge Access	12 (37.50%)	10 (34.48%)	33 (91.67%)	23 (76.67%)
Explicit False Belief	8 (25.00%)	7 (24.14%)	10 (27.78%)	5 (16.67%)

3.3.2.2 Research Question 1

Model 1: Do cultural group or gender influence Theory of Mind performance?

Overall, the full model explained significantly more variation in ToM performance than the null model (with just participant identity as a random effect; likelihood ratio test: $\chi 2 = 44.00$, df = 3, $p <$ .001). There was a significant interaction between 'gender' and 'cultural group' (see Table 3.5), which suggested that whilst boys and girls in Uganda performed similarly on the ToM scale, in the UK, girls outperformed boys (Figure 3.1) and whilst UK boys gained a higher score on average than Ugandan boys, this difference was much larger for girls.

Table 3.5

Results of final Model 1 examining whether gender or cultural group predicted 4-year-old children's ToM scores (estimates and standard errors, together with confidence limits and significance tests).

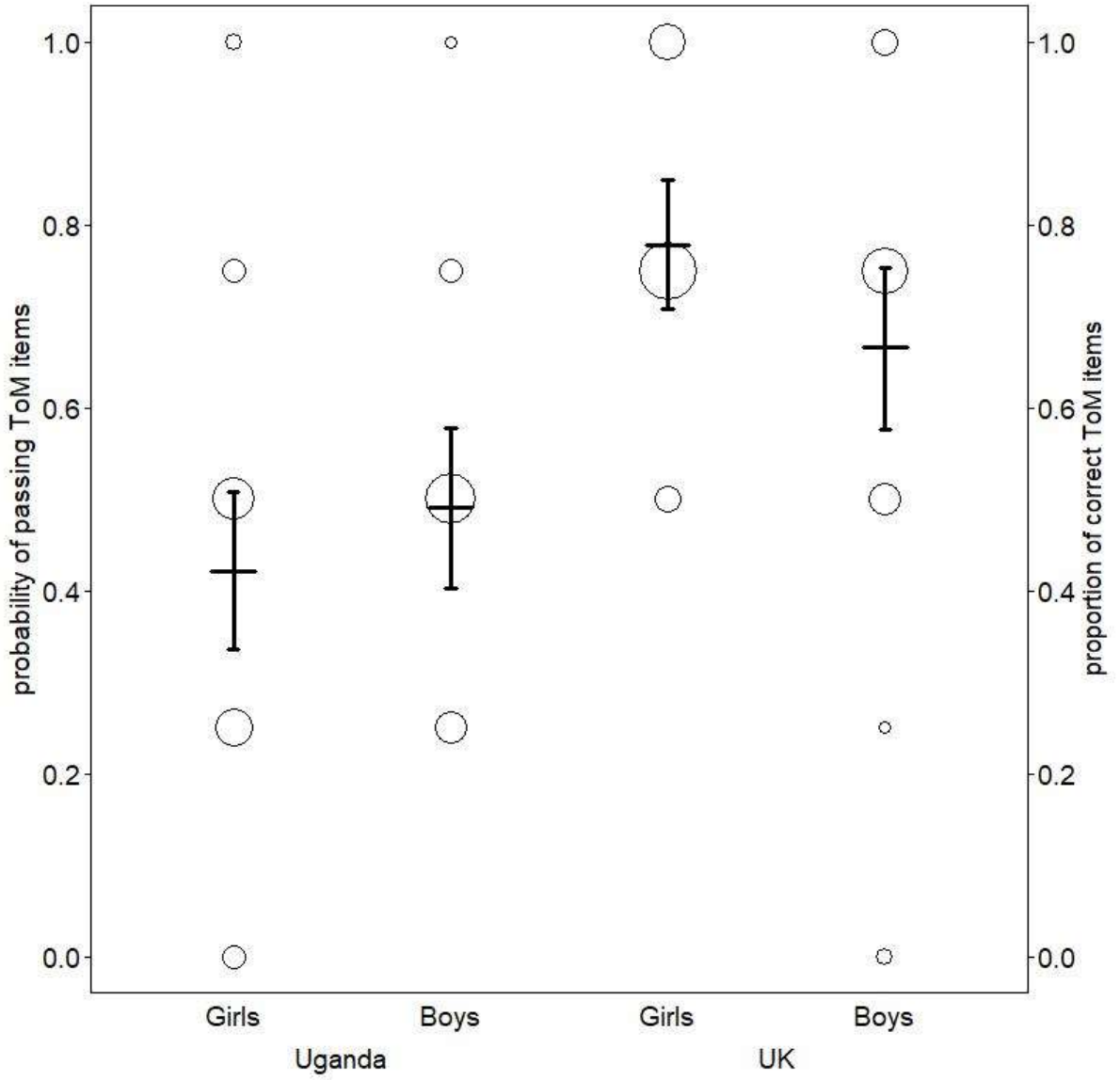
Term	Estimate	SE	Lower CI	Upper CI	z-value	p
Intercept	-0.32	0.18	-0.68	0.03	-1.76	.078
Gender[(1)]	0.28	0.26	-0.24	0.80	1.09	.277
Cultural Group[(2)]	1.57	0.27	1.11	2.15	5.83	<.001
Gender*Cultural Group	-0.84	0.38	-1.62	-0.10	-2.21	.027

[(1)] dummy coded with girls being the reference level

[(2)] dummy coded with the Ugandan sample being the reference level

Figure 3.1

Probability that a ToM item is passed for girls and boys in Uganda and the UK. Bubbles depict the raw data (proportion of tokens given) whereby bubble area is proportionate to the number of children with the same proportion of ToM items correct (range: 1 to 22). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



3.3.2.3 Research Question 2

Model 2: Does performance on the Theory of Mind task predict children's sensitivity to normative expectations in 4-year-old girls and boys in the UK or Uganda?

Overall, the full model explained significantly more variation in the proportion of tokens given to the other child than the null model, which only contained random effects (likelihood ratio test: $\chi 2 = 34.70$, df = 15, $p = .003$). In the process of generating Model 2 (i.e. dropping non-significant interaction terms from the initial model structure), the 4-way interaction between 'ToM Score', 'Approval Observe condition', 'gender' and 'cultural group' was not significant (estimate = 0.50, SE =,0.50, p =.324). There were also no significant 3-way interactions (see Appendix 3.1.4.2). The final model showed a significant interaction between 'Approval Observe condition' and 'gender' (Table 3.6), mirroring the results of Chapter 2, Study 1, which suggests that even in this reduced sample, the boys showed a marked sensitivity to normative expectations, unlike the girls. The main effect of 'cultural group' (Table 3.6), with Ugandan children sharing more overall than children in the UK, also reflected the results of Study 1.

Table 3.6

Results of final Model 2 examining whether 4-year-old children's ToM scores predict sensitivity to normative expectations in girls and boys in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.92	0.53	-0.05	2.04	1.72	.085
AO[(1)]	0.05	0.16	-0.30	0.39	0.34	.737
ToM Score	0.02	0.22	-1.69	-0.10	0.09	.929
Gender[(2)]	-0.85	0.39	-0.43	0.44	-2.18	.029
Cultural Group[(3)]	-1.42	0.46	-2.36	-0.55	-3.08	.002
AO*Gender	0.52	0.18	0.18	0.87	2.87	.004

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with girls being the reference level

[3] dummy coded with the Ugandan sample being the reference level

Model 3: Does performance on the Theory of Mind task predict children's sensitivity to normative expectations, controlling for gender and cultural group?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'gender' and 'cultural group' (likelihood ratio test: $\chi 2 = 4.02$, df = 3, $p = .259$). This reinforces the conclusion that children's mentalising ability as measured by the ToM scale is not related to their ability to navigate normative expectations.

3.4 Part 2: Research Questions 3 and 4

Against theoretical predictions, children's increasing mentalizing ability did not relate to their ability to act in line with others approval when being observed in either girls or boys from the UK or Uganda; a finding that will be further contextualised in the Discussion. Although ToM has a strong framework for its importance in children's ability to attend to normative expectations, there are other aspects of cognition in which children's emerging normativity could be grounded, such as aspects of executive functioning. This chapter will now turn to exploring this possibility, with a special focus on how sensitivity to normative expectations may be related to children's cognitive flexibility. As a reminder, it was theorized that if this relationship between normativity and cognitive flexibility exists, the children who pass the DCCS would likely share more in the Approval Observe condition, in comparison to the other conditions, due to their greater ability to compare various rules, inhibit the contextually inappropriate one and act in accordance with what is expected.

3.4.1 Methods

3.4.1.1 Participants

The sample consisted of participants who had valid sensitivity to normative expectations data (described in Chapter 2; Study 1) and valid DCCS data at 4 years. Out of the 129 participants who had valid sensitivity to normative expectations data, 6 UK participants did not complete the DCCS task, and 16 participants did not have valid DCCS data (more details can be found in the Trial Validity section; Table 3.8). Therefore, the final sample consisted of 107 participants (63 girls). More details regarding the sample of participants can be found in Table 3.7. Further information about the participants and recruitment strategies can be found in Chapter 2, Study 1 in the 'Study Populations' and 'Participants' sections.

Table 3.7

The N and mean age of the participants with valid data for the Normative Expectations Task and the DCCS task in each cultural group.

Task	Uganda							UK						
	N	Girls	Boys	Long[1]	X-Sec[2]	Mean Age[3]	SD[4]	N	Girls	Boys	Long[1]	X-Sec[2]	Mean Age[3]	SD[4]
DCCS	50	29	21	32	18	4.60	0.14	57	34	23	36	21	4.66	0.25

[1]Number of longitudinal participants in the sample

[2]Number of cross-sectional participants in the sample

[3]Age measured in years

[4]Standard deviation in years

3.4.1.2 Materials

In the DCCS task, children were presented with two trays, each with a laminated 8.5cm by 8.5cm target card fixed to the top of the tray. One target card was a picture of a red flower, and the other was a picture of a blue car (see Figure 3.2). During this game, children were asked to sort a series of test cards that shared the dimension of (i) shape or (ii) colour to the target cards, first by sorting according to one dimension ('pre-switch phase') and then switching, and sorting according to another dimension ('post-switch phase'). There were twelve test cards (all 8.5cm by 8.5cm), six of which were pictures of a blue flower, and the other six were pictures of a red car (see Figure 3.3).

Figure 3.2

The pictures of the red flower and blue car on the target cards that were affixed to the sorting trays during the DCCS task.

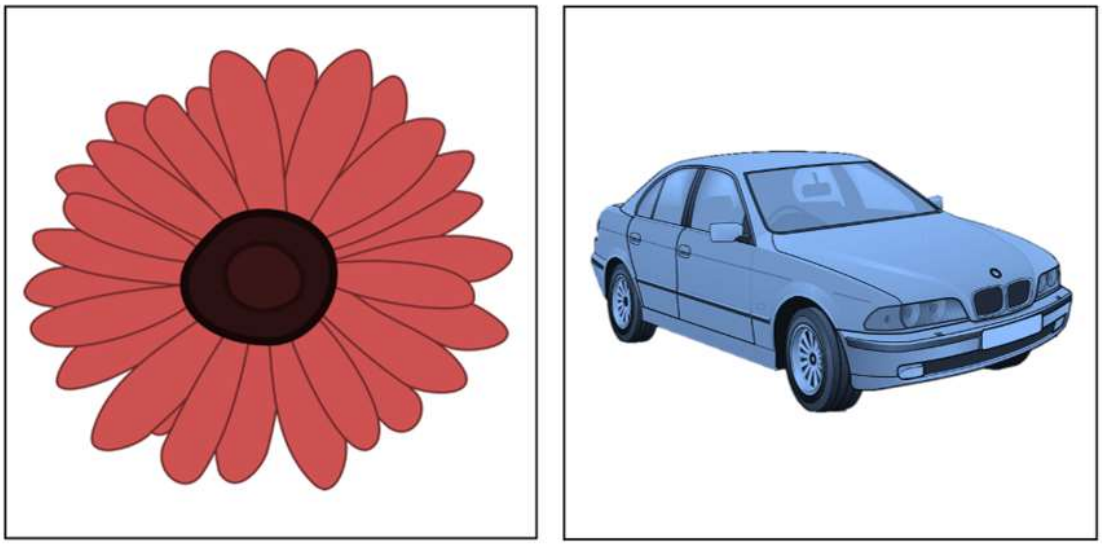
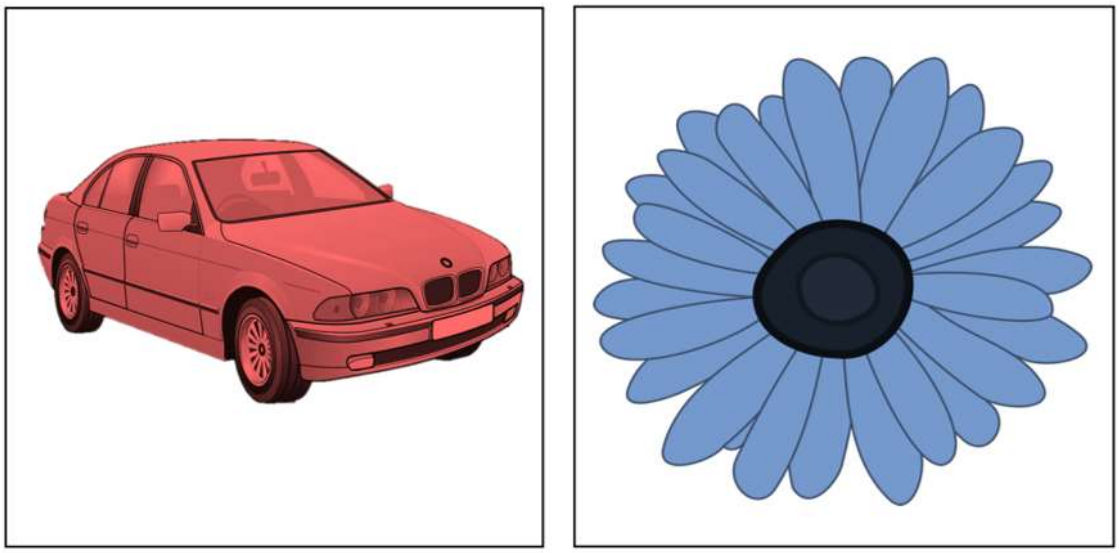


Figure 3.3

The pictures of the red car and blue flower on the test cards that were presented to the children during the DCCS task.



3.4.1.3 Rules

When the child was asked to sort according to the shape dimension, this was referred to as the Shape Game. The children were told by the experimenter that in the 'Shape Game', all the cards with blue flowers should be placed into the tray with the affixed red flower target card, and the cards with red cars should be placed into the tray with the affixed blue car target card. Alternatively, during the 'Colour Game', the children were told that the cards with the blue flowers should be placed into the tray with the affixed blue car target card, and the cards with red cards should be placed into the tray with the affixed red flower target card. The order of the rules was counterbalanced across children, meaning half the children received the colour rule instructions first, and the other half received the shape rule instructions first.

3.4.1.4 Procedure

3.4.1.4.1 Pre-Switch Phase

One rule was explained to the child, as described above, and the experimenter demonstrated where one of the test cards should go (e.g. if this was the colour rule, the experimenter would pick up a red car card whilst saying, "Here is a red car", and then place it face down in the car the with red flower target card). The experimenter then picked up the alternative test card (e.g. a blue flower), stated what was depicted on the card, and ask the child where it should go. Upon the child choosing the correct

tray, the card was placed face down into that tray by the child and the experimenter stated, “Very good. You know how to play the colour game”. If the child chose the incorrect tray, the experimenter explained why it was incorrect (e.g. “No, this one is red, so it has to go here in the colour game”), before asking the child to help them put the card in the correct tray. These two practice cards were then picked up by the experimenter, the rule was reinstated, and the test phase then began. The experimenter presented the child with six of the test cards (i.e. 3 red cars and 3 blue flowers) for the child to sort into trays. No feedback was given to the child in between cards about their performance.

3.4.1.4.2 Post-Switch Phase

Once the final card of the pre-switch phase had been sorted, the post-switch phase began. The child was informed by the experimenter that they were now going to play another game, which was always the alternative game to the one played in the pre-switch phase (e.g. if the colour game was played in the pre-switch phase, the shape game was played in the post-switch phase, and vice versa). The child was then asked to sort the remaining six test cards (3 red cars and 3 blue flowers) using the new rule.

3.4.1.5 Video Coding

3.4.1.5.1 Trial Validity

Children were included in the analysis if they had valid data from the Normative Expectation task (described in Chapter 2) and valid data on the DCCS. For valid data on the DCCS, children had to pass the pre-switch phase of the task by scoring at least 5/6 (following Zelazo, 2006), as the task is related to rule switching, and that could not be explored for any children who failed to keep the first rule in mind before switching to the post-switch rule. As highlighted in the participant’s section above, 9 children were excluded because they scored 4 or less on the pre-switch phase. Another 6 children were excluded due to experimenter error, which either consisted of the experimenter not picking up the practice cards at the start of the pre-switch phase (meaning the child only received 4 cards in that phase; n = 4), the children receiving additional instructions (e.g. being reminded of the rule before each card; n = 1) or the experimenter placing the card in a different tray to the one chosen by the child (n = 1). One further child refused to play the game as intended, so was also excluded (see Table 3.8). Other minor deviations from the protocol were not regarded as important enough to warrant exclusion of the data. For example, if the experimenter omitted repetition of the rule before the test phase, as long as the rule had already been clearly explained, we retained these participants for the analyses (N = 41).

Table 3.8
Exclusion reasons for the participants who were excluded from the final analysis related to the DCCS.

Exclusion Reason	UK		Uganda	
	Girls	Boys	Girls	Boys
The child did not pass the pre-switch phase (scoring 4 or less)	0	1	1	7
Child refused to play game	0	1	0	0
Experimenter error				
Child only received 4 pre-switch trials	0	1	2	1
Child received additional instructions	0	1	0	0
Incorrect placing of the card by the experimenter	0	1	0	0

3.4.1.5.2 Coded Variables

We coded, from the video recordings of the trials, whether the children put the card in the correct trays in both the pre-switch and post-switch phases. The number of cards out of 6 was then totalled in both phases.

Interobserver reliability: After training, two research assistants independently coded trials from 28 children (22.40% of the total sample before exclusions). Interobserver reliability between the two research assistants and the lead author was calculated by judging agreement on whether the cards (N = 336) were sorted correctly (using Cohen's Kappa; Cohen, 1960). All Cohen's kappa's were equal to 1, indicating perfect interobserver reliability. The remaining trials (N = 95) were then split between the lead author (n = 6), and the two research assistants (n = 65 and 24, respectively).

3.4.1.6 Statistical Approach

See the 'Statistical Approach' section in Part 1.

3.4.1.7 Statistical Analyses

Research Question 3: Model 4

Do cultural group or gender influence performance on the DCCS?

This analysis, along with descriptive statistics, were included to explore DCCS performance in our subgroups of children, and specifically to examine whether cultural group, gender, or the interaction between them influenced how children performed on the DCCS task. Details of model construction can be viewed in Table 3.9 (Model 4). See Appendix 3.2.1.1 for an outline of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 107 DCCS scores (pass/fail) from 107 participants.

Research Question 4: Model 5

Does performance on the DCCS task predict children's sensitivity to normative expectations in 4-year-old girls and boys from the UK and Uganda?

This question was theoretically motivated given the significant interaction between approval, observation and gender found in the 4-year-old children, and the literature suggesting that culture and its interaction with gender may influence the development of cognitive flexibility. A description of this model can be viewed in Table 3.9 (Model 5). See Appendix 3.2.1.2 for details of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 428 proportions of tokens given from 107 participants over 4 trials.

Research Question 4: Model 6

Does performance on the DCCS task predict children's sensitivity to normative expectations, controlling for gender and cultural group (rather than exploring interactions with these variables)?

This question was included as a check, to ensure that statistical power was not limiting our ability to detect a significant interaction between the Approval Observe condition and DCCS performance. Therefore, cultural group and gender were included as fixed control factors (rather than as interaction terms), meaning they were also included in the null model during the full-null comparison. A description of this model can be viewed in Table 3.9 (Model 6). See Appendix 3.2.1.3 for details of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 428 proportions of tokens given from 107 participants over 4 trials.

Due to the significantly lower pass rates of Ugandan participants on the DCCS in comparison to the UK participants, we re-ran Models 5 and 6 with a subset of the data comprising only the UK participants. This was to ensure that any lack of a relationship between normative sharing and

performance on the DCCS in the UK children was not being unduly influenced by noise in the Ugandan children's DCCS performance (as explained in the Results section and Appendix 3.2.2).

Table 3.9

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between performance on the DCCS task and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 and 3).

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
1	Binomial GLM	DCCS (Pass/Fail)	Gender*Cultural Group	NA	NA	NA	1.04	Gender = 1.00; Cultural Group = 1.00
2	Binomial GLMM	Tokens given|kept in test trials	AO[3]*DCCS *Gender *Cultural Group	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3] + DCCS)|Experimenter on Trial (Gender + Cultural Group + AO[3] + DCCS)||Order of Conditions[2]	1.16	AO[3] = 1.00; DCCS = 1.23; Gender = 1.00; Cultural Group = 1.23
3	Binomial GLMM	Tokens given|kept in test trials	AO[3]*DCCS	Cultural Group + Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3] + DCCS)|Experimenter on Trial (Gender + Cultural Group + AO[3] + DCCS)||Order of Conditions[2]	1.13	AO[3] = 1.00; DCCS = 1.23; Gender = 1.00; Cultural Group = 1.23

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

[3]AO = Approval Observe condition

3.4.2 Results

3.4.2.1 Descriptive Statistics: DCCS

The percentage of children who passed the DCCS, split by gender and cultural group can be viewed in Table 3.10.

Table 3.10
Number and percentage of children who passed the DCCS, split by cultural group and gender.

Task	Uganda		UK	
	Girls	Boys	Girls	Boys
DCCS	5 (17.24%)	3 (14.29%)	20 (58.82%)	16 (69.57%)

3.4.2.2 Research Question 3

Model 4: Do cultural group or gender influence performance on the DCCS?

Overall, the full model explained significantly more variation in DCCS performance than the null model, which only contained the intercept (likelihood ratio test: $\chi 2 = 26.72$, df = 3, $p < .001$). In the process of generating the final model (i.e. dropping non-significant interaction terms), the 2-way interaction between 'gender' and 'cultural group' was non-significant (estimate = 0.69, SE = 0.98, $p = .479$; see Appendix 3.2.1.1). Therefore, a final model was constructed with only the main effects. The model revealed a significant effect of 'cultural group' (Table 3.11), which suggested the UK children performed significantly better than the Ugandan children in the DCCS task (Figure 3.4).

Table 3.11

Results of final Model 4 examining whether gender or cultural group predicted 4-year-old children's DCCS scores (estimates and standard errors, together with confidence limits and significance tests).

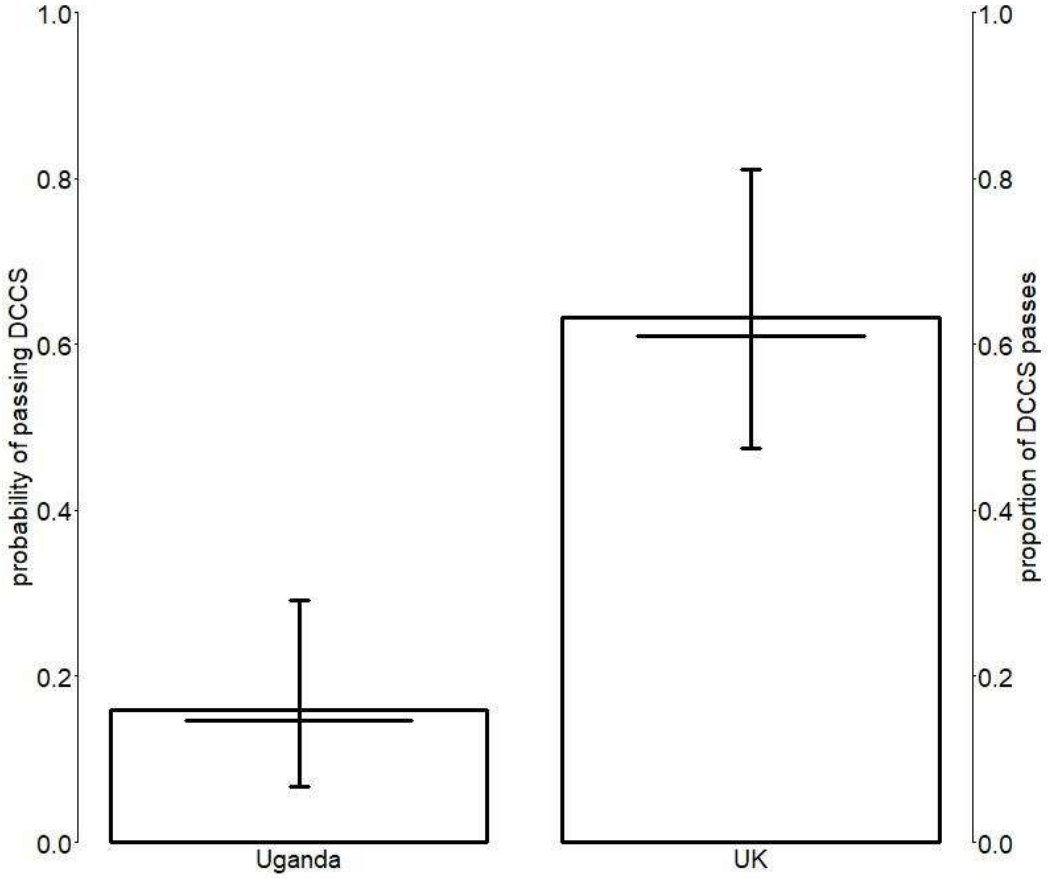
Term	Estimate	SE	Lower CI	Upper CI	z-value	p
Intercept	-1.76	0.44	-2.70	-0.96	-4.01	<.001
Gender[(1)]	0.23	0.46	-0.66	1.14	0.51	.613
Cultural Group[(2)]	2.21	0.48	1.32	3.19	4.65	<.001

[(1)] dummy coded with girls being the reference level

[(2)] dummy coded with the Ugandan sample being the reference level

Figure 3.4

Probability that the DCCS is passed in Uganda and the UK. Bars depict the raw data (proportion of DCCS passes). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



3.4.2.3 Research Question 4

Model 5: Does performance on the DCCS task predict children's sensitivity to normative expectations in 4-year-old girls and boys in the UK and Uganda?

Overall, the full model explained significantly more variation in the proportion of tokens given to the other child than the null model, which only contained the random effects (likelihood ratio test: $\chi 2 = 25.08$, df = 15, $p = .049$). In the process of generating the final model, all non-significant interactions were removed. There was no significant 4-way interaction between 'DCCS', 'Approval Observe condition', 'gender' and 'cultural group' (estimate = -1.33, SD = 1.36, $p = .328$), nor were any of the component 3-way interactions significant (see Appendix 3.2.1.2). There was a significant interaction between 'DCCS' and 'cultural group' (Table 3.12), which suggested that children from the UK and Uganda who passed the DCCS shared different amounts of tokens (Figure 3.5), although, this was not specific to condition. There was also a main effect of 'gender' (Table 3.12), suggesting that girls were more generous than boys overall (akin to what was reported in Study 1). The main effect of 'Approval Observe condition' was only a trend (Table 3.12), likely due to the reduced sample size.

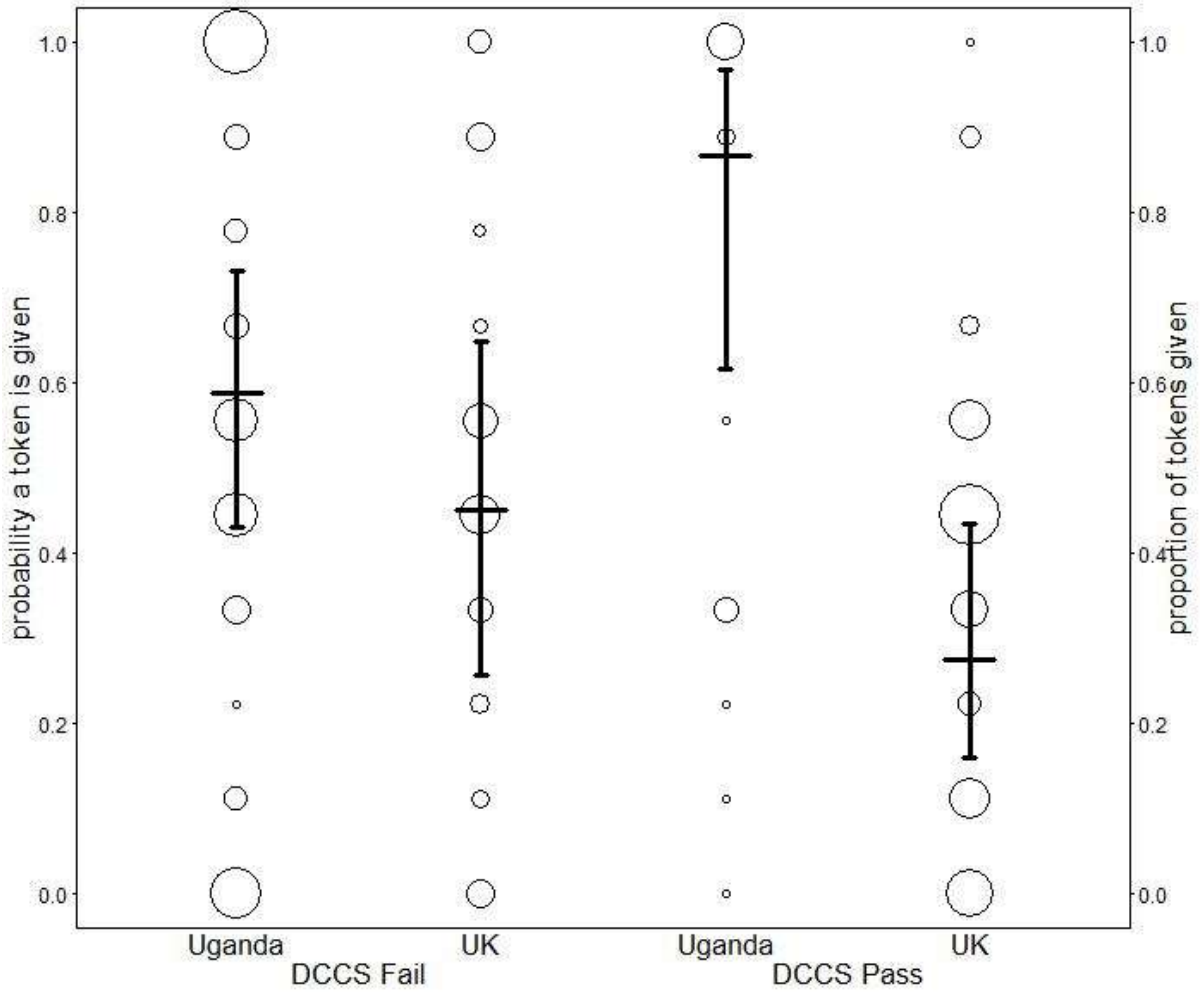
Table 3.12

Results of final Model 5 examining whether 4-year-old children's DCCS scores predict sensitivity to normative expectations in girls and boys in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	p
Intercept	0.64	0.38	-0.09	1.37	1.70	.090
DCCS[(1)]	1.50	0.81	0.04	3.14	1.85	.065
Cultural Group[(2)]	-0.56	0.55	-1.63	0.56	-1.02	.307
Gender[(3)]	-0.83	0.41	-1.67	-0.02	-2.02	.043
AO[(4)]	0.22	0.12	-0.04	0.48	1.79	.073
DCCS*Cultural Group	-2.28	0.99	-4.32	-0.46	-2.31	.021

(1) dummy coded with Fail being the reference level
(2) dummy coded with the Ugandan sample being the reference level
(3) dummy coded with girls being the reference level
(4) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

Figure 3.5
Probability that a token is given for children in Uganda and the UK. Bubbles depict the raw data (proportion of tokens given) whereby bubble area is proportionate to the number of children with the same proportion of tokens given (range: 1 to 49). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



Model 6: Does performance on the DCCS task predict children's sensitivity to normative expectations, controlling for gender and cultural group?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'gender' and 'cultural group' (likelihood ratio test: $\chi 2 = 2.98$, df = 3, $p = .394$). This further suggests that, when grouping children as one sample, cognitive flexibility, as measured by performance on the DCCS task, is not related to their ability to navigate normative expectations.

We also re-ran Models 5 and 6 with a reduced data set comprising only the UK children, to check whether any relationship between normative sharing and performance on the DCCS in UK children was not being obscured by noise relating to Ugandan children's performance on the DCCS (as explained in the 'Statistical Analysis' section). There was no significant relationship between the Approval Observe condition and DCCS performance in either model (Appendix 3.2.2), giving us confidence that the lack of this interaction was not being driven by noise in the data.

3.5 Discussion

This chapter aimed to explore how children's emerging sensitivity to normative expectations is related to their concurrent ToM or cognitive flexibility skills. In contrast to theoretical perspectives which place emphasis on the importance of these skills in the emergence of normative behaviour (e.g. Tomasello, 2019), neither ability was related specifically to children's normative sharing. Therefore, this chapter highlights the significance of testing theoretical ideas, especially across diverse populations, to gain insight into whether they are supported by empirical evidence.

ToM has been theorized to be important for normativity, given that an understanding of what others approve of depends upon an ability to represent their mental state (House, 2018). However, that was not borne out by our data. ToM abilities did not predict children's normative expectations (i.e. it did not interact with our Approval Observe variable in any model), suggesting that children's differential sharing across the conditions is not related to their mentalising ability. Furthermore, ToM did not interact with gender, which may suggest that the gender differences in sensitivity to normative expectations observed in Study 1 are not likely due to gender differences in ToM abilities. It is possible that this null result is a reflection of insufficient statistical power related to a modest sample size. Sample size in this study was constrained by the preexisting size of the longitudinal cohort and logistical constraints on available time to recruit and test additional 4-year-olds, particularly in Uganda where fieldwork time was limited. Future research should aim to replicate this study with a larger sample size. If this replication supports the lack of relationship between ToM and sensitivity to normative expectations in 4-year-olds, we would then be able to interpret these findings with more confidence.

Our findings stand in contrast to some previous studies which have found a relationship between ToM and aspects of prosociality that could be related to normativity. Links have previously been observed between ToM and prosocial behaviour (Imuta et al., 2013), social skills (Bosacki & Astington, 1999; Devine et al., 2016; Devine & Apperly, 2021; Liddle & Nettle, 2006) and fairness (Castelli et al., 2010; 2013; Sally & Hill, 2006; Takagishi et al., 2010) – all concepts which relate to normativity. However, not all research has found a link between children's ToM and fairness-related behaviours. Castelli et al., (2014) found that 8- to 10-year old's decisions during an Ultimatum game were not influenced by their ToM abilities. This study is particularly pertinent to our research question, as the authors established children's normative expectations about fair offers (i.e. they assessed children's beliefs about [i] what they deemed acceptable and [ii] what other children would believe to be an acceptable offer). The results suggested that children were proposing offers in line with normative expectations (i.e. what they thought other children would deem as fair, which in this case was an equal division or decided randomly by a coin toss). However, children struggled to accept offers that

weren't equal (even if decided with the coin toss, which they thought others would deem to be fair). This gap between children's reasoning about fair behaviour and their actual actions has also been reported in simpler paradigms with younger children (e.g. Smith, Blake & Harris, 2013), in which children can say how much they should share, but fail to do so. One interpretation of these studies is that children are aware of how they should act, but have not developed the skills necessary to implement that awareness in their actions. In this case, children were aware, and could offer, socially appropriate offers (similar to the design of our study in which children decided how much to give), but children's offers were not related to their mentalizing ability (similar to our data). Therefore, it could be plausible that ToM could relate to prosociality generally, but not to children's specific understanding of how to interpret and navigate explicitly normative situations. Although, we also did not observe a main effect of ToM on sharing behaviour, so our results do not support the link to prosociality more generally either.

However, it is worth noting that the proportion of tokens children share cannot be treated as a general measure of prosociality. Each condition in the Normative Expectation task employed some kind of priming that was theorized to influence children's sharing, such as showing the children a video of hyper generosity and including expressions of approval or indifference to that modelled generosity. Therefore, whilst we did not observe a direct link between ToM and prosociality more generally, this may be due to how the task was designed. It was outside of the scope of the research questions to test this relationship, but one way of doing so would be to examine the relationship between children's ToM abilities and the number of tokens shared in the Dictator Game, before any priming commenced, as this would act as more of a baseline measure of children's generosity.

Also, there are some notable differences in our study design to that of the aforementioned studies, such as children's age, which may potentially be influencing the role ToM plays in normativity. Prior studies which found a relationship between ToM and prosociality and fairness tested children of an older range (i.e. 5-10 years) in comparison to the 4-year-old children in our sample, which could explain the lack of a significant effect of ToM on children's sharing across the experimental conditions. It may not be until later in development, when ToM is more developed, that it begins to play a more active role in children's decision making, but early signs of normativity rely on other aspects of children's cognition. Signals of normative understanding have been evidenced in children as young as 3 years of age, such as them being able to report what they should do in sharing stations (Smith, Blake & Harris, 2013), and children's willingness to conform to, and uphold conventional norms (Li, Britvan & Tomasello, 2021; Rakoczy, Warneken & Tomasello, 2008). These have been evidenced to occur at an age prior to children's ability to represent the mental states of others; an understanding of diverse beliefs is not typically passed until between the ages of 4 to 5 years (Yu &

Wellman, 2024). Perhaps children's developing normativity shapes their understanding of others' mental states, instead of it being the other way around, and this occurs at later ages than those tested here.

In line with this argument, Andrews (2009) proposes that whilst the focus of the literature emphasizes that ToM is a prerequisite of predicting behaviour and understanding social norms, there is some evidence which suggests that understanding the normative context precedes a fully-fledged ToM. This argument is based on the principle that both humans, and nonhuman primates, can predict behaviour from a number of variables which do not include attributing mental states or beliefs to another person (e.g. facial expressions, context, stereotypes); even young children who have not yet developed a fully-fledged ToM can predict the behaviour of others. Andrews states that ToM has developed as a means to explain, rather than predict, non-normative behaviour, and as such, an understanding of normative expectations must be present before the development of ToM. In relation to our study, this would mean that children who are showing sensitivity to the experimenter's approval may be basing this on other contextual variables, and not specifically ToM. Once the social norm is understood by the child (i.e. to be generous when other people expect you to be, and when they're watching to see how you behave), ToM then develops in response to violations to this learned behavioural rule. However, it is difficult to make any conclusions based on these concurrent measures. In order to fully understand the causation of these socio-cognitive abilities, future research needs to implement longitudinal measures of both ToM and sensitivity to normative expectations over a larger age range in order to statistically explore the direction of this prediction.

Another possibility is that the relationship between ToM and normativity was not borne out by the data due to unexpected performance on the ToM scale, particularly of Ugandan children. Fitting with previous research, in the UK sample, girls performed better than boys, which may be an indication that this task worked as expected for our Western sample (as this reflects what has been reported in previous individualistic cultural samples; Yu & Wellman, 2024). In the meta-analysis by Yu and Wellman (2024), children in individualistic countries typically progress through the ToM scale in linear fashion, with 70% passing Knowledge Access (the third item) when they're around 5 years old. This is similar to what was observed in our UK sample, as high percentages of girls and boys passed the first three items. However, on average, Ugandan girls and boys gained lower scores than UK children, limiting the variation in this measure in this sample and perhaps limiting the opportunity to detect relationships between this variable and others. In Yu and Wellman's (2024) meta-analysis, although the items are not passed linearly, 70% of children from collectivist countries are still reported to pass the first three items by around the age of 5, meaning our Ugandan sample performed below the average of other societies. There are several reasons why the performance of Ugandan

children on this task may have been limited. First, previous research suggests that children whose parents prioritise obedience do less well on these tasks, in comparison to children from Western samples (Chasiotis et al., 2006; Mayer & Träuble, 2013). Previous research suggests that children from cultural settings where obedience is prioritised (and are expected to follow rules without needing to reason about mental states) have less exposure to dialect about mental states, and therefore less scaffolding of mentalising abilities, leading to a delay in the onset and trajectory of ToM abilities (Chasiotis et al., 2006). As there is evidence the longitudinal participants in our Ugandan sample identified more with relational parenting practices (where obedience is highly valued) than our longitudinal UK participants (Holden et al., 2022), this may offer an explanation. Second, it also may be possible that the structure of the task or the materials used were not suitable for our rural Ugandan sample resulting in our task underestimating their abilities. We tried to mitigate this risk through using a scale that has been used with many samples of children across the globe (Yu & Wellman, 2024) and through collaboration with the local research assistants to ensure the items used were culturally appropriate, but we may not have fully succeeded. For instance, the stories rely on attributing desires and beliefs to characters portrayed with puppets. Most UK children have the opportunity to play with puppets and dolls, and are exposed to puppets and dolls having human-like minds in books and television shows, yet the same is less likely to be true for the 4-year-olds in our Ugandan sample. So, although we made efforts to use puppets that were culturally appropriate in each setting, the reliance on puppets to tell the stories may have lacked ecological validity for the Ugandan participants. Future research that gives the children the opportunity to see real people interacting rather than puppets may increase the ecological validity of the task for these participants. Furthermore, using a more implicit measure of ToM, such as preferential-looking and anticipatory-looking tasks may be a better indicator of ToM in these samples, as research suggests that children from non-Western groups perform similarly to children from Western groups on these measures (Barrett et al., 2013).

Much like ToM, we found no interaction between performance on the DCCS and the Approval Observe condition, in explaining the number of tokens children gave away. This suggests that contrary to our predictions, children's cognitive flexibility is not supporting their sensitivity to normative expectations. We hypothesized that this aspect of executive functioning would allow children to logically apply the rules about how much to appropriately share across each condition – specifically, that children should be better able to distinguish between contrasting expectations (i.e. when the experimenter approves of or is indifferent to generosity) and apply certain behavioural rules depending upon the context (i.e. when the experimenter was watching or was not watching them share). However, this was not borne out by the data; children passing or failing the DCCS did not relate to their sharing behaviour across different experimental manipulations, suggesting that cognitive flexibility in this instance, does not support children's emerging normativity.

As with ToM, it is possible that this null result is a reflection of insufficient statistical power related to a modest sample size, and more confidence could be obtained in there being no relationship through replication with a larger sample. It is also important to consider the highly limited variation obtained on the DCCS measure in the Ugandan sample may also have limited our ability to identify relationships between DCCS and other variables. There are several potential reasons for the significantly lower pass rates of the Ugandan children, in comparison to children in the UK. One possibility is that this represents genuine differences in cognitive flexibility between our samples. A growing number of studies suggest that aspects of executive function are influenced by numerous cultural and social factors (Obradović & Willoughby, 2019). These include contextual influences, like how much scaffolding young children receive from parents at home (Fernald et al., 2011; Haft & Hoeft, 2017; McCoy et al., 2015), to nutritional and health-related factors (Black et al., 2017), which is especially pertinent to low- and middle-income countries (Obradović & Willoughby, 2019). Experience in formal education settings has been reported to minimise the gap in cognitive flexibility between high and low performing children (Yeniad et al., 2014). Therefore, given that socio-cultural factors contribute to the shaping of executive functioning in young children, and our samples represent two very diverse contextual settings, it is not unreasonable to assume validity in children's performance on the DCCS task.

With that said, another possibility to consider is that constraints of the task limited our ability to observe a genuine relationship between normativity and cognitive flexibility. Although a concerted effort was made to collaborate with local research assistants to ensure the materials were culturally appropriate, there might have been aspects of the task which were unfamiliar or unsuitable for Ugandan children (who likely have limited experience with these types of card-sorting games in comparison to children in the UK). Recent work which has attempted to translate the DCCS for use with 7–14-year-old Kenyan children has shown some difficulty, particularly with children's capacity to sort cards by shape, as they tend to understand the term 'shape' in geometric terms rather than as a broader categorical label (Duffey et al., 2022; McHenry et al., 2023). Therefore, if the task was not well adapted for our Ugandan sample, the measure of cognitive flexibility may not be reliable for those Ugandan girls and boys, and could have contributed to our non-significant interaction between performance on the DCCS and the Approval Observe condition. Furthermore, despite the pass/fail coding of the task being aligned with previous research and recommendations (e.g. Zelazo, 2006), it may have limited the sensitivity of the measure, and decreased variation. Future work would benefit from teasing apart whether this cultural variation in cognitive flexibility represent genuine differences, or whether certain task limitations prevent cognitive flexibility from being accurately measured in more rural, low-income settings. This could be achieved by implementing a more fine-tuned measure of cognitive flexibility, which captures more variation in children's executive functioning. Thought

should also be given into the transferability of this task with other cultures; perhaps a task in which children need to switch between different rules that does not involve card sorting or categorical labels would be more appropriate for these cultural groups who have less experience with these types of games and instructions. For example, one could expand on a 'Bear-Dragon' task as a more direct measure of cognitive flexibility. During this task, children are asked to imitate the actions of the bear, whilst ignoring the actions of the dragon (engaging inhibition and working memory). If instead, children learnt two competing rules that corresponded to two colours (e.g. *when I see the colour 'red' I copy the bear, and when I see the colour 'blue' I copy the dragon*), children's flexibility could be measured by which simultaneous action they copied (i.e. the bear or the dragon). Replacing the bear and the dragon with real people, instead of puppets, could also be considered to enhance children's relatability to the model.

Despite performance on the DCCS task not relating to children's sensitivity to others approval, there was a relationship between DCCS performance, cultural group and sharing. The findings suggested that, whilst general sharing was similar across cultural groups when children failed the DCCS, there were differences in sharing between our UK and Ugandan samples for children who passed the DCCS – specifically, Ugandan children tended to share more than UK children. This suggests that the way cognitive flexibility relates to prosociality might be shaped by culture. Whilst there has been little investigation into this link, findings from a study by Lan (2023) are in line with this. Parenting styles, cognitive flexibility and prosociality were compared between Chinese immigrant-origin adolescents in Italy and their non-immigrant Chinese and Italian counterparts. The results suggested that higher cognitive flexibility promoted prosociality for participants who received supportive parenting styles within the Chinese-immigrant samples, but not non-immigrant Italians or Chinese. In contrast, higher cognitive flexibility promoted prosociality for the non-immigrant Italian participants who received harsh parenting styles, but not for either Chinese sample. This suggests that cognitive flexibility interplays with cultural factors, like parenting practices, to influence prosociality. This raises the question of which aspects of Ugandan children's socialisation might enhance the relationship between cognitive flexibility and generosity, in comparison to the UK children. However, without specific data to answer this, it is only possible to speculate. Given that our sample of Ugandan participants identify more with relational parenting styles (Holden et al., 2022), it might be that children raised in this environment are having to navigate conflicting rules about how to behave which often prioritise interdependence. With greater flexibility, children may choose to behave in a way that promotes social harmony over more selfish behaviour, which would lead to better cognitive flexibility fostering prosociality within the group. This would also fit with why the UK children with higher cognitive flexibility share less: parenting styles are more autonomous in our UK sample (Holden et al., 2022), so children might be raised in a way that fosters cognitive flexibility tuned to processing more

independent, self-reliant decision making. Of course, this should be interpreted with slight caution given how the sharing paradigm was not designed to be a baseline measure of prosociality, and the priming may have influenced children's sharing behaviour. Future research should explore whether specific parenting and socialisation practices moderate the relationship between cognitive flexibility and prosociality, using a design which focuses on baseline sharing (like in a standard Dictator Game).

Future research should also explore other elements of executive functioning and their link to children's sensitivity to approval. A distinction has been made regarding hot and cool executive function (Zelazo & Carlson, 2012), which may play a role in exploring the mechanisms for children's social norm understanding. Cool executive functioning refers to situations in which cognitive control needs to be employed in relatively neutral, low-stakes environments, in which there are little rewards or punishment (Zelazo & Carlson, 2012) – typically what is examined in game-like contexts like the DCCS. In contrast, hot executive functioning refers to cognitive control employed in affective or motivationally charged, high-stakes environments (Zelazo & Müller, 2002), such as in contexts like gambling or delayed gratification. Children typically perform better during measures of inhibition when the task is cool, for example, when it concerns rewards for someone else in comparison to rewards for the self (Prencipe & Zelazo, 2005), and the development of hot executive functioning has been argued to lag behind development of cool executive functioning (Bunge & Crone, 2009; Zelazo, Qu & Kesek, 2010; Zelazo & Carlson, 2012), suggesting variation within the developmental trajectory of these aspects. Some research suggests that hot executive functioning specifically is related to prosocial and socially competent skills, as rated by teachers (Blair et al., 2004). Therefore, it may be that the Normative Expectation task was particularly 'hot' with regards to executive function, and the measure of cool executive function (i.e. the DCCS) did not translate to the inhibition of self-inclined motivations children faced during the Normative Expectation task. Perhaps including a measure of hot EF would have been more appropriate, like a delayed gratification task or the Iowa Gambling task, as the Normative Expectations task could have been viewed by the children as high-stakes and emotionally significant.

Overall, these results highlight the necessity for testing empirical predictions borne of theoretical models. Aside from relevant concurrent socio-cognitive skills, theoretical models exist which suggest our underlying psychology for social norms is based upon skills that develop in infancy. For example, although Tomasello (2019) argues that the development of ToM and cognitive flexibility precede children's normative self-governance, the crux of Tomasello's framework for how children become normative beings is centred on skills such as joint attention. Joint attention, the ability to recursively attend to a third entity with at least one other person, is believed to develop when infants turn 9 months. Tomasello (2019) proposes that related, but distinct, developmental trajectories, like for ToM

and normative self-governance, are thought to be based on an infant's ability to align and contrast perspectives. In this framework, it may be possible that ToM and cognitive flexibility co-occur with children's developing normativity, but are not direct predictors. Moreover, imitation abilities in infancy have been theorized to provide a basis for children's developing normativity (Carey, 2009; Essler et al., 2023; Tomasello & Carpenter, 2005), as social norms can often be arbitrary, and require faithful copying from individuals. These models, along with empirical tests of them, will be further discussed in Chapter 4.

Notably, the observed interaction between the Approval Observe condition and gender replicates the prior finding from Chapter 2, Study 1 in our sample of participants who had valid ToM data. Although this interaction was not present in models with DCCS, this could be explained by the greatly reduced sample size of participants who had valid DCCS data. This suggests that boys showing a pattern consistent with normative expectations, in comparison to girls, is robust even with a slightly smaller sample size, and when incorporating the new 'Approval Observe condition' variable, instead of including both approval and observation separately into the models. Therefore, this way of capturing sensitivity to normative expectations will be continued in Chapter 4.

In conclusion, this chapter presents the first empirical exploration into how sensitivity to normative expectations is related to 4-year-old children's concurrent socio-cognitive skills: ToM and cognitive flexibility. Despite anticipating that the gender effects observed in Study 1 may be a result of differences in children's underlying social cognition, neither ToM nor cognitive flexibility reliably predicted children's sensitivity to normative expectations in the way we anticipated. This raises questions about theoretical models which propose their importance and highlights the necessity of empirically testing these models, especially with diverse samples. It may be possible that ToM relates to normativity later in childhood, or that norm understanding precedes the ability to represent and explain others non-normative behaviour. Furthermore, it is possible that limitations of the tasks, and their application to our specific samples, limited the ability to detect a link between these aspects of cognition and children's sensitivity to approval when being observed. Future research should assess alternative tasks, to clarify whether these links are genuinely absent in early childhood, or whether they are an artifact of the tasks used. Also, research should explore how children's emerging prosocial normativity might be an extension of joint attention or imitation skills which develop early in ontogeny, to provide further insight into the human psychology for social norms.

Chapter 4: Longitudinal predictors of normative expectations in British and Ugandan children

4.1 Abstract

To be accepted as members of their social group, children must learn to follow the widely accepted conventions and norms of their culture. Theoretical models have made several proposals about how early cognitive mechanisms support the later development of normativity, but there has been relatively little empirical research testing these proposals, nor whether the relationship between early cognitive skills and children's sensitivity to social norms varies across cultural groups or gender. This chapter investigates how cognitive skills in infancy, including both domain specific abilities (joint attention and overimitation) along with more general measures of cognitive development, relate to children's sensitivity to normative expectations within the context of prosocial behaviour. We conducted a longitudinal study with boys and girls from diverse cultural backgrounds (the UK and Uganda) and assessed how individual variation in (i) joint attention at 15 months, (ii) overimitation at 18 months and (iii) general cognition at 11 months (using the Bayley Scale of Cognitive Development), related to individual variation in sensitivity to normative expectations when children were 4 years of age. It was hypothesised that more robust (i) joint attention or (ii) overimitation in infancy would predict greater sensitivity to normative expectations, evidenced by children sharing more in a condition when an experimenter approved of generosity and was observing their behaviour (compared to when either of those were absent). However, there was no evidence of a relationship between joint attention or overimitation and children's normative sharing. In contrast, regardless of gender or cultural group, general cognitive ability was associated with children's sharing behaviour, with lower cognitive scores linked to greater sensitivity to approval and observation. These findings suggest that domain-general cognitive processes may be more important for early normativity than specific abilities like joint attention and overimitation, challenging some leading theoretical perspectives for the mechanisms which support the human propensity for social norms.

4.2 Introduction

To become fully fledged members of their cultural group, children, who come into the world as cultural novices, need to acquire the same beliefs, skills and attitudes as group members, and follow the conventions and norms of their social group (Boyd et al., 2011; Tomasello, 2016). There is considerable variability in social norms, especially prosocial ones, across societies (Chudek & Henrich, 2011), but the age in which children begin to mimic adult-like patterns of behaviour is relatively comparable (House et al., 2013, House et al., 2020a). Therefore, it is likely that from infancy, the strategies and mechanisms employed in cultural learning also share similarities across societies (Legare & Harris, 2016). There are a number of theoretical models which aim to explain the mechanisms for how children's sensitivity to normative expectations emerges, but this chapter will focus on two of the most prominent: joint attention (Part 1) and imitation (Part 2) abilities. These are not mutually exclusive, and further empirical testing is essential to determine whether either, or both, help to explain how domain specific cognitive mechanisms support the development of normative behaviour. To ensure the specificity of any predictive relationships we observe between joint attention and/or overimitation and sensitivity to normative expectations, and to test whether domain-general cognitive processes play a role in children's normative prosociality, we will also explore whether a general measure of cognitive development in infancy predicts children's later normative behaviour (Part 3). This will allow us to detect whether sensitivity to normative expectations is specifically being driven by relationships with joint attention and overimitation, or whether this might reflect more general developmental processes.

4.2.1 Joint Attention and the Development of Normative Expectations

Tomasello (2019) argues that the human capacity for normative behaviour is built upon our unique, evolved abilities for joint attention: the ability of two or more people to coordinate and share attention about an external entity (Tomasello, 1999). In their first year of life, infants become able to engage in joint attention, which requires them to attend to the perspectives of social partners. This forms the basis of Tomasello's theory of Shared Intentionality; infants begin to understand that they are sharing attention about the same thing, but at the same time, are holding individual perspectives about it. This capacity for joint attention is then theorised to mature into cooperative communication and collaborative activities; toddlers not only begin to attend to mental content, but begin to actively align others' perspectives with their own to communicate effectively and achieve goals together, marking the beginning of children understanding that there are ways things *should* be done and recognising the roles and expectations of others and themselves in achieving those goals. This understanding and navigation of perspectives is critical for joint goals, and allows young children to begin to understand and track when certain perspectives are 'right' or 'wrong'. When children reach school age, these

components of social cognition, communication and cultural learning form the basis of cooperative thinking and reasoning – the cooperative thinking and reasoning which underpin many of the cooperative achievements which make humans unique. There is then a shift from their own perspectival view to a more objective one, and they become increasingly able to think in terms which are not associated with any one person's perspective or intentional state. Tomasello (2019) argues that this ability to consider an objective perspective sets the stage for normative self-governance; the principle that there is an objectively reasonable way to think or act in certain situations, and this can be compared with one's own attitudes or behaviour.

This theoretical framework provides a clear prediction that joint attention forms the foundation of later acquired social skills, including sensitivity to normative expectations. However, before we consider the empirical support for this theoretical framework, it is important to acknowledge that joint attention is a multi-faceted phenomenon that is defined and operationalised in very different ways in the literature (Carpenter & Call, 2013; Carpenter & Liebal, 2011; Graham et al., 2021; Sipsova & Carpenter, 2019). Firstly, most research specifically focuses on skills related to joint attention, which largely fall into two categories: responding to joint attention, and initiating joint attention (Mundy & Newell, 2007). Responding to joint attention skills relate to the ability to follow another's gaze direction or gestures to coordinate visual attention on the same external entity as a partner. Initiating joint attention refers to the ability to direct another's attention towards a third entity. Whilst joint attention skills can be used for individualistic or competitive reasons, they also facilitate engagement in joint attention events with a social partner, where attention about an external entity is shared. Joint attention events are, however, operationalised very differently across the literature and can include simple definitions such as two or more individuals attending to the same external entity (e.g. Craig et al., 2004; Najnin & Banerjee, 2018; Povis & Crowley, 2015; Richardson et al., 2007; Yu & Smith, 2013), or both partners using gaze alternation to coordinate attention towards their partner and an external entity (e.g. Gaffan et al., 2010; Nelson et al., 2008). However, these operationalisations of joint attention events have been criticised due to them not capturing the 'jointness' of the attention being shared (Carpenter & Call, 2013; Sipsova & Carpenter, 2019). For instance, gaze alternation between a partner and external entity could occur for selfish or competitive reasons, such as monitoring others attention to and approach towards a high value reward in order to compete with them for it. When considering the predictive power of joint attention for normativity it seems crucial to try and capture the 'jointness' of the interaction, as for infants to develop shared intentionality, there must be processing of both another's perspective but also the shared experience (Tomasello, 2019). To address this, more sophisticated and conservative operational definitions of joint attention events have been put forward, which aim to capture whether two individuals know that they are attending to the same thing (Carpenter & Liebal, 2011; Sipsova & Carpenter, 2019). Usually, this

definition requires both parties to communicate about the entity they are sharing attention about (Carpenter & Liebal, 2011; Sipsova & Carpenter, 2019), which acts as a concrete and explicit signal that they are jointly engaged, and mutually aware, that they are both sharing attention about an external entity. This operational definition is the most rigorous and offers the greatest chance of capturing true joint attention events, and should in theory, relate to children's sensitivity to normative expectations in the most salient way. Communication is also an important aspect of developing normative self-governance (Tomasello, 2019), as it elicits understanding that behaviours or attitudes are viewed differently, and have notions of approval or disapproval associated with them.

Tomasello's (2019) theory for how joint attention leads to later normativity has received some indirect empirical support. From early in infancy, joint attention skills, such as gaze following and coordinating attention between an adult and object, have been reported to facilitate communicative competence (Carpenter et al., 1998), indicating its importance for understanding others' intentions. Additionally, from the age of 18 months, infants will attempt to reengage a partner to reestablish a shared goal, unlike chimpanzees (Warneken, Chen & Tomasello, 2006), providing support for the idea that once joint attention is established, children begin to understand the role each agent plays, and will cooperatively communicate to achieve that goal. Then, as predicted by Tomasello's (2019) theory, as children age, they become more able to hold an objective perspective that they can compare individual perspectives against. For example, until the ages of between 4 and 5 years, children feel conflicted in saying that an object is both right-side-up and upside-down when they and an adult see the object from opposite sides of a table (Flavell et al., 1981). Similarly, 4-year-olds, but not 3-year-olds, understood that an object (viewed through different coloured filters by them and an adult) could appear simultaneously blue to them and green to the adult across the table (Moll et al., 2013). Whilst all these studies provide support that children are progressing through the stages of developing normativity outlined by Tomasello (2019), there is yet to be direct evidence that joint attention is a pivotal building block for the later emergence of children's sensitivity to normative expectations. Strong evidence in support of this would require a longitudinal study examining joint attention in infancy and relating it to children's later normative behaviour.

If joint attention in infancy influences children's emerging normativity, then factors which influence the development of joint attention might also lead to variation in children's sensitivity to others' approval. Specifically, joint attention is involved in attending to and coordinating others' perspectives, which theoretically ought to support children's understanding of what it means when perspectives are labelled 'right' and 'wrong' (namely that these labels can have consequences which arise during social interactions), and help children to develop a sensitivity to what others do and do not approve of (i.e. normative expectations). Therefore, children who engage in joint attention earlier, or more

robustly in infancy, might be expected to develop an earlier or stronger sensitivity to the implications of others' approval and disapproval of their behaviour. As previously established in earlier Chapters, two important factors to consider for the present research programme are (i) gender and (ii) culture, and so it is important to consider how these are related to joint attention as this could be essential for interpreting the relationship between joint attention and normative expectations.

Previous research suggests that there may be gender differences in joint attention abilities in infancy, with girls outperforming boys. In infancy, girls have been reported to engage in longer bouts of joint attention (measured by extended periods of gaze alteration between the mother and object; Raver & Leadbeater, 2014). Girls also show more frequent initiating and responding to joint attention (Mundy et al., 2007; Olafsen et al., 2006) and engage in more bouts of joint attention whilst playing with their mothers, as characterised by reciprocal play and turn-taking, in comparison to boys (Saxon & Reilly, 1999). Recent research also suggests that whilst boy infants more frequently shifted attention between entities (e.g. hands, face and objects), girls experienced longer periods of joint attention in which both they and their parent were simultaneously looking at an object together (Perkovich & Yoshida, 2024). Whilst these results point to girls being more advanced in their joint attention skills and engaging in joint attention events differently to boys during infancy, there is considerable variation in the operationalised definition of joint attention events throughout this research, and it is difficult to make firm judgements about whether some of these operationalisations truly reflect the shared experience of joint attention. Yet, these studies nonetheless suggest that gender may be important to consider when exploring the relationship between joint attention and normativity.

Although much research relating to joint attention falls victim to the persistent Western sampling bias evident in developmental psychology (Bard et al., 2021; Nielson et al., 2017), cross-cultural findings suggest some universality in the development of joint attention. Callaghan et al., (2011) explored joint attention engagement in Canadian infants, and infants from traditional settings such as rural Peru and India, by measuring infant's responses to bids for their attention, followed by gaze alternation between an object and partner. They found that infants performed similarly across cultural groups. Childers, Vaughan & Burquest (2007) explored time spent in joint attention events with a play partner in rural Nigerian infants and American infants. During this study, multiple levels of joint attention were coded during naturalistic observations, with the top tier representing 'coordinated joint attention'. For this level, infants needed to be actively involved with their caregiver, and to gaze alternate between their caregiver and the object of their caregiver's focus. Nigerian infants spent nearly double the amount of time in 'coordinated joint attention'. However, the authors did use the American toys for the Nigerian participants, so this could have heightened coordination through novelty. In a more recent exploration of responding to joint attention skills (as measured by gaze

following), infants from Sweden and Bhutan were also found to show similar trajectories (Astor et al., 2022). However, in this study, the authors also explored the effects of maternal post-partum depression on this trajectory and found that whilst Bhutanese infants' joint attention skills were unrelated to post-partum depression, Swedish infants were negatively affected. The authors suggest that aspects of an interdependent culture in which social networks are prioritised, like in Bhutan, might act as a protective buffer to the impact of maternal depression, unlike the WEIRD culture in Sweden. This might suggest a more complicated picture, in which culture may interplay with other factors or influences to affect the trajectory of joint attention. It is important to note that none of the studies listed above operationalised joint attention events according to the most rigorous definition where the shared aspect of the interaction is manifest through communication, so our picture of how engagement in joint attention events develops across cultural groups is far from complete. However, the limited data available broadly suggests that there is some similarity in the development of joint attention skills across cultural groups.

To summarise, the theoretical roadmap for how joint attention shapes children's later normativity has been clearly expressed, predominantly by Tomasello (2019). Whilst cross-cultural research suggests some uniformity in joint attention skills development, research has highlighted how gender may be an important factor in infants' joint attention abilities. As such, it seems reasonable to question whether joint attention plays a role in shaping children's sensitivity to normative expectations, and whether this relationship might be influenced by individual or cultural factors. However, joint attention is not the only early cognitive skill that has been put forward as a foundation of human normativity.

4.2.2 Overimitation and the Development of Normative Expectations

Our motivation to conform to normative expectations may also be an extension of the imitation abilities which start to develop during the first year of life (Carey, 2009; Essler et al., 2023; Tomasello & Carpenter, 2005). Imitation allows children to rapidly acquire skills and knowledge in a way that avoids costly trial-and-error learning (Bekkering, WohlschlaÈger & Gattis, 2000), and often, the tendency to imitate is so prominent that children show a tendency to copy all actions by an adult regardless of necessity - a behaviour which is now known as overimitation (Lyons, Young & Keil, 2007).

Humans' propensity to imitate faithfully has been linked to our species' unique capacities for culture. Despite a recent consensus that culture is widespread amongst many animal species (Whiten, 2021), it remains evident that these do not match the rich and intricate nature of cultural behaviours and traditions of the human species (Henrich, 2016; Tomasello, 2019; Whiten et al., 1999). One theoretical proposal is that our species' capacity for culture is supported at least in part by our

predisposition to overimitate, which allows humans to accurately learn cultural behaviours (including social norms) that accumulate and evolve over time to become more sophisticated or adaptive to local environments (Legare & Nielsen, 2015). Developmentally, overimitation in humans has been widely documented: it emerges from around 18 months of age (Nielsen, 2006) and increases over the course of childhood (Nielson & Tomaselli, 2010; McGuigan et al., 2007; McGuigan & Whiten, 2009). Moreover, the robustness of overimitation in humans has been demonstrated across multiple contexts, including naturalistic observations with adults and children (Whiten et al., 2016). In contrast, research suggests that non-human primates opt to ignore irrelevant actions in favour of efficiency, and as such, it has been reported that chimpanzees (Horner & Whiten, 2005), bonobos (Clay & Tennie, 2018) and orangutans (Nielsen & Susianto, 2010) do not overimitate (although see Huber et al., 2018; 2020 for evidence of overimitation in domestic dogs, and an opposing argument by Allen & Andrews, 2024). Overall, however, the evidence points to more robust overimitation in humans than other animals, which would be consistent with the idea that it is related to uniquely human psychological systems.

A large proportion of the literature on overimitation has focused on pinpointing the exact motivations for it, and debates have ensued as to whether faithful copying is more instrumentally or socially driven (Kim et al., 2018; Lyons et al., 2007; McGuigan et al., 2007; Nielsen, 2006; Nielson & Blank, 2011; Nielson, Simcock & Jenkins, 2008; Shimpi et al., 2013; Watson-Jones et al., 2014). Despite these alternative theories being pitted against each other frequently in the research, it is plausible that they are not mutually exclusive. Even some of the earliest accounts of imitation suggest that it might hold both a cognitive and social function (Užgiris, 1981), and more recent accounts of overimitation offer a similar argument in terms of motivation (Hoehl et al., 2019; Over & Carpenter, 2013). Consistent with the idea that these theories are not mutually exclusive comes a third argument: that the purpose of overimitation is inherently norm-based (Allen et al., 2021; Kenward et al., 2011; Keupp et al., 2013; Király et al., 2013), allowing children to learn cultural practices which are crucial to becoming competent members of their social group. Both the social and instrumental accounts of overimitation include elements of norm-learning (Allen & Andrews, 2024), as both involve children needing to decide what *should* be copied when learning how to interact with objects and people in a culturally appropriate way. This account of overimitation has clear parallels with normative expectations, as following normative expectations about how to behave means that individuals are attending to their beliefs about what *should* be done (either based on other people's behaviour, or what other people approve of).

There is strong evidence for an association between imitation and normativity. For example, after seeing a model produce relevant and irrelevant actions to retrieve a reward, children from 3 years of age will not only overimitate causally irrelevant actions, but they also spontaneously protest

omissions of the irrelevant actions when a puppet retrieved the reward, suggesting that they interpreted the irrelevant actions as normative (Keupp et al., 2013). Also from the age of 3, children can observe a single action and infer a social norm (Schmidt et al., 2016), indicating that behaviours children imitate can easily be represented by them as normative, and as such, object to violations of these norms. Additionally, from the age of 5 years, children overimitate more when it's accompanied by a normative statement (i.e. about how *we* as a group do things; Clay, Over & Tennie, 2018). Taken together, these studies suggest that overimitation is directly associated with normative behaviour. However, the question remains as to whether imitation serves as a springboard for the emergence of normativity later in life: do imitation abilities just co-occur with our capacity to follow and uphold social norms, or is our tendency to imitate a socio-cognitive precursor which extends into our capacity for normativity?

Little empirical work has been employed to explore the relationship between imitation and the human capacity for normativity, but evidence thus far suggests that normative behaviours may be rooted in our propensity for imitation. Essler et al., (2023) reported that 18-month-olds' tendencies to imitate were related to (i) self-distress in response to norm violations and (ii) subsequent normative protest to those norm violations. The interpretation is that a stronger capacity for imitation leads to better mental representations of normative behaviour (also supported by Zelazo & Lourenco, 2003). These mental representations, in turn, result in distress from the infants when behavioural actions conflict their mental image of how something *should* be done, and so they attempt to resolve that distress by verbally correcting the third party. This finding therefore suggests that imitation is specifically related to the acquisition and preservation of social norms. In support of this claim, Hardecker and Tomasello (2017) found that whilst both 2- and 3-year-old children imitated norm enforcement (i.e. correcting a third-party's interaction with an object), only the 3-year-old children were able to generalise this normative protest to a novel action. This further suggests that children's developing normativity is based on their capacity for imitation - as children age, they are better able to integrate their observations of what has previously been done to what *should* be done in future instances. This can also be theoretically applied to infants' understanding of approval: stronger imitation abilities can lead to better mental representations of what is deemed acceptable (or not) within their specific social group, and as such, an infant with a stronger tendency for imitation will be better able to create a richer, fine-tuned cognitive account of socially desirable behaviours.

As theorised, if normativity is grounded in our propensity to imitate and overimitate, then factors influencing overimitation could play an indirect role in the development of normative expectations. For the same reasons described above, for this project it is important to consider the influence of both

(i) culture and (ii) gender on overimitation, as this could be essential for interpreting the relationship between overimitation and normative expectations.

Similarly to joint attention, there remains a persistent WEIRD sampling bias in the overimitation literature (Hoehl et al., 2019), but the few cross-cultural studies which have been conducted suggest that overimitation is a broadly human phenomenon. Children living across the globe have been found to demonstrate similar engagement in overimitation in comparison to their Western counterparts, including children of South African Bushmen and Australian Aborigines (Nielson, Mushin, Tomaselli & Whiten, 2014), and children from Vanuatu (Clegg & Legare, 2016a) and Japan (Taniguchi & Sanefuji, 2017). With that said, during some studies, cultural context has been reported to influence children's overimitation. For example, although Chinese-American and Caucasian-American children copied inefficient tool use similarly after observing one model demonstrate using it, Chinese-American children imitated inefficient tool use more than Caucasian-American children when a consensus (i.e. three models) demonstrated using the inefficient tool (Corriveau et al., 2017). American and ni-Vanuatu children both overimitated more following a normative, conventional cue in comparison to one that was outcome-based, but ni-Vanuatu children overimitated more in the outcome-based scenario than American children (Clegg & Legare, 2016). Some have suggested that there is a quantitative cultural difference instead of a qualitative one – children from urban Germany showed more overimitation than children from rural Hai||om, despite a similar number of children demonstrating the presence of overimitation (Stengelin et al., 2020). Furthermore, in a study comparing two non-Western small-scale cultures from the Central African Republic (Aka hunter-gatherers and Ngandu horticulturalists), 4- to 7-year-old Aka children overimitated less than Aka adults and 4- to 7-year-old Ngandu children, who in turn, performed similarly to those of 23- and 30-month-old WEIRD children (Berl & Hewlett, 2015). Therefore, although overimitation is widely demonstrated across cultures, there seems to be some variation in the quantity of, and motivation for, overimitation in childhood. It seems plausible to propose that the frequency of skills involved in developing normativity may vary between cultures, as some cultures might encourage third-party observation, whereas other more interactive strategies might be more permissible in others (Legare & Harris, 2016). Overall, this research warrants exploration into whether the diverse cultural samples in this study show variation in overimitation.

In contrast to culture, there has been extremely little direct empirical exploration into the effect of gender on overimitation. One study, which explored overimitation in 5- to 12-year-old French and Serbian children, found that across both cultures, boys overimitated more than girls (Frick et al., 2017). This was an unexpected finding which the authors could not situate within the wider research context, as it was unclear in prior studies whether any influence of gender had been omitted, or

whether gender had not been included as a variable to begin with (Frick et al., 2017). Since this explicit discussion about gender influences, little research has directly explored the effect of gender on overimitation. One recent study explored overimitation in 3- to 5-year-old Catalonian girls and boys, and found that overimitation increased with age in boys, whereas overimitation remained stable (and high) in girls from an earlier age (Barbarroja et al., 2024). Other recent studies have found no difference in overimitation in young girls and boys (Schleihauf & Hoehl, 2021). Given the small number of relevant studies and inconclusive empirical evidence (much of which has primarily focused on older children), it is warranted to explore gender differences in overimitation in infancy within our samples, but it is difficult to make clear predictions about what we would expect to observe.

4.2.3 Domain-General Cognition and the Development of Normative Expectations

Both joint attention and overimitation are specific aspects of cognition for which the link to normativity is grounded in theory and indirect evidence. However, there may be other domain-general aspects of cognition that are related to normativity, which are not captured by either of those early abilities, such as memory, general attentional abilities, reasoning or pattern recognition. Recently, arguments have been put forward which have questioned whether human normativity is borne completely out of specialised processes, and which instead suggest that the human ability to comprehend social norms may also rely on domain-general cognitive processes combined with feedback from social interactions during development (Heyes, 2024). For example, through development, children may be able to recognise and memorise patterns of behaviour (using domain-general processes) that elicit social approval or disapproval (via social interaction), which leads to normative expectations being understood, and social norms being imitated and enforced. However, to date, there have been no empirical tests of whether normativity relies on domain-general processes. As highlighted above, when considering whether general cognitive processes form the ontogenetic roots of normativity, it is important to consider other factors which might lead to individual variation in domain-general processes, like how gender and culture may influence general cognitive development.

The Bayley Scales of Cognitive Development is a well-established tool for measuring many different skills and abilities in infancy and toddlerhood, including early cognitive skills (Del Rosario et al., 2021). Prior research has found that girls achieve higher average scores on the Bayley Cognitive Scale from 7 to 36 months in both Western and non-Western settings (Hua et al., 2019; Krogh & Væver, 2019; Suhonen et al., 2018), though a more recent study has suggested that although girls do perform significantly higher than boys, the effect size is negligible (Caemmerer et al., 2024). Studies exploring cultural differences in the Bayley Cognitive Scales have mostly aimed at determining whether the scales are valid for those cultures (Hua et al., 2019; Pendergast et al., 2018; Sun et al.,

2019), instead of comparing across cultures, with a couple of exceptions. One study found that Sri Lankan infants had higher cognitive scores at 12 months in comparison to the US norms, but lower cognitive scores at 24 months (Godamunne et al., 2014). Furthermore, another study found that German infants achieved higher scores at 3 months and 9 months than Cameroonian Nso infants, but there were no differences between the groups at 6 months (Vierhaus et al., 2011). Overall, these suggest that there might be cultural differences in cognition across infancy, and girls are likely to outperform boys, so it is necessary to consider how gender and culture might influence the role general cognitive development plays in the emergence of children's sensitivity to normative expectations.

4.2.4 Motivation for the Current Study

In order to explore whether joint attention, overimitation and/or general cognitive abilities support the emergence of normativity, specifically children's emerging sensitivity to normative expectations, a longitudinal design was needed so as to capture individual variation in both (i) children's cognition in infancy and (ii) children's sensitivity to normative expectations. Furthermore, exploring these cross-culturally in two very different samples will allow us to detect if the importance of any of these mechanisms is culturally bound, or present in both samples. To do this, we used a longitudinal cohort of children in the UK and Uganda who had already completed joint attention and overimitation tasks in infancy and completed the Normative Expectations task at 4 years. Due to the reduced size of the longitudinal sample (i.e. cross-sectional participants cannot be included), each infant measure was assessed separately, beginning with joint attention (Part 1) and then moving onto overimitation (Part 2). Whilst, as argued, both joint attention and overimitation are theoretically important for children's understanding of normative expectations, to ensure that these reflect domain specific relationships with children's emerging normativity, and are not simply artifacts of general developmental processes, it is important to explore how general developmental processes might be related to children's normative sharing. Therefore, the relationship between normativity and a general measure of cognitive abilities in infancy will also be explored (Part 3).

To assess joint attention, when infants were 15 months old mother-infant dyads were presented with unfamiliar animal calls (auditory stimuli), and the number of joint attention events the dyad engaged in was then measured. We adopted a tiered approach to explore the occurrence of joint attention events, which included analysing the most rigorous definition of a joint attention event (in which participants know together that they are sharing attention) and more liberal definitions of these events (where the sharing or joint nature of the interaction is less clear). We adopted this approach for two reasons. The first reason was due to the uncertainty about how common events defined with the most rigorous operationalisations of joint attention would be. Since recent debates about definitions of joint

attention events, research employing strict operationalisations is scarce, and especially so considering we were exploring auditory joint attention (instead of visual joint attention), across two distinct and partly under-researched societies. Ample individual variation is required to detect relationships between variables of interest, so we also coded for the more liberal definitions of joint attention events because prior studies suggest that children do vary in their engagement of events defined in this manner. The second reason was to ensure that our approach was comprehensive. Although we were predicting the most rigorous definition of joint attention events would relate to children's later normativity (as both involve aspects of shared attention and mutual knowledge), there is a possibility that engagement in more simply defined joint attention events could nonetheless be important building blocks which still support the emergence of normativity, so including them in the analysis allows for a fine-grained exploration into the predictive aspects of joint attention events. However, before using joint attention as a predictor for normative expectations, we firstly used joint attention as an outcome variable to assess how it is influenced by cultural group and gender. This leads to our first research question:

Research Question 1: Is engagement in joint attention events at 15 months predicted by cultural group or gender?

Due to the majority of research conducted suggesting some universality in joint attention abilities in infancy (Callaghan et al., 2011; Childers, Vaughan & Burquest, 2007), we hypothesized that our UK and Ugandan samples of children would engage in joint attention events at similar rates at 15 months. However, the evidence regarding gender is less conclusive; whilst many Western studies suggest that girls outperform boys in measures related to joint attention in infancy (Mundy et al., 2007; Olafsen et al., 2006; Perkovich & Yoshida, 2024; Raver & Leadbeater, 1995; Saxon & Reilly, 1999), less is known about how this might generalise to non-WEIRD groups. Therefore, whilst we expected girls to engage in higher rates of joint attention events than boys in our UK sample, this research question was exploratory in terms of what effects of gender might transpire in Uganda.

Research Question 2: Does the rate of joint attention events at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with cultural group or gender?

We hypothesised that if children's sensitivity to normative expectations (e.g. observation and observer approval) during the sharing task is predicted by the rate at which they engaged in joint attention events during infancy, this would suggest that the early developmental trajectory of engagement in joint attention facilitates the later development of normative expectations. As described in Chapter 3, we operationalise "normative expectations" by specifically focusing on comparing sharing behaviour

in the Approval Observe condition in comparison to other conditions. In practical terms, evidence of this relationship would result in an interaction between the Approval Observe condition and the rate of joint attention in predicting the proportion of tokens given away. This would support the Shared Intentionality model of the development of normativity (Tomasello, 2019), as it would suggest that infants who more frequently engage in joint attention events become better able to represent and contrast alternative viewpoints, eventually leading to better normative self-governance.

Because (i) Tomasello's Shared Intentionality model does not imply that there should be differences between cultures in the relationship of joint attention and normative behaviour, (ii) studies suggest that there are broad similarities in joint attention abilities across cultural groups in infancy (Callaghan et al., 2011), and (iii) cultural differences in children's normativity were not observed in Chapter 2, we do not predict cultural differences in the relationship between joint attention and normative expectations. However, if cultural group is observed to play a role in the rate of infant engagement in joint attention events, in the model addressing whether joint attention predicts normativity we will include cultural group in the interaction term to further explore how culture might shape this relationship. In contrast, due to the gender differences found in early children's normativity in Chapter 2 (i.e. boys showed the pattern of sensitivity to normative expectations, whereas girls tended to share more generously overall), it is important to consider how gender might influence the relationship between early socio-cognitive skills and developing normativity. As such, gender will be included in the interaction term to assess whether any observed relationship between normativity and joint attention is expressed differently in girls and boys. However, given this is the first empirical test of this relationship, it would be premature to make strong predictions about the role gender might play, and as such, this research question was exploratory.

To assess overimitation, 18-month-old infants were presented with a transparent puzzle box (based on the classic stimuli from Horner & Whiten, 2005), and a model demonstrated three causally irrelevant actions before demonstrating one causally relevant action to retrieve the reward. Firstly, we assessed overimitation as an outcome variable to understand the influence of cultural group and gender on the likelihood of infants engaging in overimitation:

Research Question 3: Is overimitation at 18 months predicted by cultural group or gender?

Whilst overimitation has been observed across a vast range of diverse cultural groups (Clegg & Legare, 2016; Nielson et al., 2014; Taniguchi & Sanefuji, 2017), some studies suggest that cultural variation exists, predominantly in the quantity of overimitation (Berl & Hewlett, 2015), or the context in which overimitation occurs (Clegg & Legare, 2016; Corriveau et al., 2017). Therefore, although we are expecting both samples to demonstrate overimitation, it is unclear from previous research whether

there will be any cultural variation in our Ugandan and UK participants. Furthermore, given the inconclusive and limited evidence about the influence of gender on overimitation (e.g. Barbarroja et al., 2024; Frick et al., 2017; Schleihauf & Hoehl, 2021), particularly in infancy, it is unknown how gender might influence infant's overimitation in these samples. As such, we took an exploratory approach to this question.

Research Question 4: Does overimitation at 18 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with cultural group or gender?

We hypothesised that if children's sensitivity to normative expectations during the sharing task is predicted by their overimitation during infancy, this would add weight to the claim that imitation, and specifically overimitation, is an important building block in children's developing normativity, as hypothesised by many researchers thus far (Carey, 2009; Essler et al., 2023; Tomasello & Carpenter, 2005). As suggested by Essler et al., (2023), children's ability to imitate is hypothesized to be related to them creating a better mental representation of normative information (i.e. more understanding about what others approve of). Given that children can infer a social norm from witnessing one action (Schmidt et al., 2016), it seems reasonable that children will be able to infer a social norm of generosity by (i) seeing a model act generously whilst playing the game and (ii) hearing the experimenter express approval of generosity. If these theoretical predictions are supported, we would expect to see in our statistical models a significant interaction between the Approval Observe condition and overimitation in predicting sharing behaviour. If gender and culture are found to have a salient influence on infants' overimitation at 18 months (RQ3), we will explore further how this variation might affect the predictive relationship between overimitation and children's sensitivity to normative expectations. However, analyses of how gender and cultural group interact with overimitation to support normativity were exploratory.

Finally, to assess children's general cognitive development in infancy, at 11 months we used the Cognitive Scale of the Bayley Scales of Infant and Toddler Development – Third Edition (Bayley-III; Bayley, 2006). Prior research suggests that cognition from 9 months (using this measure) predicts children's IQ at 5 years (Klein-Radukic & Zmyj, 2023), showing that this scale has been used to predict children's characteristics at a much later date in development. This scale has also been used with non-WEIRD participants (Hua et al., 2019; Pendergast et al., 2018; Sun et al., 2019) including Kenya (McHenry et al., 2021), which corroborates its successful application outside of a WEIRD context. This allowed us to explore the specificity of any observed relationships between joint attention and/or overimitation and normativity, and test whether domain-general processes may be important, leading to our final set of research questions:

Research Question 5: Is cognition at 11 months predicted by cultural group or gender?

This question is included to follow the same logic as the other research questions, in that the analyses will determine which factors should be included in the predictive model. Although not comprehensive, prior research hints that girls might achieve higher scores on the Bayley Cognitive Scale in infancy (Hua et al., 2019; Krogh & Væver, 2019; Suhonen et al., 2018). Research also suggests that there may be differences between cultural groups (Godamunne et al., 2014; Vierhaus et al., 2010), however, this is likely quite specific to the cultural groups being compared. Therefore, due to us not knowing whether these results are generalisable to our Ugandan participants, our approach to this question remains exploratory.

Research Question 6: Does cognition at 11 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with cultural group or gender?

We hypothesized that if children's normative sharing in the Normative Expectation task was predicted by cognitive abilities at 11 months, any observed relationship between other variables might reflect general developmental processes, rather than specific skills upon which normativity is built. No previous research has explored the link between general cognitive ability and sensitivity to normative expectations. As such, this question was exploratory, and we made no prior predictions.

4.3 Part 1: Research Questions 1 and 2

4.3.1 Methods

4.3.1.1 Participants

The sample consisted of participants who had valid data from the Normative Expectations task at 4 years (described in Chapter 1; Study 1) and valid joint attention data at 15 months. Out of the 78 longitudinal participants who had valid Normative Expectations data, 38 UK participants, and 36 Ugandan participants had valid joint attention data at 15 months, and so the final sample consisted of 74 participants (36 girls). Further details regarding the sample can be viewed in Table 4.1.

Table 4.1

The N and mean age of the participants with valid data for the Auditory Joint Attention task in each cultural group.

Task	Uganda					UK				
	N	Girls	Boys	Mean Age[1]	SD[2]	N	Girls	Boys	Mean Age[1]	SD[2]
Auditory Joint Attention	36	17	19	1.26	0.01	38	19	19	1.24	0.03

[1]Age measured in years

[2]Standard deviation in years

4.3.1.2 Procedure and Materials

4.3.1.2.1 Research Design and Procedure Overview

The data in this study were collected from a longitudinal sample, both including participants from the UK and Uganda. Data were collected by local researchers, who were fluent in the participant's language in both the UK and Uganda, during 2-to-3-hour home visits, which usually took place outside in Uganda, whereas it was most commonly inside in the UK. The Auditory Joint Attention experiment was conducted during a home visit at 15 months, and the Overimitation task was conducted during a home visit at 18 months. The Normative Expectations task was conducted at a home visit at 4.5-years (as described in Chapter 2).

4.3.1.2.2 Auditory Joint Attention Task

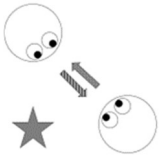
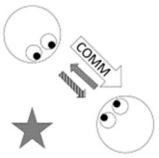
To present exciting and novel auditory stimuli that the dyad could share attention about, eight 'weird noises' from an unfamiliar animal were presented to the dyad via a hidden Bluetooth speaker (Betron KBS08 model). The stimuli consisted of a series of 1 - 8 second animal noises from either i) a zebra, (ii) a peacock or (iii) an orca (counterbalanced across participants). Where these sounds were obtained from can be found in Appendix 4.1.1.1. All sounds were cut to the appropriate length and equalised in RMS amplitude in Raven Pro 1.6 software. These animals were chosen as their calls were likely to be unfamiliar to both testing groups, and so should have been novel to the participants. Mothers were informed the experimenter was about to play some "weird noises", and asked to react to them as if they had heard something unexpected and novel, using culturally-appropriate examples; in the UK, mothers were told to react as if they had heard "the siren of an ambulance", whereas in Uganda, mothers were told to react as if they had heard "a plane flying over". Mothers were instructed to sit facing their infant, and to play as they naturally would (which could include non-noisy toys). When the task began, the stimuli were played to the dyad over a period of around 3 minutes, with pauses being left between each sound file to give the mother and child time to communicate about the sound. Trials were terminated if the infant lost attention or became distressed. The experiment was filmed by two experimenters positioned behind two cameras from different angles: one always focusing on the mother and one always focusing on the infant. A third camera on a tripod was sometimes used if the testing space was small and limited the movement of the experimenters attempting to film all shared looks between the dyad and towards the speaker.

4.3.1.2.3 Measures of Joint Attention

Most measures of joint attention require visual attention on the external stimulus before visual attention to the partner (gaze alternation between object/event and the partner). However, when the stimulus is auditory and loud enough to be salient, it can be assumed that everyone in the area heard

the sound, so visual orientation towards it is not necessary (Battich et al., 2020). Therefore, we looked for instances of mutual gaze between mother and infant during or in the 3 seconds after the sounds were playing as triadic attention events (where we assumed the mutual gaze was about the salient auditory event that was occurring or just occurred). To be confident that the mother and infant knew they were attending together to the sound, we looked for communication (facial expression, gesture, vocalisation, speech) from the infant and mother during or just after mutual gaze. We identified events in which just one partner communicated and then events where both partners communicated, giving us our most rigorously defined joint attention event, where the jointness of the interaction was most apparent (Carpenter & Liebal, 2011; Graham et al., 2021; Sipsova & Carpenter, 2019; see Table 4.2).

Table 4.2
Joint attention levels coded for the Auditory Joint Attention task

Category	Response to stimulus	Sequence of behaviours	
(i) Triadic Attention	Both hear sound (orientation to sound not necessary as it is salient enough to assume everyone in the area heard it)	Then engage in mutual face direction (coordination of attention)	
(ii) Triadic Attention with Communication	Both hear sound (orientation to sound not necessary as it is salient enough to assume everyone in the area heard it)	Then engage in mutual face direction with at least ONE individual communicating	
(iii) Shared Attention	Both hear sound (orientation to sound not necessary as it is salient enough to assume everyone in the area heard it)	Then engage in mutual face direction with BOTH individual communicating	

4.3.1.3 Video Coding

4.3.1.3.1 Trial Validity

One longitudinal participant did not complete the Auditory Joint Attention task at 15 months, so could not be included in the analyses. Three additional participants were excluded from the final dataset (2 UK; 1 Ugandan). For one participant, the mother received incorrect instructions and so the data were not valid. The other two participant exclusions were due to an insufficient number of stimuli being presented. The protocol required that each participant was presented with 8 stimuli over the course of the trial. However, on 12 trials more than 8 stimuli were presented (n = 11 experimenter error; n = 1 speaker technical issue) and on two trials fewer than 8 stimuli were presented. Only trials where 7-9 stimuli were presented (and were not interrupted by third parties) were included in the final dataset. If there were more than 9 stimulus presentations during the trial, only the first 9 presentations were considered for coding (N = 2 trials). If there were fewer than 7 presentations, those participants were excluded (N = 2; trials were terminated early due to infant distress or interference by third party).

4.3.1.3.2 Auditory Joint Attention Coding

The first phase of video coding was completed by the lead author and a team of three research assistants using The Observer XT 14 software (Noldus). The second phase was completed sequentially by another coder and the lead author, also using The Observer XT 14 software. Table 4.3 contains definitions of the terms used in the auditory joint attention coding, and details of what occurred during each phase of coding. Phase 1 consisted of the initial video coding, in which basic coding of the Experimental Period, stimulus presence and gaze of the mother and infant took place (Table 4.3). Phase 2 consisted of identifying periods of mutual gaze that occurred during or just after stimulus presentation, and coding what (if any) communication occurred during or near the period of mutual gaze (Table 4.3). The full coding scheme for the auditory joint attention experiment is in Appendix 4.1.1.2.

Table 4.3
Overview of the Auditory Joint Attention task coding procedure

Phase	Step	Definition	Purpose
1	1) Code Experimental Period	Began when the first stimulus was presented until 30 seconds after the last stimulus was played, or when the experimenter said the trial was finished. Any periods in which there were disruptions (e.g. the mother or infant were interacting with a third person, the infant strayed from the testing area, or the infant became distressed; see 'Not Coded' in the coding scheme listed in Appendix 4.1.1.2) were not included in the Experimental Period.	Isolate the period of testing in which the stimuli were presented to the mother-infant dyad uninterrupted, and shared attention may occur.
	2) Code Stimulus Presence	Stimulus presence or absence (i.e. when a 'weird noise' was and was not being played) were coded throughout the Experimental Period	Identify periods in which shared attention was most likely to be attributed to the stimuli
	3) Code Gaze Direction	Both mother and infant gaze direction were coded continuously during the Experimental Period, including gaze to the social partner's face, gaze towards the stimulus (if the location of the speaker was known), gaze to the experimenter, gaze to other individuals, gaze elsewhere and periods where gaze direction was unclear. For the purpose of this study, only (i) 'gaze to the social partner's face' and (ii) 'gaze direction unclear' were used. We used 'gaze to the social partner's face' to identify periods of mutual gaze (i.e. when the mother's gaze was directed to the infant's face, and when the infant's gaze was directed towards the mother's face). We used 'gaze direction unclear' to calculate the period in which we could have	Identify when the partner was looking towards the infant's face, and the infant was looking towards the partner's face.

Phase	Step	Definition	Purpose
		detected mutual gaze if it had occurred (e.g. removing periods where gaze of infant or partner was unclear).	
2	4) Identify Mutual Gaze Codable Periods and code Mutual Gaze	Cases of mutual gaze were identified by extracting instances of overlap between when the infant looked at the partner's face, and the partner looked at the infant's face (see Code Gaze Direction in Phase 1). We only recorded cases of mutual gaze that occurred during the stimulus being present (Code Stimulus Presence in Phase 1), or within 3 seconds of the stimulus offset. Figure 4.1A shows an example of mutual gaze after the offset of the stimulus, whereas Figure 4.1B shows an example of mutual gaze occurring during the stimulus presentation.	Find candidate joint attention events. It was theorised that any joint attention events occurring i) during the sound being played and ii) shortly after the sound finished playing, would likely be a direct result of the 'weird noise', and not some other entity the mother-infant dyad wished to share attention about.
	5) Identify Communication Codable Periods and code Communication	In order to determine if one or both partners communicated about the stimulus, we first identified Communication Codable Periods. These periods started at the onset of mutual gaze until 3 seconds after the offset of mutual gaze (Figure 4.1A and 4.1B). In some instances, the onset of the next stimulus triggered the Communication Codable Period to be cut short as a new Mutual Gaze Codable Period had begun, and as such, the target of any mutual gaze or communication was likely to be the most recent stimulus; see Figure 4.1C). Once Communication Codable Periods were identified (Figures 4.1A - C), the presence of any vocalisations, gestures and facial expressions produced by the infant and the partner were coded within these periods. The production of any type of communication was	Determine the level of joint attention that occurred (i.e. whether it was triadic attention, triadic attention with communication or shared attention).

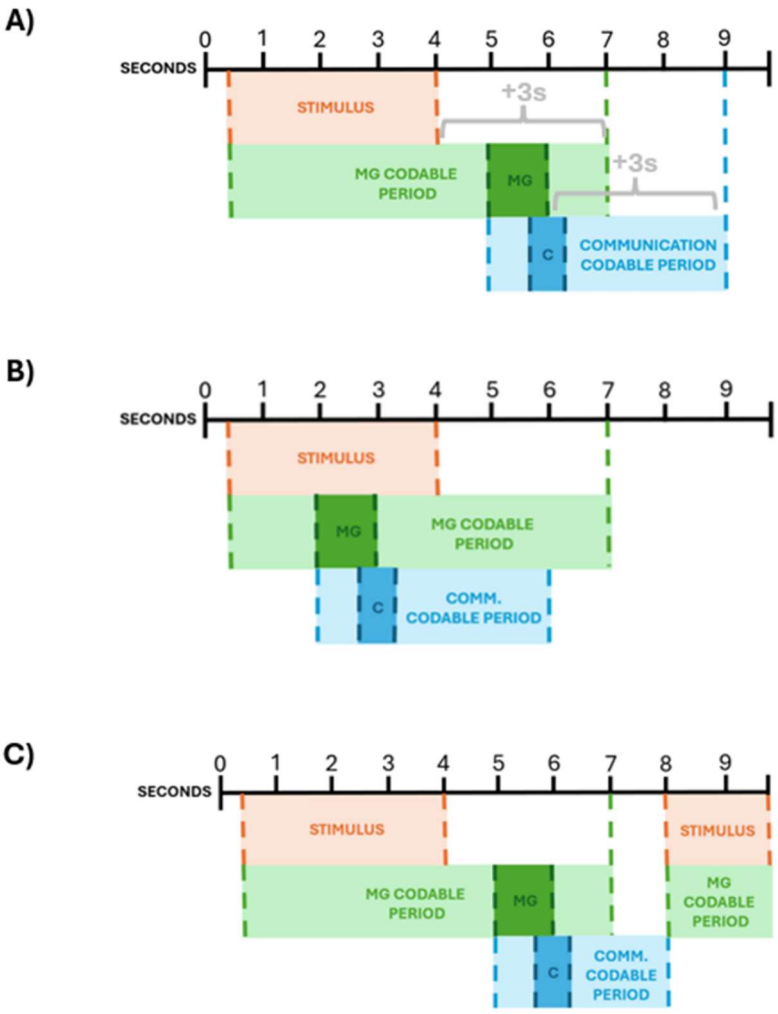
Phase	Step	Definition	Purpose
		considered to be sufficient to code the infant or partner as having communicated about the stimulus.	

Interobserver reliability of phase 1: To assess reliability, 20.00% of videos (N = 18 / 90) were double coded by the lead author and the three research assistants (6 videos each). Cohen's Kappa's (Cohen, 1960) for both the duration/sequence and frequency/sequence of (i) the Experimental Period, (ii) the infant's gaze to the partner's face alongside periods of stimulus presence, and (iii) the partner's gaze to the infant's face alongside periods of stimulus presence were between 0.71 - 0.99, indicating excellent agreement (see full details in Appendix 4.1.1.3).

Interobserver reliability of phase 2: To assess reliability, the lead coder and the lead author both coded 20.94% of the mutual gaze and subsequent Communication Codable Periods (N = 94 / 449 mutual gaze events). Cohen's Kappa's were 0.75 for mutual gaze, 0.77 for the presence/absence of infant communication and 0.70 for the presence/absence of partner communication, indicating excellent reliability.

Figure 4.1

Three example timelines of Auditory Joint Attention event coding



Note. A) For a triadic attention event, mutual gaze must have occurred during the stimulus presentation or within 3 seconds after the offset of the stimulus (Mutual Gaze Codable Period), as can be seen in light green, 'MG CODABLE PERIOD'. If mutual gaze occurred during this period (as can be seen in dark green, 'MG'), the Communication Codable Period began at the onset of mutual gaze, until 3 seconds after its offset as can be seen in light blue, 'COMMUNICATION CODABLE PERIOD'. 'C' refers to communication being present, so depending upon whether one or both individuals communicated, this would be coded as triadic attention with communication or shared attention, respectively. B) A second example, where MG occurs during the presentation of the stimulus. C) If another stimulus is presented before the Communication Codable Period ends, it is cut short, and the new coding episode begins from the onset of the new stimulus.

When calculating the final rates of each level of joint attention event to be used in the statistical models, we wanted to calculate a per-minute rate for each level of joint attention event which accounted for the available time in which participants *could* have shared attention about the stimuli during the task. To do this we first calculated the number of (i) triadic attention events, (ii) triadic attention with communication events and (iii) shared attention events. We then calculated the Total Codable Period, by summing each stimulus duration plus 3 seconds (i.e. each Mutual Gaze Codable Period) throughout the Experimental Period, and subtracting any time which one/both individuals gaze was unclear (i.e. thereby ensuring that it was not assumed that no mutual gaze was occurring just because an individual's gaze was not clear). We then calculated three separate joint attention event rates by dividing the number of each type of joint attention event by the Total Codable Period.

4.3.1.4 Statistical Approach

The statistical approach mirrored that of Chapter 3, including the use of the 'Approval Observe condition' variable, instead of analysing approval and observation separately in the models (as done in Chapter 2).

Research Question 1: Model 1

Is engagement in joint attention events at 15 months predicted by cultural group or gender?

To analyse how the frequency of joint attention events varied with cultural group (UK and Uganda) or gender (girls and boys), and their interaction, a GLM was used with a negative binomial error structure and log link function (McCullagh & Nelder, 1989). The most rigorous definition of joint attention event, shared attention, was included as the outcome variable (number of shared attention events). The length of time (in seconds) in which shared attention events could be coded (Total Codable Period) varied between participants and so was included as an offset term (McCullagh & Nelder, 1989). The model was not overdispersed or zero-inflated. Details of model construction can be viewed in Table 4.4 (Model 1). See Appendix 4.1.2.1 for an outline of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 74 shared attention frequencies from 74 participants.

Research Question 2: Model 2

Does the rate of joint attention events (shared attention events) at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

This model tested the more rigorous definition of joint attention; the rate of shared attention events (in which both partners needed to communicate during the joint attention event). A description of this model can be viewed in Table 4.4 (Model 2). See Appendix 4.1.2.2 for model stability parameters and

details of the process followed to reach the final model. The sample consisted of 296 proportions of tokens given, from 74 participants over 4 trials.

Research Question 2: Model 3

Does the rate of joint attention events (triadic attention with communication events) at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

This model was included to explore whether the rigorous definition of joint attention used (i.e. both partners needing to communicate) was reducing individual variation in our predictor variable and impacting our ability to observe a relationship between joint attention and normativity. Therefore, the rate of triadic attention events, in which only one partner communicated, was used as a predictor instead of the rate of shared attention events. A description of this model can be viewed in Table 4.4 (Model 3). See Appendix 4.1.2.3 for model stability parameters and details of the process followed to reach the final model. The sample consisted of 296 proportions of tokens given, from 74 participants over 4 trials.

Research Question 2: Model 4

Does the rate of joint attention events (triadic attention events) at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

As in Model 3, this model was included as the most liberal test of joint attention events, in which attention is operationalised as both partners engaging in mutual gaze in response to presentations of the stimuli. Neither partner needed to communicate in this instance, increasing the individual variation in this measure. A description of this model can be viewed in Table 4.4 (Model 4). See Appendix 4.1.2.4 for model stability parameters and details of the process followed to reach the final model. The sample consisted of 296 proportions of tokens given, from 74 participants over 4 trials.

Research Question 2: Model 5

Does the rate of joint attention events (shared attention events) at 15 months predict sensitivity to normative expectations when children are 4 years old, controlling for cultural group and gender?

This question was included as a check, to ensure that statistical power was not limiting our ability to detect a significant interaction between the Approval Observe condition and the rate of joint attention events at 15 months (as measured by shared attention events) in a complex model which included an interaction with gender. Therefore, cultural group and gender were included as fixed control factors (rather than as interaction terms) and were also included in the null model for the full-null comparison. A description of this model can be viewed in Table 4.4 (Model 5). See Appendix 4.1.2.5

for model stability parameters and details of the process followed to reach the final model. The sample consisted of 296 proportions of tokens given, from 74 participants over 4 trials.

Table 4.4

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between Joint Attention and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 – 5).

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
1	Negative Binomial GLM	Number of Shared Attention Events	Gender* Cultural Group	Offset(Total Codable Period)	NA	NA	0.99	Gender = 1.00; Cultural Group = 1.00
2	Binomial GLMM	Tokens given|kept in test trials	AO^3*Rate of Shared Attention Events *Gender	Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO^3+ Rate of Shared Attention Events) ||Experimenter on Trial[2] (Gender + AO^3+ Rate of Shared Attention Events + Cultural Group)||Order of Conditions[2]	0.99	AO^3 = 1.00; Rate of Shared Attention Events = 1.03; Gender = 1.01; Cultural Group = 1.02
3	Binomial GLMM	Tokens given|kept in test trials	AO^{3*} Rate of Triadic Attention with Communication Events* Gender	Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO^3+ Rate of Triadic Attention with Communication Events)||Experimenter on Trial[2] (Gender + AO^3+ Rate of Triadic Attention with Communication	1.01	AO^3 = 1.00; Rate of Triadic Attention with Communication Events = 1.00; Gender =

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
						Events + Cultural Group)||Order of Conditions[2]		1.00; Cultural Group = 1.01
4	Binomial GLMM	Tokens given|kept in test trials	AO[3]* Rate of Triadic Attention Events* Gender	Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3]+ Rate of Triadic Attention Events) ||Experimenter on Trial[2] (Gender + AO[3]+ Rate of Triadic Attention Events + Cultural Group)||Order of Conditions[2]	1.01	AO[3] = 1.00; Rate of Triadic Attention Events = 1.00; Gender = 1.00; Cultural Group = 1.00
5	Binomial GLMM	Tokens given|kept in test trials	AO[3]*Rate of Shared Attention Events[4]	Gender + Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3]+ Rate of Shared Attention Events) ||Experimenter on Trial[2] (Gender + Cultural Group + AO[3]+ Rate of Shared Attention Events)||Order of Conditions[2]	0.99	AO[3] = 1.00; Rate of Shared Attention Events = 1.02; Gender = 1.01; Cultural Group = 1.02

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts were unidentifiable and therefore excluded from the model

[3]AO = Approval Observe condition

4.3.2 Results

4.3.2.1 Research Question 1

Model 1: Is engagement in joint attention events (shared attention events) predicted by cultural group or gender?

Overall, the full model did not explain significantly more variation in the frequency of shared attention at 15 months than the null model, which omitted the fixed effects of gender, cultural group and their interaction (likelihood ratio test: $\chi2 = 3.80$, df = 3, p =.284). This indicates that cultural group, gender, or their interaction, did not have a significant influence on participants' shared attention event frequency, so as such, it was unnecessary to consider the coefficients of this model further.

4.3.2.2 Research Question 2

Model 2: Does the rate of joint attention events (shared attention events) at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'cultural group' and the random effects (likelihood ratio test: $\chi2 = 11.49$, df = 7, p = .119). Therefore, we can assume that neither the 'Approval Observe condition', 'rate of shared attention events' or 'gender', nor the interactions between them, significantly influenced the proportion of tokens children gave. As such, the coefficients were not further explored.

Model 3: Does the rate of joint attention events (triadic attention with communication events) at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'cultural group' and the random effects (likelihood ratio test: $\chi2 = 12.67$, df = 7, p = .103). Therefore, we can assume that neither the 'Approval Observe condition', 'rate of triadic attention with communication events' or 'gender', nor the interactions between them, significantly influenced the proportion of tokens children gave. As such, the coefficients were not further explored.

Model 4: Does the rate of joint attention events (triadic attention events) at 15 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'cultural group' and the random effects (likelihood ratio test: $\chi 2 = 11.64$, df = 7, $p = .113$). Therefore, we can assume that neither the 'Approval Observe condition', 'rate of triadic attention events' or 'gender', nor the interactions between them, significantly influenced the proportion of tokens children gave. As such, the coefficients were not further explored.

Model 5: Does the rate of joint attention events (shared attention events) at 15 months predict sensitivity to normative expectations when children are 4 years old, controlling for cultural group and gender?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'gender', 'cultural group' and the random effects (likelihood ratio test: $\chi 2 = 5.76$, df = 3, $p = .124$). Therefore, we can assume that the 'Approval Observe condition', 'rate of shared attention events', nor their interaction, significantly influenced the proportion of tokens children gave. As such, the coefficients were not further explored.

4.4 Part 2: Research Questions 3 and 4

Against theoretical predictions, none of the three levels of joint attention events in infancy that we examined related to children's sensitivity to normative expectations. This will be discussed in more detail in the Discussion. Another possibility, which has yet to be explicitly tested, is that children's emerging normative prosociality may be an extension of imitation abilities that are acquired early in development (Carey, 2009; Essler et al., 2023; Tomasello & Carpenter, 2005). Such is the propensity to imitate, that children are often reported to faithfully copy all actions in achieving a goal, regardless of relevancy. This chapter will now turn to exploring this tendency to overimitate, with a focus on whether overimitation in infancy is related to children's sensitivity to normative expectations later in childhood. As stated in the introduction, it was theorized that if this relationship between normativity and overimitation exists, the children who overimitate would be more likely to share more in the Approval Observe condition in comparison to the other conditions.

4.4.1 Methods

4.4.1.1 Participants

The sample consisted of participants who had valid data for the Normative Expectations task at 4 years (described in Chapter 2) and valid overimitation data at 18 months. Out of the 78 longitudinal participants who had valid data for the Normative Expectations task, 73 participants had valid overimitation data at 18 months participants (36 girls; see Table 4.5 for details).

Table 4.5

The N and mean age of the participants with valid data for the Overimitation task in each cultural group.

Task	Uganda					UK				
	N	Girls	Boys	Mean Age[1]	SD[2]	N	Girls	Boys	Mean Age[1]	SD[2]
Overimitation	33	17	16	1.51	0.02	40	19	21	1.49	0.02

[1]Age measured in years

[2]Standard deviation in years

4.4.1.2 Procedure and Materials

4.4.1.2.1 Overimitation Task

The procedure was modelled on Horner and Whiten (2005). Participants were presented with a large transparent puzzle box which contained a toy (a plastic egg) and a stick which could be used as a tool to retrieve the toy. The puzzle box had two openings: one functional, and the other non-functional. One experimenter (E1) sat holding the stick next to the infant, who usually sat on or beside their mother, with the puzzle box in front of them. The mother was asked to remain distracted throughout the trial, so that she would not influence the infant's behaviour in any way. The set-up of the task in the UK and Uganda can be viewed in Figure 4.2. The trial began with E1 trying to get the infant's attention by saying their name and, "Look!". Once E1 had the infant's attention, they demonstrated three non-functional actions and one functional action to retrieve the reward from the box. The non-functional actions were (i) tapping the stick on the palm of their upturned hand, (ii) moving the stick in a circle motion around the top of the box and (iii) tapping the stick in the top, non-functional opening of the box (the order was counterbalanced across children). The final, functional action was to put the stick in the functional opening of the box and poke the toy, which fell through a vertical tube out the bottom of the puzzle box and could then be retrieved by E1. Then, E1 showed the toy to the child by saying, "Look!", whilst a second experimenter (E2) discreetly reloaded another toy into the puzzle box outside the infant's field of view. The puzzle box was then placed back down in its position in front of E1 and the infant, and E1 gave the child the stick and said, "You try!". The infant was then allowed to interact with the box for 1 minute. If the child did not attempt to retrieve the toy, E1 repeated, "You try!" every 10 seconds. The trial ended when (i) the child had retrieved the egg, (ii) the child unsuccessfully tried to get the egg for 1 minute, or (iii) the child did not engage in the task for 1 minute. The experiment was filmed using two cameras for later video coding.

Figure 4.2
The set-up of the Overimitation task in the UK (left) and Uganda (right).



4.4.1.3 Video Coding

4.4.1.3.1 Trial Validity

No trials were excluded from the final dataset.

4.4.1.3.2 Overimitation Coding

From the videos, imitation of the three irrelevant actions was manually coded and entered onto a Microsoft Excel Spreadsheet. Initially, an overimitation score was calculated for each participant by summing the number of imitated irrelevant actions (range 0-3). However, due to no participants imitating more than one of the irrelevant actions (i.e. all participants only ever overimitated one action), coding was transformed into a binary variable which categorised whether overimitation occurred in the trial ('Yes' or 'No').

Interobserver reliability: After training, all trials were coded by a research assistant and, a second research assistant coded a random selection of 51.9% (N = 42) of the total trials. Interobserver reliability between the two research assistants was calculated by judging agreement on whether children overimitated). Cohen's Kappa (Cohen, 1960) for overimitation status was equal to 1, indicating perfect interobserver reliability.

4.4.1.4 Statistical Approach

The statistical approach was the same as Part 1, and followed the outline set out in Chapter 2, Study 1.

Research Question 3: Model 6

Is overimitation at 18 months predicted by cultural group or gender?

Details of model construction can be viewed in Table 4.6 (Model 6). See Appendix 4.2.1.1 for an outline of the process followed to reach the final model and stability parameters of the final model. The sample consisted of 73 overimitation statuses from 73 participants.

Research Question 4: Model 7

Does overimitation at 18 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with cultural group or gender?

A description of this model can be viewed in Table 4.6 (Model 7). See Appendix 4.2.1.2 for model stability parameters and details of the process followed to reach the final model. The sample consisted of 292 proportions of tokens given, from 73 participants over 4 trials.

Research Question 4: Model 8

Does overimitation at 18 months predict sensitivity to normative expectations when children are 4 years old, controlling for cultural group or gender?

This question was included as a check to ensure that statistical power was not limiting out ability to detect a significant interaction between the Approval Observe condition and the presence of overimitation. Therefore 'gender' and 'cultural group' were included as a fixed control effect (rather than in the interaction term) and were also included in the null model during the full-null comparison. A description of this model can be viewed in Table 4.6 (Model 8). See Appendix 4.2.1.3 for stability parameters and details of the process followed to reach the final model. The sample consisted of 292 proportions of tokens given from 73 participants over 4 trials.

Table 4.6

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between overimitation and (i) gender and cultural group (Model 1) and (ii) sensitivity to normative expectations (Models 2 and 3).

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
6	Binomial GLM	Overimitation (Yes/No)	Gender*Cultural Group	NA	NA	NA	1.06	Gender = 1.00; Cultural Group = 1.00
7	Binomial GLMM	Tokens given|kept in test trials	AO[3] *Overimitation *Gender *Cultural Group	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3]+ Overimitation) ||Experimenter on Trial[2] (Gender + AO[3]+ Cultural Group + Overimitation) ||Order of Conditions[2]	1.13	AO[3] = 1.00; Overimitation = 1.32; Gender = 1.00; Cultural Group = 1.32
8	Binomial GLMM	Tokens given|kept in test trials	AO[3]* Overimitation	Gender + Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3]+ Overimitation) ||Experimenter on Trial[2] (Gender + AO[3]+ Cultural Group + Overimitation) ||Order of Conditions[2]	1.10	AO[3] = 1.00; Overimitation = 1.30; Gender = 1.00; Cultural Group =1.30

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts were unidentifiable and therefore excluded from the model

[3]AO = Approval Observe condition

4.4.2 Results

4.4.2.1 Research Question 3

Model 6: Is overimitation at 18 months predicted by cultural group or gender?

Overall, the full model explained significantly more variation in the presence of overimitation at 18 months than the null model, which omitted the fixed effects of gender, cultural group and their interaction (likelihood ratio test: $\chi 2 = 20.86$, df = 3, $p < .001$). The interaction between 'cultural group' and 'gender' was non-significant (estimate = -0.86, SE = 1.43, $p = .549$; Appendix 4.2.1.2), so this interaction was omitted in the process of reducing the model. In the final model there was a significant main effect of cultural group (Table 4.7), which suggested that significantly more UK participants engaged in overimitation than Ugandan participants (Figure 4.3).

Table 4.7
Results of final Model 6 examining whether overimitation is predicted by gender or cultural group (estimates and standard errors, together with confidence limits and significance tests).

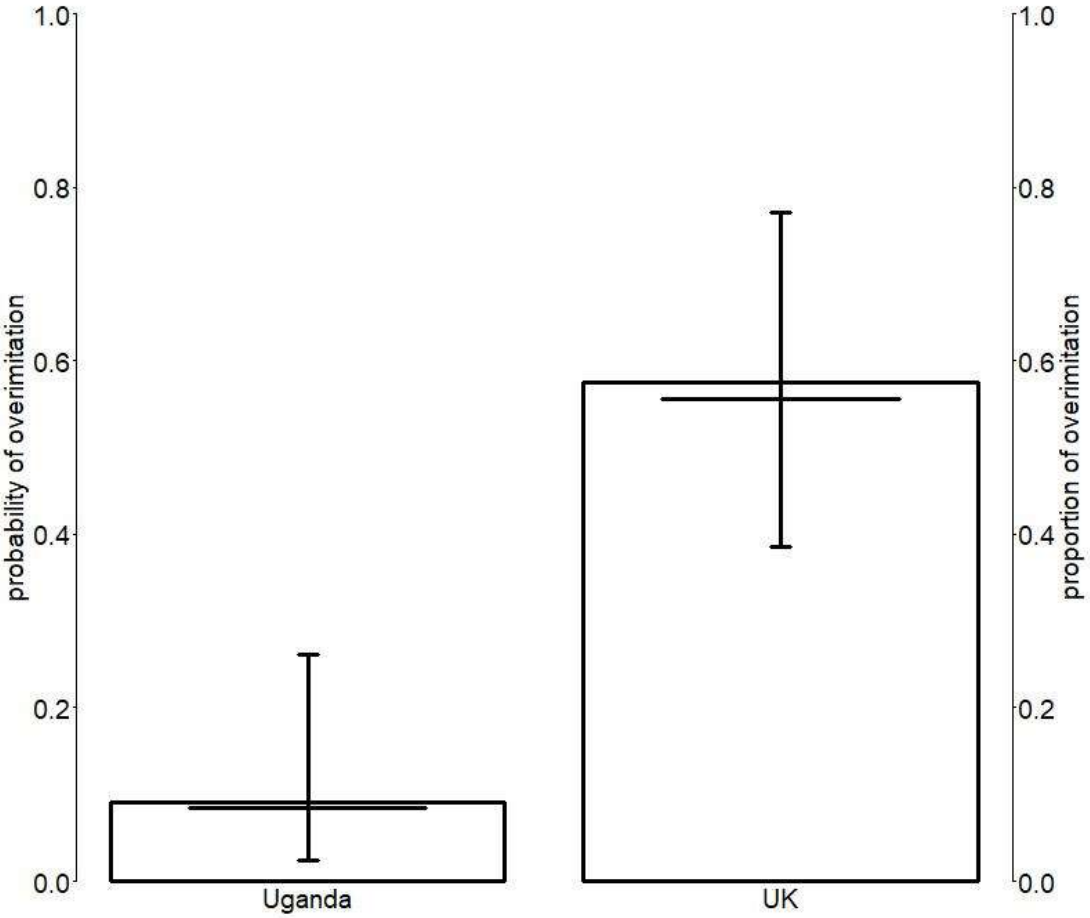
Term	Estimate	SE	Lower CI	Upper CI	z-value	p
Intercept	-2.38	0.67	-3.92	-1.21	-3.54	<.001
Cultural Group[(1)]	2.60	0.68	1.38	4.14	3.80	<.001
Gender[(2)]	0.15	0.57	-0.97	1.27	0.27	.790

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

Figure 4.3

Probability that overimitation was present in Uganda and the UK. Bars depict the raw data (proportion of participants that overimitated). Horizontal lines with error bars depict the fitted model and its bootstrapped confidence limits.



4.4.2.2 Research Question 4

Model 7: Does overimitation at 18 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with cultural group or gender?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which only contained random effects (likelihood ratio test: $\chi 2$ = 23.54, df = 15, p = .073). Therefore, we can assume that in this reduced sample, neither the 'Approval Observe condition', 'overimitation', 'gender' or 'cultural group', nor their respective interactions, significantly influenced the proportion of tokens children gave.

Model 8: Does overimitation at 18 months predict sensitivity to normative expectations when children are 4 years old, controlling for cultural group or gender?

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model, which contained gender, cultural group and the random effects (likelihood ratio test: $\chi 2 = 2.15$, df = 3, $p = .541$). Therefore, we can assume that in this reduced sample, neither the 'Approval Observe condition' and 'overimitation', nor their interaction, significantly influenced the proportion of tokens children gave.

4.5 Part 3: Research Questions 5 and 6

Like joint attention, overimitation did not reliably relate to children's sensitivity to normative expectations later in childhood, which will be discussed in more detail in the Discussion. This raises the possibility, in line with Heyes' (2024) suggestions, that normativity, at this stage, may be more related to general developmental processes, instead of being built upon specific socio-cognitive mechanisms. Specifically, general cognition might capture aspects about early development that joint attention or overimitation do not, such as domain-general processes, and it might be these aspects of cognition which relate to the emergence of normativity in the age range we are testing. This chapter will now explore how general cognition in infancy relates to children's sensitivity to normative expectations in childhood. As this question was exploratory, no prior predictions were made

4.5.1 Methods

4.5.1.1 Participants

The sample consisted of participants who had valid data for the Normative Expectations task at 4 years (described in Chapter 1) and (i) valid data for the Bayley-III Cognitive Scale at 11 months. Out of the 78 longitudinal participants who had valid data for the Normative Expectations task, 64 participants had valid Bayley-III Cognitive Scale data at 11 months (28 girls; see Table 4.8 for details). More information about exclusion can be viewed in the Trial Validity section.

Table 4.8

The N and mean age of the participants with valid data for the Bayley Cognitive Scale in each cultural group.

Task	Uganda					UK				
	N	Girls	Boys	Mean Age[1]	SD[2]	N	Girls	Boys	Mean Age[1]	SD[2]
Bayley Cognitive Scale	26	10	16	0.94	0.03	38	18	20	0.93	0.02

[1]Age measured in years

[2]Standard deviation in years

4.5.1.2 Procedure and Materials

4.5.1.2.1 Bayley Cognitive Scale

The materials for the Bayley-III Cognitive Scale were used from the standardised Bayley-III administration kit, as detailed in the administration manual (Bayley, 2006). The scale comprised 91 items which assess various cognitive abilities that progress in difficulty. These cognitive abilities include, but are not limited to problem-solving skills, attention and memory. The average test re-rest reliability coefficient for the Cognitive Scale is .91, indicating excellent internal consistency (Bayley, 2006).

The Bayley-III Cognitive Scale was administered to each child by an experimenter at the 11-month timepoint. The experimenter sat across from the infant, at opposite sides of a table (see Figure 4.4 for the set-up in the UK and Uganda). The age of the infant determined which item to begin administration with. The infant was required to pass the first three consecutive items in order to establish their basal score, otherwise, the experimenter would go back to the previous start point and administer the first three consecutive items. This repeated until the infant passed the first three items. Once the basal level was established, the experimenter worked progressively through the items until the infant failed 5 items in a row, at which point administration was stopped.

Figure 4.4
Administration set-up of the Bayley Cognitive Scale in the UK (left) and Uganda (right).



4.5.1.3 Video Coding

4.5.1.3.1 Trial Validity

In total, 14 longitudinal participants were excluded from the analyses. Five participants (1 UK; 4 Ugandan) did not complete the Bayley-III Cognitive Scale at 11 months. The additional 9 participants (2 UK; 7 Ugandan) attempted the scale at 11 months but needed to be excluded from analyses because the task had to be stopped before 5 consecutive items had been failed (and as such, the data were invalid). This was either due to experimenter error (n = 1), infant distress (n = 4) or the infant not completing the items due to tiredness or distractibility (n = 4).

4.5.1.3.2 Bayley Scale Coding

Coding was completed in real time by a second experimenter who observed the Bayley-III being administered. Scoring followed the procedure of the Bayley-III administration manual (Bayley, 2006). Infants were scored 1 for every item they passed, and 0 for every item they did not pass. The Cognitive Score for each participant was calculated by summing the number of items passed with all items prior to the basal level (as it was assumed the infant would have passed all these items).

4.5.1.4 Statistical Approach

The statistical approach was the same as Part 1, and followed the outline set out in Chapter 2, Study 1.

4.5.1.5 Statistical Analysis

Research Question 5: Model 9

Is cognition at 11 months predicted by gender or cultural group?

A description of this model can be viewed in Table 4.9 (Model 9). See Appendix 4.3.1.1 for model stability parameters and details of the process followed to reach the final model. The sample consisted of 64 Bayley-III Cognitive Scores from 64 participants.

Research Question 6: Model 10

Does cognition at 11 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

Due to no differences reported between cultural groups in early cognition, cultural group was included as a fixed control variable (and as such, was included in the full and null model). A description of this model can be viewed in Table 4.9 (Model 10). See Appendix 4.3.1.2 for model stability parameters and details of the process followed to reach the final model. The sample consisted of 256 proportions of tokens given from 64 participants over 4 trials.

Table 4.9

Model construction, dispersion parameters and measures of collinearity between predictor variables for the final model exploring the relationship between cognition and (i) gender and cultural group (Model 9) and (ii) sensitivity to normative expectations (Model 10).

Model	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes	Dispersion Parameter	VIF[1]
9	Binomial GLM	Items correct| incorrect	Gender*Cultural Group	NA	NA	NA	0.59	Gender = 1.01; Cultural Group = 1.01
10	Binomial GLMM	Tokens given|kept in test trials	AO[3]*Cognitive Score *Gender	Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO[3]+ Cognitive Score) ||Experimenter on Trial[2] (Gender + AO[3]+ Cultural Group + Cognitive Score) ||Order of Conditions[2]	1.00	AO[3] = 1.00; Cognitive Score = 1.01; Gender = 1.01; Cultural Group =1.01

[1]VIF = Variance Inflation Factor

[2]The correlations between random slopes and intercepts were unidentifiable and therefore excluded from the model

[3]AO = Approval Observe condition

4.5.2 Results

4.5.2.1 Research Question 5

Model 9: Is cognition at 11 months predicted by gender or cultural group?

Overall, the full model did not explain significantly more variation in cognition at 11 months than the null model, which omitted the fixed effects of gender, cultural group and their interaction (likelihood ratio test: $\chi2 = 1.12$, df = 3, $p = .772$). Therefore, no further exploration of the coefficients was warranted, and we can assume gender and cultural group (or their interaction) did not influence cognition at this point in infancy.

4.5.2.2 Research Question 6

Model 10: Does cognition at 11 months predict sensitivity to normative expectations when children are 4 years old, and does this vary with gender?

Overall, the full model explained significantly more variation in the proportion of tokens given to the other child than the null model, which contained cultural group and the random effects (likelihood ratio test: $\chi2 = 17.42$, df = 7, $p = .015$). In the process of reaching the final model, the model was reduced to the 2-way interactions, given the non-significant 3-way interaction between 'Approval Observe condition', 'gender' and 'Cognitive Score' (estimate = 0.46, SE = 0.32, $p = .114$; Appendix 4.3.1.2). There was a significant 2-way interaction between 'Cognitive Score' and 'Approval Observe condition' (Table 4.10). Upon visualisation, this suggested that participants with lower cognitive scores at 11 months, gave more tokens away in the Approval Observe condition compared to the other 3 conditions (showing sensitivity to normative expectations) at 4 years old (Figure 4.5). There was also a significant 2-way interaction between 'Approval Observe condition' and 'gender' (Table 4.10), mirroring the results of Chapter 2, Study 1.

Table 4.10

Results of final Model 10 examining whether sensitivity to normative expectations at 4 years is predicted by cognition at 11 months in girls and boys, controlling for cultural group (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.91	0.62	-0.32	2.17	1.46	.145
AO[(1)]	-0.03	0.21	-0.59	0.45	-0.14	.891
Cognitive Score[(2)]	0.25	0.59	-0.97	1.47	0.42	.678
Gender[(3)]	-1.15	0.63	-2.44	0.15	-1.83	.067
Cultural Group[(4)]	-1.16	0.65	-2.51	0.13	-1.77	.077
AO*Cognitive Score	-0.45	0.16	-0.76	-0.11	-2.90	.004
AO*Gender	0.70	0.27	0.09	1.29	2.60	.009
Cognitive Score*Gender	0.35	0.71	-1.87	1.03	-0.50	.619

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

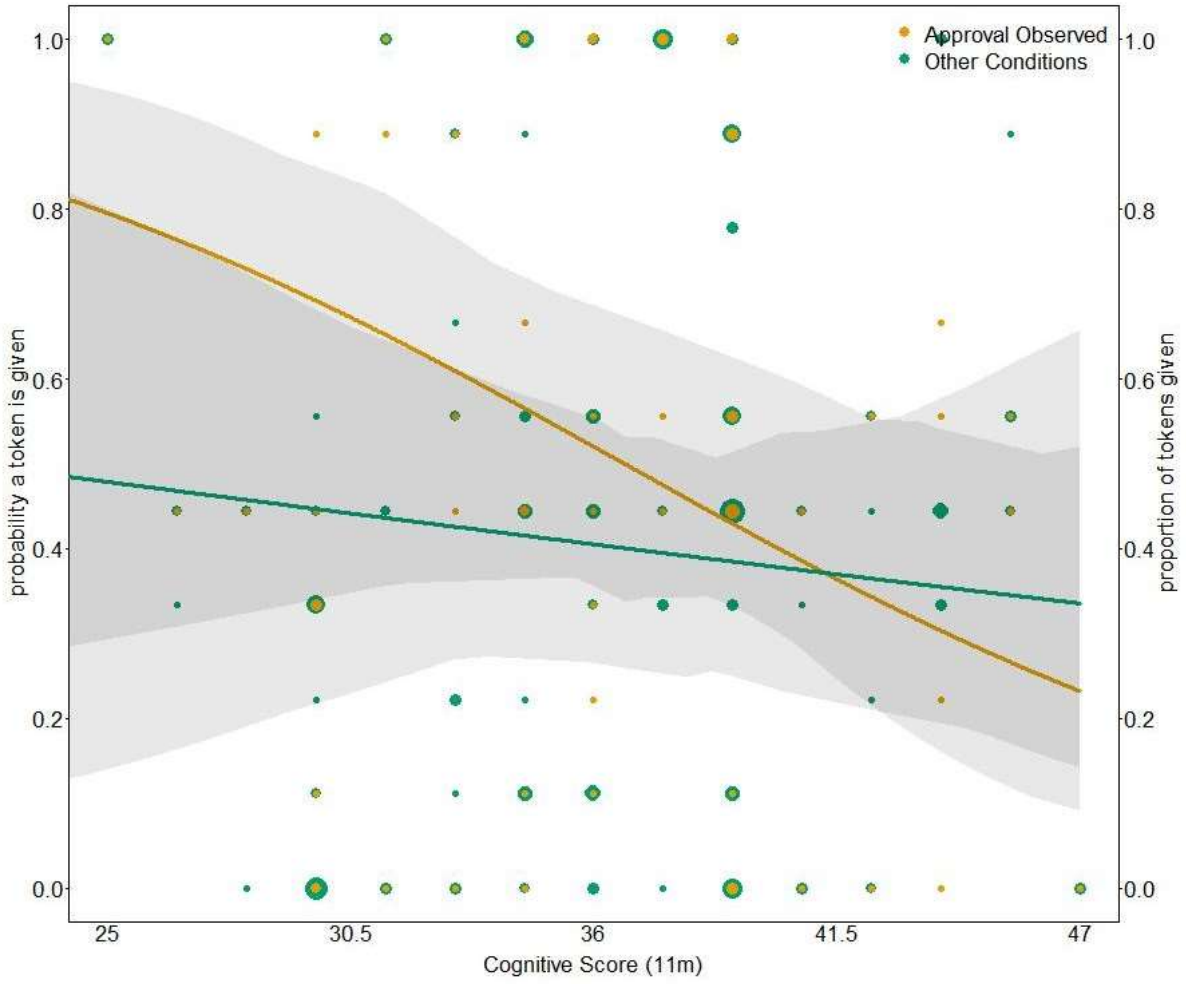
[(2)] z-scored (mean = 0; SD = 1)

[(3)] dummy coded with girls being the reference level

[(2)] dummy coded with the Ugandan sample being the reference level

Figure 4.5

Probability that a token is given in the Approval Observed condition and other conditions. Bubbles depict the raw data (proportion of tokens given) whereby bubble area is proportionate to the number of observations (range: 1 to 17). Lines depict the fitted model, and grey areas depict the bootstrapped confidence limits.



4.6 Discussion

This chapter aimed to explore how children's sensitivity to normative expectations at 4 years of age is related to their early cognition in infancy. In contrast to predictions based on theoretical models of the development of human normativity, neither joint attention (at 15 months) nor overimitation (at 18 months) predicted children's later normative prosociality (at 4 years). A measure of children's general cognitive abilities at 11 months, however, was related to children's sharing across approval and observation manipulations. Infants with lower Bayley-III Cognitive Scores were more sensitive to normative expectations, evidenced by them sharing more in the Approval Observe condition in comparison to other conditions, whilst infants with higher scores shared a similar proportion of tokens across all conditions.

Given the theoretical importance of joint attention for shared intentionality, and later normative self-governance (Tomasello, 2019), the lack of a relationship between joint attention and sensitivity to normative expectations is surprising. However, it is important to note that with a small (and capped) sample size and high proportions of zeros in the data, it is possible that our power to detect relatively small effect sizes may have been limited. We attempted to mitigate this concern by using a statistical model which only considered the interaction between the rate of joint attention events and the Approval Observe condition, which still provided no evidence of a relationship. The null findings might also have been contributed to by the measure we used. Although three levels of auditory joint attention events that varied in rigour were included to comprehensively test the relationship between joint attention events and normativity, the measures were only based on one dyadic task. Auditory joint attention has received less empirical attention compared to visual joint attention, and though theoretically there should be no conceptual differences between modes of joint attention, perhaps there was something specific about using a task that relied on auditory stimuli that limited the sensitivity of our measure. Future research should use a battery of joint attention tasks, including both auditory and visual modalities, to (i) test joint attention abilities more comprehensively and (ii) determine what differences there might be in the predictive value of joint attention elicited from different stimuli.

If replication with a larger sample and a more comprehensive battery of joint attention improves confidence in these null results, it would have implications for theoretical models of normativity. Joint attention appears to be foundational for many socio-cognitive skills in WEIRD samples, like Theory of Mind and language acquisition (e.g. Baldwin, 1995; Brandone & Stout, 2022; Charman et al., 2000; Kristen et al., 2011; Sodian & Kristen-Antonow, 2015). Yet, this result would suggest no direct link between joint attention abilities in infancy, and children's later normative prosociality, as posited by Tomasello's (2019) model. This may indicate that (i) joint attention is not as foundational for

normativity as previously thought, or (ii) joint attention may be indirectly related to normativity in a multi-step developmental pathway. For example, it is possible that joint attention might predict early cooperation, which then in turn could predict normative sensitivity. In this sense, joint attention does not have a direct influence on children's normative prosociality, but facilitates later skills that do, such as engaging in shared goals and commitments. Indeed, joint attention skills have been found to contribute to successful cooperation in 1- to 2-year-old children (Wu et al., 2013). Furthermore, responding to joint attention at 12 months has been found to predict the coordination of joint action at 18 months, by elevating responsiveness to nonverbal suggestions (Białek et al., 2013). Together, these studies support the idea that coordination of attention is important for achieving shared goals. However, the strongest evidence for a multi-step developmental pathway would come from a longitudinal study in which (i) joint attention, (ii) cooperation and (iii) normativity were measured over time.

Also contrary to our predictions, overimitation at 18 months did not predict children's later sensitivity to normative expectations. The relationship between imitative abilities and normative behaviour has not only been theorised (Carey, 2009; Tomasello & Carpenter, 2005), but has been empirically explored – Essler et al., (2023) found that 18-month-old infants' imitation abilities were related to their normative protest. Therefore, it is surprising that our results did not align with these accounts. As stated above, it may be possible that the effect sizes were too small to be detected with power afforded by the size of the longitudinal sample. The null result may also be a result of the limitations of the task. Firstly, the task was initially designed for use with 4-year-old children, so the apparatus might have been too advanced for testing at 18 months. Research suggests that the earliest signs of overimitation are evident from 18 months (Nielsen, 2006), although research typically focuses on children aged 3 to 5 (when overimitation is considered to be more prominent, e.g. Keupp, Behne & Rakoczy, 2013). Whilst it was important to test infants at a time that developmental variation would be high (assumed to be at the emergence), perhaps the measures need to be more sensitive when testing a younger age range. During this paradigm, infants only overimitated one out of the three possible causally irrelevant actions, meaning overimitation was reduced to a binary measure (i.e. whether infants overimitated or not). The action that was copied was a transitive action, meaning physical contact was made with the box). Intransitive actions are typically imitated less than transitive ones, both in young children (Clay et al., 2018) and infants (Kim et al., 2015). Perhaps including more transitive actions would have increased overimitation and subsequent sensitivity of the measure and variation in the data.

Individual variation was also dramatically reduced in the Ugandan infants – only three Ugandan infants overimitated, compared to 23 British infants. On one hand, this could represent a genuine

difference in children's early social cognition, perhaps reflecting variation in cultural learning techniques across societies (Legare & Harris, 2016). For example, perhaps children in the UK learn more through imitation, but Ugandan children learn more through third-party observation. On the other hand, it may be an artifact of using a WEIRD task in a non-WEIRD setting. For example, research conducted with this sample suggests that our Ugandan infants are less likely to engage in social object play in infancy than our UK sample (Marshall et al., *in prep.*), so interacting with the apparatus in a social context may have been highly unusual for them and prevented them from engaging. This highlights both (i) the importance of ensuring tasks are culturally relevant (Broesch et al., 2019) and (ii) the difficulty of untangling possible explanations for infant behaviour. Future research could use a battery of tasks to measure imitation abilities (similar to Essler et al., 2023), to provide insight into whether the cultural difference in overimitation that we observed reflects genuine variation in cultural learning, or issues with ecological validity.

If replication improves confidence in this null result, it will indicate that overimitation in infancy is not directly related to gaining social approval during normative sharing. One possibility is that overimitation might be more important for conventional norms, rather than social norms. Conventional norms are less concerned with approval and more describe how something is done within the social group. Essler et al., (2023) found that infants' imitation related (albeit indirectly) to normative protest when a puppet sorted toys in the wrong boxes. Critically, this normative protest is not related to social approval in the same way we have conceptualised it, as it is more related to a descriptive norm (i.e. rules about where coloured toys go), rather than an injunctive norm (i.e. reliant on expectations and approval about what should be done). Therefore, action imitation is more aligned with the content of descriptive norms; they are both usually concerned with what others are doing. During our Normative Expectation task, we aim for children to infer an injunctive norm about what the experimenter approves of, so maybe overimitation is less salient in this context.

Alternatively, perhaps overimitation does relate to normativity, but not in the ages we've explored. Most studies which point to a link between overimitation and normativity have been conducted with children that are at least 3 years of age (Clay et al., 2018; Keupp et al., 2013; Schmidt et al., 2016) – perhaps overimitation undergoes a transformative period in between infancy and childhood where the focus or motivation changes. It has been argued that overimitation is motivated by a variety of needs depending upon context (Over, 2020), and perhaps these needs and motivations change substantially as children progress through childhood. Maybe in infancy, overimitation functions as a way for infants to learn how to interact with the world around them, but later transitions into something more normative around 3 years of age as children also begin to understand more about social norms. To fully understand the structure and developmental trajectory of overimitation, and to be able to

understand if and when it relates to normativity, future research should aim to develop a measure of overimitation that can be successfully used across a range of ages and cultural groups (Over, 2020).

Unlike joint attention and overimitation, there was a relationship between performance on the Bayley-III Cognitive Scale and sensitivity to normative expectations, although perhaps in a surprising direction. Lower performance at 11 months was associated with greater sensitivity to approval whilst observed at 4 years of age. Specifically, children shared more in the Approval Observe condition, in comparison to the other conditions, when they scored lower on the scale, whereas children with higher scores differentiated less between the conditions. This suggests that children who were relatively less advanced in terms of cognitive development at 11 months of age were likely to give more in the Approval Observe condition compared to other conditions at age 4 years. Though this could suggest a link between more general cognitive abilities and the development of normative expectations, the direction of this finding is not consistent with what we might expect based on the theoretical model of Heyes (2024). Heyes (2024) argues that domain-general cognitive processes, which are not innately specialised for normativity, play a salient role in developing 'cognitive gadgets' that become specialised for social norm following through social and cultural experiences. This model ought to predict a positive relationship between domain-general cognitive abilities and children's normativity, so on the surface our results seem to contradict this model. One possibility is that the Bayley Cognitive Scale, as a whole, is not a direct measure of the specific domain-general cognitive abilities that are most relevant for the acquisition of the "cognitive gadget" for normativity, and so it is not an appropriate measure for testing the model's predictions. This would mean that further empirical tests of this model would require greater specification of which specific domain-general cognitive abilities are most relevant, so appropriate empirical measures for infants can be developed. However, whilst the direction of these results does not necessarily support Heyes' (2024) theory, these results do suggest that domain-general cognitive processes should not be overlooked in developing theoretical perspectives about the mechanisms that support normativity.

One interpretation of this result may be that those children with relatively slower cognitive development (at least in infancy) rely on more external social cues to guide behaviour, whereas those whose cognitive development has progressed more quickly tend to focus less on those external cues (i.e. be less attentive to the experimenter's approval or observation). This could be important if children with less advanced cognitive development deliberated *less* on the cues about the experimenters (their approval or observation). Research suggests that cognitive deliberation (i.e. thinking about something) decreases the effect of injunctive norms (Melnyk et al., 2011), which is likely driven by the fact that injunctive norms specifically have to do with external motivation (Cialdini & Goldstein, 2004), and focusing attention on that external motivation can cause some

individuals to think the opposite of the intended message. For instance, if individuals feel they're being manipulated or persuaded to act in a certain way to gain approval, it might have the opposite effect. Although this line of research is very different to what was studied in this thesis, it could provide some insight into our results. Perhaps children with less advanced cognitive development deliberated less on the experimenter's cue as a form of external motivation, and instead relied on the normative content of the cue to inform their sharing behaviour. In contrast, perhaps those with more advanced cognitive development could have engaged more thoughtfully with the cue, and questioned why they were being expected to behave in certain ways, leading to less differentiation between the conditions.

A different account might be that these results can be explained as reflecting children's developing intuitiveness to act fairly. Early in childhood, between ages 4 to 6 years, children's fair choices have been found to be slower and more deliberate, whereas selfish choices have a shorter reaction time (Chajes et al., 2022). In contrast, older children between ages 7 to 9 years showed no differences in reaction time between choices, and they were more likely to choose the fair distribution in comparison to the selfish one (Chajes et al., 2022). These results suggest that fairness becomes more intuitive as children age, and becomes less of a deliberated decision. Perhaps this process may be supported by general cognitive development, and children with more advanced cognition early in infancy are quicker to transition to intuitive fairness, whereas children whose cognition develops more slowly still carefully consider when to act fairly or more selfishly, resulting in a larger differentiation between sharing in the Approval Observe condition in comparison to other conditions.

One possible reason we did not find a relationship between joint attention or overimitation, like we did for cognitive score, might be related to the stability of the measures over time. Though, in general, measures of infant cognitive ability don't consistently predict cognitive measures later in childhood (Sternberg et al., 2001) recent research suggests that the Bayley-III Cognitive Scale does have predictive value for IQ over 4 years later (Klein-Radukic & Zmyj, 2023). This indicates that, for this measure at least, it could be possible that individual variation in cognitive development in infancy indeed tells us something about variation in cognitive development in childhood. This would be consistent with why cognitive score predicted normativity over 3 years later. Measures of joint attention and overimitation might not be as stable over time, however, this has not been empirically tested in the same way general cognition has, as most studies have only tested stability within infancy (e.g. Mundy et al., 2007), and not into childhood. Future research would benefit from longitudinally assessing joint attention and overimitation across development, to determine the stability and predictive value of these abilities.

Aside from the relationship with normativity, we can make some inferences about these abilities in infancy. Firstly, joint attention did not vary across subgroups of culture or gender, suggesting there may be some universality in its emergence. This aligns with previous cross-cultural studies which found no differences in joint attention skills (Callaghan et al., 2011; Childers, Vaughan & Burquest, 2007). No gender differences were evident, which contrasts previous research with Western samples suggesting that girls are more advanced in joint attention skills earlier in childhood (Mundy et al., 2007; Olafsen et al., 2006; Perkovich & Yoshida, 2024; Raver & Leadbeater, 1995; Saxon & Reilly, 1999). Perhaps if we had measured specific aspects of joint attention, like the length of each bout, or details about initiation, we could have observed different results. Secondly, cognition also did not vary between groups, contrasting previous cross-cultural studies which found differences in cognition (Godamunne et al., 2014; Vierhaus et al., 2010). Also, in contrast to prior research (Hua et al., 2019; Krogh & Væver, 2019; Suhonen et al., 2018), girls did not outperform boys in general cognitive ability. This suggests that differences or similarities in cognitive development might be bound to certain group comparisons. This also points to the suitability of the Bayley-III Cognitive Scale (despite its WEIRD nature) for some non-WEIRD groups, if an effort is made to ensure materials are culturally appropriate. Lastly, overimitation varied with cultural group -- the UK displayed more overimitation than Uganda. Although overimitation was observed much less in Uganda, it was not completely absent, supporting the idea that overimitation is present across diverse groups (Clegg & Legare, 2016; Nielson et al., 2014; Taniguchi & Sanefuji, 2017), but differs in terms of quantity or the context in which it is elicited (Clegg & Legare, 2016; Corriveau et al., 2017). Overall, it is interesting that two out of three infancy measures were not influenced by cultural group, and none of them were influenced by gender. This might represent a time in development in which cognition has not yet been greatly influenced by socialisation, or socialisation is similar for these facets of cognition in these two samples.

In conclusion, this chapter presents the first empirical exploration into the early mechanisms that support normative prosociality in childhood. Despite predictions based on existing theoretical models, neither joint attention (Tomasello, 2019) nor imitation abilities (Carey, 2009; Essler et al., 2023; Tomasello & Carpenter, 2005) in infancy predicted children's later sensitivity to normative expectations. A more general measure of early cognition played a surprising role, with lower cognitive scores in infancy being associated with higher sensitivity to the experimenter's approval and observation during children's sharing behaviour, indicating that domain-general processes may be important for the later development of normative behaviour. Results should be replicated with a larger sample size to improve confidence in the results, but if confirmed, they challenge pivotal accounts of normative emergence, and highlight the need for further research into how children develop sensitivity to normative expectations.

Chapter 5: General Discussion

This chapter explores the developmental, cognitive and cultural dimensions of children's early sensitivity to normative expectations. I start with a synthesis of key findings of Chapters 2 – 4, before discussing broader theoretical implications and outlining key considerations for future research.

5.1 Summary of Findings

The main aim of the current thesis was to explore the emergence of children's sensitivity to normative expectations, in girls and boys from two distinct cultural groups: the UK and Uganda. In Chapter 2, I found that 4-year-old boys, in comparison to 4-year-old girls, showed a pattern of sensitivity to the experimenter's approval when being observed. This was evidenced by boys demonstrating elevated levels of generosity when observed by an experimenter who approved of generosity but sharing statistically less than the girls when (i) the experimenter was not watching, or (ii) the experimenter was indifferent to generosity. Amongst alternative possibilities, these gender differences could have been driven by differences in socialisation relating to gender or differences in the underlying mechanisms which support the emergence of normativity. Chapter 2 also explored how the developmental trajectory of sensitivity to normative expectations changed between 5 and 11 years of age, but we found no evidence of children's sensitivity to normative expectations over this wider age range. Instead, children overall seemed to share more when the experimenter approved of generosity, but did not change their sensitivity to approval across different levels of observation. This could represent children's internalisation of normative information (i.e. relying less on external cues to motivate generosity) but could also have been driven by the methodological issues, such as the use of visible cameras and strong approval statements. Interestingly, and in contrast to our predictions, there were no cultural differences in children's sensitivity to normative expectations between the UK and Uganda between 4 and 11 years of age. This was surprising, given the differences of both groups' attitudes towards infant socialisation (Holden et al., 2022) and norm conformity (Uz, 2015).

Chapter 3 explored the relationship between 4-year-old children's concurrent social and cognitive skills and their sensitivity to normative expectations. I found that neither Theory of Mind, nor cognitive flexibility (an aspect of executive functioning) were related to children's developing sensitivity to approval and observation. These results were not aligned with our predictions or theoretical perspectives, which posited that children need to be able to represent others' mental states in order to consider what another believes is 'right' or 'wrong' (House, 2018), or have the ability to track normative information across scenarios, and resolve conflicts between their own desires and beliefs about what should be done (House & Tomasello, 2018).

Finally, Chapter 4 examined whether children's sensitivity to normative expectations at 4 years was predicted by various cognitive abilities in infancy. I found that engagement in joint attention events at 15 months, and overimitation at 18 months, did not predict children's later normative sharing. However, general cognitive ability at 11 months, as measured by the Bayley-III Cognitive Scale, did relate to children's sensitivity to normative expectations. Those with a lower cognitive score shared more in the condition in which the experimenter approved of generosity and was observing sharing behaviour, in comparison to the conditions in which experimenter approval or observation were not featured.

5.2 Theoretical Implications

It is important to note that the null findings from Chapters 3 and 4 may have been a result of the specific tasks used, or statistical power of the analyses. For example, as discussed throughout Chapters 3 and 4, it is possible that the relationship between children's early normativity and these cognitive abilities could have been detected more robustly through the use of alternative tasks, like exploring hot or cool executive functioning, or focusing on visual joint attention instead of auditory joint attention (which has received little empirical attention). Furthermore, the size of the longitudinal sample was predetermined and so could not be increased, which may have limited our power to detect relatively small effect sizes. Therefore, these results warrant further exploration. If, however, replication gives us confidence in these null findings, they have important implications for theoretical models of the human psychology for social norms. Tomasello (2019) argues that joint attention forms the foundation for later social-cognitive abilities, including normative thinking, so the lack of a relationship between early joint attention and children's later normativity is not directly consistent with a direct relationship. Also, Theory of Mind also does not relate to children's early sensitivity to social norms, which represents another step in Tomasello's (2019) model for how joint attention facilitates later normative self-governance. Whilst these results on their own do not undermine Tomasello's (2019) theory, they do highlight the need for further empirical support and indicate that future research should prepare for small effect sizes. Similarly, imitation has been theoretically proposed to facilitate representations of normative behaviour (Essler et al., 2023), but these null results, if supported with replication, would stand in contrast to this. Overall, if these null findings were replicated in future studies, it would suggest that these theoretical models regarding the ontogenetic roots of normativity may be mischaracterised, or missing important details about the nature of the relevant developmental processes. For example, if joint attention and imitation are related to the development of normativity, their causal role may be nonlinear or otherwise complex in a way that requires further specification by theoretical models. However, another possibility is that other cognitive mechanisms besides joint attention or imitation may be important to consider.

Given the finding that general cognitive ability in infancy relates to children's normative prosociality (despite the nature of the relationship being somewhat surprising), this suggest that at least some aspects of early infant cognition might help support the later development of normativity in early childhood. One possibility is that domain-general mechanisms might be a key driving force for how social cues are processed and subsequently guide normative behaviour. This resonates somewhat with a recent model of norm psychology, proposed by Heyes (2024), who suggests that the human propensity for social norms is based primarily on domain-general processes, which shape domain-specific explicit processes (like normative reasoning) through social interactions with group members. Heyes's (2024) theory would predict that earlier, more robust development of domain-general processes (like memory, attention or pattern recognition), paired with rich social interactions that signal which behaviours gain social approval, would lead to earlier, more robust normative behaviour. Whilst the direction of our results do not support this (i.e. a negative relationship between cognitive score in infancy and normativity in early childhood), our findings do support the concept that domain-general processes in infancy could have an effect on later normative behaviour in childhood, and question whether specialised mechanisms (like joint attention) play a unique role in the development of norm sensitivity.

Though, on the surface, these theoretical accounts do propose which mechanisms support the emergence of normativity and how they might complement each other, further specificity is likely necessary for empirically testing these ideas. For instance, does Tomasello's (2019) theory posit differences between joint attention in different mediums (i.e. auditory or visual joint attention)? Are there particular components of Theory of Mind or executive functioning which we assume to be more relevant than others? Which specific domain-general processes are the most important for processing social cues early in life? Answers to these questions would permit more coherent, grounded empirical tests of theory, and facilitate subsequent insight into the foundations of normativity.

With that said, a recent perspective by Schmidt, Vaish and Rakoczy (2024) questions the need to frame theoretical models of normativity as a stark choice between rich, nativist-like accounts (towards which Tomasello's (2019) theory leans) and more domain-general accounts (like Heyes's (2024) theory). The authors argue that there is ample space for middle-ground positions, in which infants may begin with cognitive structures richer than Hayes proposes, but leaner than Tomasello suggests. This may be a useful perspective in which to interpret our results regarding the developmental trajectory of normativity. The absence of a relationship between normative sharing and specialised mechanisms, like joint attention, is not consistent with the idea that specialised mechanisms provide the basis from which normativity is directly built. However, the negative relationship between early cognition and later normative sharing complicates the simplistic and straightforward domain-general

explanation. Together, these findings, if replicated, would support this idea that neither extreme account (strongly nativist or domain-general) fully captures the developmental processes at play, supporting calls for more middle-ground theories of norm psychology.

5.3 Normative Expectations Task

As discussed at multiple points throughout this thesis, there is scarce research exploring how children's behaviour relates to their beliefs about what others approve of, and this is especially so when considering the cognitive mechanisms which support the emergence of children's sensitivity to normative expectations. Upon reflection, one possible reason for why this has not been examined previously could be because of the difficulty associated with creating a working paradigm that teases apart children's underlying motivation for their sharing behaviour. The Normative Expectation task presented in this thesis underwent significant piloting before it was finalised (although in the UK only, as discussed later). All previous versions still considered the intersection between approval and observation, but used very different materials. For example, in one of the earliest versions, we explored children's tendency to cheat during a game in which they gained rewards by rolling a die. Children could only win if the die landed on one of the six sides (marked with a star), making the odds stacked against them. We were interested in whether children would cheat by telling the experimenter that they rolled the winning side across different levels of observation, and across different attitudes towards cheating stated by the experimenter. Despite a carefully planned design, the children found ways to cheat that fell into certain 'grey areas' of the rules, like manipulating the die roll or dropping instead of rolling the die. To try and prevent such manipulation of the task rules, next we developed a tower with multiple individual drawers, each containing a token. A 'colour rule' meant that the colour of the inside of the drawer determined whether the child should keep the token or give it to another child, and the free play rule meant that they child could decide whether to keep or give away each token, regardless of the colour of the drawer. Unfortunately, in this version of the paradigm the colour rule was too strongly adhered to and resulted in low levels of cheating and a lack of variation in behaviour across participants. This highlighted to us the importance of developing a paradigm that would generate individual variation in responses. This was achieved in a number of ways in the final task, such as reducing the strength of the approval prime and increasing the number of tokens children shared to 9, so that we could have confidence there would be enough variation in children's sharing behaviour over the 4 trials. This was especially important, because the longitudinal sample was capped in terms of sample size, so we attempted to maximise statistical power by improving measure sensitivity instead of sample size. Therefore, although designing a methodologically sound and appropriate task for the varied age ranges was challenging, these efforts underscore the importance of tackling complex conceptual questions through empirical work. Both careful piloting and purposeful, justified decisions that are guided by both theoretical relevance and

practical constraints, are essential in developing sensitive measures that can meaningfully capture these nuanced motivations of behaviours across childhood.

The development of this novel paradigm can also shed light on previous experimental designs that have explored children's sensitivity to observation. Past paradigms, although a useful starting point, have implicitly assumed that changes in children's prosocial behaviour when observed are driven by children's desire to gain reputational status (e.g. Engelmann, Hermann & Tomasello, 2012; Leimgruber et al., 2012). However, for children to be truly motivated by reputational gains, they would need to have an understanding of whether a behaviour is socially approved of, which has largely been ignored by previous observation research. In Chapter 2, I found that 4-year-old boys (though not 4-year-old girls) generally shared more when the experimenter approved of generosity. This highlights the significance of children's beliefs about what others' think they should do, which seemed to be a more consistent driver of prosociality during this period of childhood (age 4-11 years) than observation. Of course, we cannot infer what beliefs children had prior to the experiment – they could have been aligned with the approval statement, or very different, so it cannot be said with confidence that the children in this experiment had different beliefs about socially desired behaviour to those children in the previous research. What can be said though, is that children's beliefs about approval are important to account for, or at least consider, when exploring children's prosociality across different levels of observation. Tackling this issue is no easy feat (as alluded to during the discussion of piloting above), but to give an example, future research could include standardised approval (or disapproval) statements so that they knew children were entering the experiments with similar information about how others believe they should act.

5.4 Cultural Variation

Despite finding no evidence for cultural differences in how children adjusted their sharing behaviour in response to their social expectations, there were substantial cultural differences across most of the predictor tasks used throughout this thesis. Interestingly though, for the most part, these differences became more substantial as children aged. At age 4 years, both in the Theory of Mind task and our measure of cognitive flexibility, children in the UK outperformed children in Uganda at 4 years. In infancy, though Ugandan children overimitated significantly less than children in the UK at 18 months, for the other two infancy measures – joint attention and general cognition – children performed similarly across the two cultural groups. One reason for this may be that these measures early in infancy (up to 15 months) are capturing foundational elements of cognition, which have not yet been shaped by diverse experiences of socialisation or formal education settings. The tasks at older ages (i.e. when the children were 4 years) were more complex than the tasks in infancy, and could have placed a higher demand on other aspects of cognition, like language, emotion

understanding, or aspects of executive functioning – all which could be shaped by early life experience. For instance, it has been suggested that formal education in early childhood can scaffold executive functioning (Blair, 2016; Davidson et al., 2023). School attendance is mandatory in the UK by the time children reach 5 years, and all 3- to 4-year-olds are entitled to free early education prior to that. The education system is very different in Uganda; there are no free pre-school options, and although state primary schools are free to attend there is no mandatory starting age, and (due to the fact that attendance requires greater provision of scholastic material from parents than in the UK) in practice this means that most children in this rural area of Uganda enter formal educational settings later than children in the UK. Therefore, it may be possible that the elevated cognitive flexibility of our UK participants could be driven by early experiences with schooling, in comparison to our Ugandan sample, who may not have had as much experience in schools. Given that executive functioning has been found to relate to Theory of Mind (Carlson & Moses, 2011; Carlson, Moses & Breton, 2002; German & Hehman, 2006; Leslie et al., 2005; Marcovitch et al., 2015; Powell & Carey, 2017), this could also explain why the children in the UK generally outperformed Ugandan children on the Theory of Mind scale. Therefore, these findings could suggest that while many aspects of early cognition may be (somewhat) universal, socio-cultural factors may increasingly shape cognitive development as children age.

5.5 Key Considerations for Future Cross-Cultural Research

There is also a possibility that the cultural differences observed in many of the predictor tasks were artifacts of the study design. Firstly, the paradigms used were designed in WEIRD settings, by researchers in those settings, which might have impacted the construct validity of the tasks in non-WEIRD settings. The way instructions are given and received, stimuli are presented, and behaviour is observed and coded can be heavily influenced by cultural attitudes, such as those towards authority or appropriate ways to act in social interactions (Hruschka, 2020). Therefore, it is of upmost importance that researchers are actively ensuring the generalisability of methods when studying cultural settings different to their own (Broesch et al., 2021). One important first step, as argued by Broesch et al., (2021), is to collaborate with members of the community to ensure the materials are culturally appropriate – something we aimed to achieve during this study. For example, most of the Theory of Mind items were changed to be culturally relevant and familiar to children in our rural Ugandan sample (e.g. cat was replaced with goat, garage was replaced with house, biscuit was replaced with Mandazi). The original pictures of the DCCS (a boat and a rabbit; Zelazo, 2006) were replaced with images commonly recognised in both cultural groups (a car and a flower). Where possible, decisions were made collaboratively with research assistants in Uganda to incorporate the perspectives of those from the study community. All tasks were also translated into local languages and back translated to ensure the quality of instructions remained consistent across groups. These efforts represent key steps

to ensuring cultural validity in the task designs, and set a precedent for valuing contributions of local researchers, who perhaps do not have discipline-specific training, but have a considerable wealth of expertise regarding the cultural context (Urassa et al., 2021).

However, it is a possibility that these efforts were not substantial enough to create truly ecologically valid materials for our Ugandan participants. Firstly, whilst we aimed to ensure active communication was always open, honest and respectful, upon reflection, it could be argued that those exchanges about methodological considerations are themselves shaped by culturally specific norms around politeness and social harmony. These dynamics might constrain the extent to which participating members feel comfortable challenging assumptions or expressing disagreements. Therefore, what appears to be consensus in methodological alignment may, in fact, reflect a desire to maintain social cohesion. This appears to be an underrepresented concern in the literature, so future research should outline clear strategies to maximise the efficiency of communication when implementing cross-cultural research.

Secondly, due to the COVID pandemic, testing was unable to commence in Uganda until after data collection in the UK was about half-way complete. This meant that, whilst we attempted to make the stimuli as culturally appropriate as possible, there was relatively little flexibility in changing the actual measures to test the constructs we were interested in by the time preparation had started in Uganda. Instead of developing the tasks from the bottom-up in Uganda, we needed to adapt the WEIRD tasks for use with our Ugandan participants; there was flexibility in the content of the stimuli, but the design of the task needed to be consistent. For example, puppets were shown to the children in the Theory of Mind items, so we used racially matched puppets across cultural groups to increase relatability and changed the stimuli as previously discussed. However, it may be possible that the actual concept of a puppet is foreign to Ugandan children, who may have less experience with these kinds of toys or pretend or symbolic play more generally, which in turn, may impact their ability to attribute mental states to them. This highlights the importance of ensuring tasks are suitable for the cultural groups in which testing is taking place (Broesch et al., 2020).

Together, these points highlight the ongoing tension between ecological validity and comparability in cross-cultural research. On one hand, it is imperative to use methods which genuinely capture the constructs of interest across diverse groups, which may require tailoring methods to align with cultural differences in familiarity, socialisation or cultural attitudes. Yet, if these methods diverge too far between groups, the results may no longer be comparable. On the other hand, if comparability is prioritised and tasks are adapted for use in another culture, there is an elevated risk that the methods will not be ecologically valid for all the participants being tested, which could introduce confounds and lead to less meaningful interpretations. Whilst there is no concrete answer to this inevitable

tension, one way for researchers to navigate this would be to begin developing methods in the culture less familiar to the research team. Constructing methods in this way would allow researchers to judge the familiarity and relevance of those developed methods within their own culture. This is perhaps less likely to generate oversights than attempting to revise WEIRD methods for non-WEIRD contexts. For instance, fully developing the tasks in Uganda, ensuring that co-production was equitable (Urassa et al., 2021), and instructions and stimuli were rooted in local cultural contexts, and then assessing suitability for use in the UK might have struck a better balance of (i) comparability and (ii) ecological validity within both cultural groups. Furthermore, an extensive piloting period in Uganda would have allowed for more confidence in the construct validity of the measures. As stated, this was not possible due to multiple unforeseen challenges that arose, such as the COVID pandemic and subsequent Ebola outbreak in Uganda. If this had been an option, we might have had more insight into instruction framing or other culturally specific assumptions. For example, tokens were used throughout many of the games, like in the Normative Expectations task, which the participants could exchange for prizes at the end of the testing sessions. This is not a dissimilar approach to that used in other cross-cultural studies (e.g. House et al., 2020a) and training was undertaken with the tokens in both cultural groups to ensure children understood that more tokens meant more prizes. However, there was an implicit assumption that the exchange of tokens would be interpreted similarly by children across the cultural groups, and that they understood the tokens would be exchanged for something exciting. Instead, it might have been more ecologically valid to use a prize that did not need to be exchanged, like sweets, although this raises new challenges such as individualised preferences, satiety and health considerations. Another example could relate to the models who featured in the Normative Expectation priming videos. Whilst an effort was made to ensure children observed an unfamiliar model from their own ethnolinguistic background (meaning there were 3 versions of the prime in Uganda), it is unknown how children identified with those models. Given the models were all adult women, cultural attitudes and norms might differ towards observing models who are the same, or a different gender across the UK and Uganda. Co-constructing these materials from the ground up in Uganda would have allowed us to ask these important questions prior to the beginning of data collection. Therefore, one recommendation for future cross-cultural research would be to adopt a similar approach to the one described – begin developing methods in the culture less familiar to the research team and prioritise genuine collaboration. Where this is not possible, researchers should be transparent about where their research lies on the spectrum of ecological validity and comparability, and be explicit about any barriers they faced and how this might have influenced the results.

Despite these challenges, it is imperative that research continues to build up a more global picture of children's development, to address the persistent bias of testing, and relying on findings from WEIRD contexts (Nielsen et al., 2017). Besides the theoretical benefits, cross cultural research should also aim

to practically benefit those communities that take part. This can be achieved through various means, such as reimbursement for the families or individuals that take part, or at the community level. Participant compensation can be difficult to navigate, especially for projects that span multiple societies with varying socioeconomic status (Sansom et al., 2020). Debates are still ongoing regarding what constitutes 'fair' reimbursements, and whether these could be perceived as financial incentives (Samsom et al., 2020). In our study, families were reimbursed in culturally appropriate ways, and in Uganda were informed by discussions with our host institution and our Ugandan research assistants. In the UK, families received payment for participation, whereas Ugandan participants received household items (e.g. soap, oil, sugar or smearing oil) as advised. Children in both groups could exchange their tokens for culturally appropriate prizes, such as glowsticks, stickers or small toys in the UK, or balls, dolls or toy cars in Uganda. We also hosted celebration events for the participating families at the end of the project, to ensure the research was equally disseminated in both communities. These examples highlight how fair, appropriate and culturally relevant reimbursements can be achieved during cross-cultural research. The benefits of the project also reached further than the participating families. Lasting connections have been built between one of the local Ugandan primary schools and schools throughout the UK, which has led to fundraising and sponsorships for some Ugandan children. The project aided the restoration of the Ugandan school's library – books were collected from participating families in the UK and people around York to increase library stock and encourage children's education. These examples are representative of how research projects can, and should, aim to have meaningful, real-world benefits for the communities involved, particularly for those conducted in lower-income settings.

The final consideration recommended for future cross-cultural studies relates to site selection. In response to the call to conduct research beyond WEIRD societies (Henrich et al., 2010; Nielsen et al., 2017), many researchers have turned to including non-WEIRD societies for comparisons, but have not included theoretical justifications for why those specific sites have been chosen (Broesch et al., 2021). In order to avoid binning these non-WEIRD samples together in a 'West versus the rest' approach, it is important for researchers to consider specific cultural contexts, and clearly justify why those samples have been chosen (Broesch et al., 2021). As discussed in Chapter 1, the UK and Uganda were chosen due to cultural variation in infant socialisation and attitudes towards norm compliance, ensuring that the research was situated in relevant cultural contexts. Despite this consideration, there were still some aspects that were unknown which could have further aided interpretation of the results. For instance, little was known about specific socialisation of different genders in these samples. Knowing this might have clarified why the young boys were showing sensitivity to normative expectations, unlike the girls, who were more generous overall, and whether socialisation was having a similar influence across both samples. When exploring children's sensitivity to social

norms in the future, it would also be interesting to compare sites with stronger or weaker expectations about child behaviour. For example, prior research has found that Sami parents (an indigenous group living in Scandinavia, Finland and Russia) encourage more child autonomy, and impose fewer rules and regulations on their children than do Norwegian parents (Javo et al., 2004). Despite being in geographically similar areas, these two sites show variation in expectations about how young children should behave, which could influence children's normative development. This example helps to demonstrate how future research should be considering theory when selecting sites, instead of choosing tokenistic sites, simply because they fall into the category of non-WEIRD societies.

5.6 Key Considerations for Future Social Norms Research

Whilst this thesis contributes novel insights into how children's prosocial behaviour is shaped by normative expectations, there are other dimensions of social norm development which were outside the scope of this thesis. Two particularly relevant areas for future research exploring social norm development are (i) the structure and salience of reference networks, and (ii) the role of emotions in motivating or regulating norm following behaviour.

Reference networks refer to the range of people that an individual cares about when making particular decisions (Bicchieri, 2016). Both an individual's empirical and normative expectations are based on people within their reference network, and there is a lot of individual variation in who might be included (Bicchieri, 2016). For example, a devout member of a religious community might only care about beliefs of fellow believers, instead of their neighbours, while others may rely more on friends or co-workers. This variation makes it challenging to empirically isolate reference networks in experimental work, but there have been some successful attempts to isolate members of these, such as using salient peer members in schools to shift collective norms around bullying (Paluck & Shepherd, 2012).

Little attention has been devoted to understanding the developmental trajectory of children's reference networks, yet it is an important factor to consider to understand variability in response to normative information, as some research suggests that children do not respond to all social cues equally. For example, infants and young children have been found to selectively imitate the actions of ingroup members more than outgroup members (Buttelman et al., 2013; Howard et al., 2014), although ingroup preferences might be less pronounced in children from socially disadvantaged groups (e.g. Shutts et al., 2011), suggesting diversity in how young children evaluate group membership. Three- and four-year-old children have shown a preference to learn from adults who others have preferentially attended to, rather than an adult who was ignored (Chudek et al., 2012), suggesting subtle socio-contextual cues might impact how individuals are perceived early in childhood.

Furthermore, early theoretical accounts propose that young children attend to the direction of their parents, while older children become increasingly influenced by the beliefs of their peers (Utech & Hoving, 1969). Although explicit tests of this are scarce, age might be an important determinant of who children choose to preferentially attend to. Based on this line of research, we could infer that there may have been variability for how children processed the information from the experimenter (and model) in the Normative Expectations task. This could be explored by systematically varying characteristics of the experimenters and models involved, such as gender-matching or using mothers or peers across different age groups. Observing children's prosocial behaviour across various manipulations would allow insight into the trajectory of how reference networks might change across development.

Lastly, emotions are undoubtedly a large part of social interactions, especially normative ones, which were not explored during this thesis. Emotions can serve two main functions in relation to social norms: signals and motivators. Firstly, emotions can act as strong signals of social approval or disapproval. For example, when participants are shown groups responding with anger to a norm transgression, they can accurately infer what the social norm was (Hareli et al., 2013; 2015), suggesting that individuals are able to reconstruct normative expectations from witnessing certain emotional responses to behaviour. Secondly, emotions play a motivating role in sustaining social norm following. Embarrassment and self-conscious emotions, like shame or guilt, can deter individuals from violating social norms as they wish to avoid negative emotions that may arise as a result of their behaviour (Staller & Petta, 2001). Some argue that guilt arises when there are deviations from personal norms or standards, whereas shame is elicited when social norms are violated and there is focus on how others will evaluate behaviour (Packard & Shultz, 2023). Also, it is not just negative emotions which can promote compliance – complex emotions like pride are thought to make us feel positive about achieving both personally, and socially valued behaviours (Packard & Shultz, 2023). Therefore, at least in adults, emotions are thought of as enforcers of normative behaviour.

In recent years, research has turned to exploring how emotions can motivate children's prosocial behaviour. For example, from 3 years of age, children have been reported to engage in more reparative behaviour in a condition where they caused a harmful outcome in comparison to when someone else was responsible for the outcome, or when the outcome was not harmful (Vaish, Carpenter & Tomasello, 2016), suggesting they felt guilt because of their actions and sought to repair any damage to the relationship. Similarly, 3-year-old children's arousal (as measured by pupil dilation) lowered if (i) the harm was caused and repaired by someone else, or (ii) they were able to personally repair accidental harm they had caused. However, arousal remained high if they could not

personally resolve the situation if they had caused the harm (Hepach, Vaish & Tomasello, 2017), which again, signals that children's prosocial behaviour can be motivated by feelings of guilt from violating personal norms.

Although guilt relates more to violations of moral norms, work has also been conducted on other complex emotions have been linked to reputation management in children. For example, in an experiment by Hepach et al., (2023), 2-year-old children showed similar emotional valence (as measured by postural elevation) across situations they personally helped or witnessed helping by someone else, but 5-year-olds displayed more positive emotions when they were the ones to provide the help. This suggests that the underlying motivation to help is different across development: early in childhood, motivation to help may be more intrinsically rewarded, whereas later, children's helping becomes more strategically motivated. The older children also displayed more positive emotions when observed by two experimenters, in comparison to when helping was anonymous. Together, these studies suggest that emotions play an important role in children's prosocial, and likely normative, behaviour. Future work should more directly examine the emotional processes the underlie children's responses to social norms and how these processes develop across childhood.

5.7 Conclusion

In this thesis, I have contributed to knowledge in the field of developmental psychology by exploring (i) how children's prosocial behaviour may be shaped by their developing sensitivity to normative expectations, and (ii) how this phenomenon varies across gender and cultural contexts (the UK and Uganda). I developed a novel paradigm designed to assess children's emerging responsiveness to approval and observation, offering new insight into the motivations underlying children's sharing behaviour. I also tested key theoretical accounts of the mechanisms supporting the emergence of normative expectations, using both longitudinal and cross-sectional methods. In doing so, this work sheds light on the psychological foundations of social norm acquisition. Importantly, this research underscores the value of cross-cultural approaches in developmental science. By implementing comparable measures of children's social and cognitive abilities across diverse communities, this thesis highlights cultural variability, and the need for methodological sensitivity and collaboration when studying populations beyond WEIRD contexts. Taken together, this work contributes to a growing understanding of how children's sensitivity to social norms is characterised in childhood. Continuing to explore children's normative sensitivities across a broader range of demographic, contextual and emotional dimensions will be vital for building a more comprehensive theory of the origins of human prosociality.

Appendix 2: Supplementary Material for Chapter 2

2.1 Appendices for Study 1

2.1.1 Materials and Procedure

2.1.1.1 Difference in the Protocol Between the Two Sites

Due to pandemic travel restrictions, initial experimental design was limited to the UK. Piloting suggested that the UK sample were very capable of extracting information from the video during the video comprehension phase, and no further comprehension questions regarding the Generous Prime were added during the priming phase. However, later piloting in Uganda suggested that children found it more challenging to explicitly report information during the visual comprehension phase, and so in Uganda participants were asked two comprehension questions: First, they were asked "What happened in the video?", and if they did not explicitly state that the model gave more tokens to the other person, they were asked a binary choice question, "Did she give more to herself, or more to the other person?". If participants got this question incorrect, they were asked to watch the video again and the same questions were repeated (this repetition was omitted for 3 participants due to experimenter error, but they were included in the analysis). As testing had already begun in the UK, it was not possible to modify the UK protocol with these changes. Instead, in the UK we added comprehension questions to the end of the protocol: if UK children failed the video comprehension phase (which was infrequent; N = 11), then after the last trial they were shown the Generous Prime and were asked the same two comprehension questions described above, to check their understanding of the video (this was missed for 3 participants due to experimenter error, but these participants were included in the analysis). This allowed for consistent protocols within each site, even though protocols were slightly different across sites.

After the video had concluded in the UK, or after the comprehension questions in Uganda, the experimenter then presented the participants with the 'Generous Prime': this was information about whether they (i) approved of how the model played the game in the video ('Approval' condition) or (ii) were indifferent to how the model played the game in the video ('Indifference' condition). In the 'Approval' condition, the experimenter stated, "I like the way that person played the game", whereas in the 'Indifference condition, the experimenter stated "I don't care if people play like that person. I like it when people play however they want to". These statements were carefully constructed based on piloting data in the UK to retain information about approval, whilst encouraging substantial individual variation for analyses. Whether the 'Approval' or 'Indifference' condition came first (i.e. during the First or second Priming phase) was counterbalanced across participants.

2.1.1.2 Deviations from Protocol

Order of Trials: The procedure was designed so that observed and unobserved trials were the same within children but counterbalanced across children, except 7 children who received different orders of observation within trials due to experimenter error. As the order of trials was included as a random effect, these trials were still included in the analysis.

Checkerboard Video Comprehension: Due to experimenter error, one participant in Uganda was not asked the follow-up questions for the checkerboard comprehension.

Generous Prime Comprehension: Due to experimenter error, one Ugandan girl was not shown the Generous Prime video again or asked follow-up questions in the Approval or Indifference Condition after not responding to the first set of questions, one Ugandan boy was not shown the Generous Prime video again or asked follow-up questions in the Approval Condition after not responding to the first set of questions, and another Ugandan boy in the Approval Condition was not asked the Generous Prime questions after watching the Generous Prime video for the second time (as he incorrectly answering the first set of questions). These participants were nonetheless included in the main analysis given that the questions were designed to determine whether children needed further instruction, and not as exclusion criteria. However, in the additional analyses that only comprised children who passed the Generous Prime comprehension (RQ3; see Appendix 2.1.3), the latter three participants were also excluded. Furthermore, two UK participants were not shown the Generous Prime video again or asked the comprehension questions despite failing the checkerboard comprehension, and another participant in the UK was not shown the Generous Prime video at the end of testing and instead was only asked the comprehension question again (which was answered correctly). As above, these participants were excluded in the additional analyses of children who passed the Generous Prime comprehension (Appendix 2.1.3).

2.1.2 Statistical Approach and Analyses

2.1.2.1 Additional Details Regarding the Process of Reaching the Final Model: Model 1 (RQ1)

To explore whether baseline sharing behaviour in the Training DG varied with cultural group (Uganda/UK) or gender (Girls/Boys) between 4 and 5 years, a GLM was run with the proportion of tokens given and kept in the Training DG as the outcome variable, and cultural group and gender (as well as the interaction between them) included as predictor variables in the model (Model A in Table A.2.1). However, as this model was overdispersed (dispersion parameter = 3.22), a beta binomial model was run instead (Model B in Table A.2.1). This beta binomial model was then checked for overdispersion, and found to still be overdispersed (dispersion parameter = 3.15), which suggested there was still a higher-than-expected amount of variability in the data. Therefore, as described in the Statistical Approach section, an Observation Level Random Effect was included in the model to account for some unexplained variation in the data (Gelman & Hill, 2007). Put simply, this meant including a factor which numbered each row of data in the model as a random intercept, and running a GLMM as described in the Statistical Approach section. The GLMM model was underdispersed (dispersion parameter = 0.31), indicating tests were now more conservative (Model C in Table A.2.1). The model stability parameters for Model 1 can be viewed in Table A.2.2).

The full-null model comparison revealed a significant effect (likelihood ratio test: $\chi 2 = 13.91$, df = 3, $p = .003$), indicating that cultural group or gender were influencing the proportion of tokens given in the Training DG. The interaction between cultural group and gender was non-significant (see Table A.2.3), so a reduced model was run which consisted of cultural group and gender as main effects (see Table 7 in the Results section of Chapter 2; Study 1).

Table A. 2.1

Model construction prior to the final Model 1 presented in the Statistical Analysis section (Table 2.2).

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Random Intercepts	Random slopes	Dispersion Parameter
A	Binomial GLM	Proportion of Tokens Given and Kept in Training DG	Cultural Group*Gender	NA	NA	3.22
B	Binomial GLMM	Proportion of Tokens Given and Kept in Training DG	Cultural Group*Gender	NA	NA	3.14
C (final)	Binomial GLMM	Proportion of Tokens Given and Kept in Training DG	Cultural Group*Gender	Observation Level Random Effect (obsID)	NA	0.31

Table A. 2.2

Model stability parameters for Model 1 (in Table 2.2), exploring 4-year-old children's baseline sharing in the Training DG in the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Variable	Original	Min	Max
Intercept	0.40	0.31	0.51
Cultural Group[(1)]	-1.59	-1.70	-1.51
Gender[(2)]	-0.80	-0.92	-0.69
Cultural Group*Gender	1.09	0.96	1.21
obsID@(Intercept)@NA	1.60	1.55	1.63

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

Table A. 2.3

Results of full Model 1 exploring 4-year-old children's baseline sharing in the Training DG in the UK and Uganda (estimates and standard errors, together with confidence limits, and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0..40	0.32	-0.27	1.05	1.25	.213
Cultural Group[(1)]	-1.59	0.45	-2.47	-0.65	-3.57	<.001
Gender[(2)]	-0.80	0.47	-1.77	0.18	-1.71	.088
Cultural group* Gender	1.09	0.65	-0.30	2.47	1.69	.091

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

2.1.2.2 Additional Details Regarding the Process of Reaching the Final Model: Model 2 (RQ2)

To analyse how the proportion of tokens given to the unfamiliar child varied with condition (Approval or Indifference), observation (Observed or Not Observed) and cultural group (UK and Uganda), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these three factors (all considered as categorical predictors), and their 3-way interaction were included as fixed effects into the model. Additionally, gender was controlled for by including it as a fixed effect in both full and null models. As all the correlations between the random slopes were estimated to have coefficients of exactly 1, which indicated that it was likely not identifiable (Matuschek et al., 2017), we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model which led to an only minor reduction in model fit (log-likelihoods, model with correlation parameters included: -1054.59, df = 35; model without correlation parameters: -1057.81, df = 19). The process of reaching the final model discussed in the Results section can be viewed in Table A.2.4. The model stability parameters for Model 2 can be viewed in Table A.2.5.

The model did not show a significant three-way interaction between 'approval', 'observe' and 'cultural group' (see Table A6), so a reduced model was run without this interaction (see Model B in Table A.2.4). The reduced model results suggested there was a trend towards the interaction of 'observe' and 'approval' (Table A.2.7), and visualisation of the data (collapsing across the UK and Uganda) indicates a trend towards the pattern of normative expectations as previously defined (see Figure 2.3 in Chapter 2). With that said, given that this interaction was not significant, a final reduced model was run with only the main effects (see Model C in Table A.2.4), which is primarily discussed in the Results Section (RQ2).

Table A. 2.4

Model Reduction Prior to the Final Model 2 Presented in the Results Section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	Observation *Approval *Cultural Group	Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
B (reduced)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	Observation*Approval + Observation*Cultural Group + Approval*Cultural Group	Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
C (final)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	Observation + Approval + Cultural Group	Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]

[1]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

Table A. 2.5

Model stability parameters for Model 2 (in Table 2.2), exploring 4-year-old children's sensitivity to normative expectations, controlling for gender, in the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.13	0.94	1.39
Observe[(1)]	-0.30	-0.58	0.06
Approval[(2)]	-0.27	-0.36	-0.11
Cultural Group[(3)]	-1.79	-2.06	-1.49
Gender[(4)]	-0.76	-1.01	-0.46
Observe*Approval	0.52	0.25	0.79
Observe*Cultural Group	0.45	-0.04	0.72
Approval*Cultural Group	0.55	0.26	0.63
Observe*Approval*Cultural Group	-0.45	-0.82	-0.08
Pariticpant ID@Intercept@NA	2.12	2.02	2.47
Experimenter on Trial@Intercept@NA	0.07	0.00	0.15
Experimenter on Trial@Gender@NA	0.00	0.00	0.44
Experimenter on Trial@Approval@NA	0.11	0.00	0.30
Experimenter on Trial@Observe@NA	0.35	0.20	0.55
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order @Gender@NA	0.00	0.00	0.26
Condition Order @Cultural Group@NA	0.00	0.00	0.03

Condition Order @Approval@NA	0.31	0.11	0.38
Condition Order @Observe@NA	0.10	0.00	0.17

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

Table A. 2.6

Results of full Model 2 (Model A in Table A.2.4), exploring 4-year-old children's sensitivity to normative expectations, controlling for gender, in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	p
Intercept	1.13	0.37	0.39	1.83	3.03	.002
Observe(1)	-0.39	0.24	-0.77	0.15	-1.22	.221
Approval(2)	-0.27	0.22	-0.67	0.16	-1.21	.225
Cultural Group(3)	-1.79	0.43	-2.70	-0.96	-4.16	<.001
Gender(4)	-0.76	0.39	-1.53	-0.04	-1.96	.050
Observe*Approval	0.52	0.23	0.06	0.95	2.27	.023
Observe*Cultural Group	0.45	0.30	-0.17	1.10	1.50	.133
Approval*Cultural Group	0.55	0.23	0.08	1.04	2.33	.020
Observe*Approval* Cultural Group	-0.45	0.30	-1.06	0.16	-1.50	.132

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

Table A. 2.7

Results of reduced Model 2 (Model B in Table A.2.4), exploring 4-year-old children's sensitivity to normative expectations, controlling for gender, in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	p
Intercept	1.07	0.37	0.32	1.86	2.87	.004
Observe(1)	-0.16	0.23	-0.60	0.32	-0.71	.478
Approval(2)	-0.14	0.20	-0.53	0.27	-0.67	.505
Cultural Group(3)	-1.68	0.42	-2.52	-0.78	-3.96	<.001
Gender(4)	-0.77	0.39	-1.55	-0.01	-1.96	.050
Observe*Approval	0.26	0.15	-0.07	0.56	1.73	.084
Observe*Cultural Group	0.22	0.26	-0.31	0.78	0.85	.393
Approval*Cultural Group	0.32	0.18	-0.03	0.69	1.78	.075

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

2.1.2.3 Additional Details Regarding the Process of Reaching the Final Model: Model 3 (RQ3)

To analyse how the proportion of tokens given to the unfamiliar child varied with condition (Approval or Indifference), observation (Observed or Not Observed), cultural group (UK and Uganda) and gender (girls and boys), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these four factors (all considered as categorical predictors), and their interaction included as fixed effects into the model. We included 'approval', 'observe' and 'gender' as random slopes within 'experimenter on trial', and we included 'approval', 'observe', 'gender' and 'cultural group' as random slopes within the 'order of conditions'. As all the correlations between random slopes within 'order of condition' were estimated to have coefficients of exactly 1, we removed the correlation parameters of these random slopes from the model which led to an only minor reduction in model fit (log-likelihoods, model with correlation parameters included: -1034.86, df = 42; model without correlation parameters: -1037.41, df = 32). The process of reaching the final model discussed in the Results section can be viewed in Table A.2.8. The model stability parameters for Model 3 can be viewed in Table A.2.9.

The 4-way interaction between 'observe', 'approval', 'cultural group' and 'gender' was non-significant (Table A.2.10), so a reduced model that only comprised the 3-way interactions was run (Model B in Table A.2.8), which is discussed in the Results Section (RQ3).

Table A. 2.8

Model Reduction Prior to the Final Model 3 Presented in the Results Section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	Observation *Approval *Cultural Group*Gender	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
B (final)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	Observation*Approval*Gender + Observation*Cultural Group*Gender + Approval*Cultural Group*Gender + Observe*Approval*Gender	NA	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]

[1]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

Table A. 2.9

Model stability parameters for Model 3 (in Table 2.2), exploring 4-year-old children's sensitivity to normative expectations, controlling for gender, in the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.12	0.71	1.48
Observe[(1)]	0.12	-0.24	0.51
Approval[(2)]	-0.52	-0.70	-0.09
Cultural Group[(3)]	-1.77	-2.59	-0.91
Gender[(4)]	-0.75	-1.32	0.04
Observe*Approval	0.22	-0.24	0.67
Observe*Cultural Group	0.08	-0.34	0.46
Approval*Cultural Group	0.92	0.49	1.32
Observe*Gender	-1.13	-2.31	-0.35
Cultural Group *Gender	0.82	0.10	1.72
Approval*Gender	-0.08	-1.25	0.95
Observe*Approval*Cultural Group	-0.50	-0.94	-0.13
Observe*Cultural Group*Gender	0.78	-0.19	1.84
Approval*Cultural Group*Gender	1.01	0.22	2.19
Observe*Approval*Gender	-1.15	-2.08	-0.46
Observe*Approval*Gender*Cultural Group	0.07	-0.98	1.02
Pariticpant ID@Intercept@NA	2.18	2.01	2.58

Term	Original	Min	Max
Experimenter on Trial@Intercept@NA	0.09	0.00	0.16
Experimenter on Trial@Gender@NA	0.08	0.01	0.35
Experimenter on Trial@Approval@NA	0.29	0.20	0.42
Experimenter on Trial@Observe@NA	0.41	0.25	0.58
Experimenter on Trial@Intercept@Gender	0.32	-1.00	1.00
Experimenter on Trial@Intercept@Approval	0.27	-1.00	0.98
Experimenter on Trial@Intercept@Observe	0.41	-0.97	1.00
Experimenter on Trial@Gender@Approval	1.00	-0.99	1.00
Experimenter on Trial@Gender@Observe	-0.73	-1.00	0.99
Experimenter on Trial@Approval@Observe	-0.76	-1.00	-0.31
Condition Order@Intercept@NA	0.00	0.00	0.05
Condition Order@Gender@NA	0.00	0.00	0.17
Condition Order@Cultural Group@NA	0.00	0.00	0.10
Condition Order@Approval@NA	0.36	0.22	0.50
Condition Order@Observe@NA	0.14	0.01	0.19

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

Table A. 2.10

Results of full Model 2 (Model A in Table 2.A.8), exploring 4-yeaer-old children's sensitivity to normative expectations, controlling for gender, in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	1.12	0.43	0.27	1.98	2.61	.009
Observe[(1)]	0.12	0.30	-0.50	0.76	0.42	.677
Approval[(2)]	-0.52	0.30	-1.13	0.10	-1.73	.083
Cultural Group[(3)]	-1.77	0.57	-2.93	-0.61	-3.08	.002
Gender[(4)]	-0.75	0.63	-2.07	0.44	-1.20	.230
Observe*Approval	0.22	0.29	-0.41	0.86	0.77	.438
Observe*Cultural Group	0.08	0.37	-0.74	0.88	0.22	.826
Approval*Cultural Group	0.92	0.33	0.23	1.64	2.77	.006
Observe*Gender	-1.13	0.34	-1.82	-0.46	-3.29	.001
Cultural Group *Gender	0.82	0.36	0.17	1.51	2.32	.021
Approval*Gender	-0.08	0.85	-1.75	1.52	-0.10	.921
Observe*Approval*Cultural Group	-0.50	0.39	-1.39	0.46	-1.28	.201
Observe*Cultural Group*Gender	0.78	0.49	-0.12	1.77	1.59	.113
Approval*Cultural Group*Gender	1.01	0.44	0.11	1.97	2.27	.023
Observe*Approval*Gender	-1.15	0.45	-2.11	-0.24	-2.54	.011
Observe*Approval*Gender*Cultural Group	0.07	0.63	-1.31	1.35	0.11	.913

[(1)] dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

2.1.3 Additional Analyses with the 4-Year-Old Participants who Passed the Generous Prime Comprehension

2.1.3.1 Excluding Participants Who Did Not Pass the Norm Comprehension

Despite the comprehension questions not being intended for exclusion purposes, it was clear that the Ugandan children struggled with some aspects of the task. Therefore, in order to check the robustness of the results, the analysis for Model 3 (i.e. the most complex model which included all other lower-order interactions, as well as the interaction between approval, observation, cultural group and gender) was rerun with only participants who passed the Generous Prime comprehension on the first attempt for both the Approval and Indifference conditions. This reduced the sample to 64 UK children (36 girls) and 35 Ugandan children (20 girls), meaning that the sample consisted of 396 proportions of tokens given and kept from 99 participants. The same statistical approach was used as described in the Statistical Analysis section for Model 3 (RQ3).

2.1.3.2 Results of Model 3 Excluding Participants Who Did Not Pass the Norm Comprehension

Model 3: Is sensitivity to normative expectations evident in 4- to 5-year-old children, and does it vary across cultural groups and genders?

Overall, Model 3 explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi 2 = 60.79$, df = 15, $p<.001$). The 4-way interaction between 'observe', 'approval', 'cultural group' and 'gender' was non-significant (estimate = -.02, SE = 0.83, $p = .798$), so a reduced model that only comprised the 3-way interactions was run (with the same structure as Model B in Table A.2.8; see Table A.2.11 for model estimates). The significant interaction between 'observe', 'approval' and 'gender' remained (estimate = 0.89, SE = 0.36, $p = .013$; Table A.2.11), suggesting that the finding that boys show the pattern of sensitivity to normative expectations, unlike girls, is robust and is unlikely to be driven by children not understanding the normative information presented to them. The significant interactions between 'observe', 'cultural group' and 'gender', and 'approval', 'cultural group' and 'gender' also remained significant (Table A.2.11), further suggesting the gender differences observed in Ugandan children's responses to observation and approval are robust, and again, not related to whether the children answered the Generous Prime questions incorrectly.

Table A. 2.11

Results of reduced Model 2 (Model B in Table 2.A.8), exploring 4-year-old children's sensitivity to normative expectations, exploring gender, only including participants who passed the Generous Prime comprehension on the first attempt (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	1.00	0.48	0.53	2.19	2.09	.037
Observe[(1)]	-0.06	0.36	-0.69	0.67	-0.16	.875
Approval[(2)]	-0.51	0.30	-0.79	-0.24	-1.69	.092
Cultural Group[(3)]	-1.64	0.59	-2.73	-1.29	-2.80	.005
Gender[(4)]	-0.47	0.77	-1.78	-0.28	-0.61	.542
Observe*Approval	0.43	0.33	0.14	1.07	1.28	.201
Observe*Cultural Group	0.28	0.44	-0.69	0.87	0.65	.518
Approval*Cultural Group	0.93	0.33	0.32	1.32	2.78	.005
Observe*Gender	-1.39	0.37	-1.66	-0.99	-3.74	<.001
Cultural Group *Gender	-0.16	0.95	-0.57	0.91	-0.17	.868
Approval*Gender	1.47	0.48	1.22	1.95	3.07	.002
Observe*Approval*Cultural Group	-0.73	0.38	-1.53	-0.30	-1.93	.054
Observe*Cultural Group*Gender	1.25	0.39	0.71	1.67	3.17	.002
Approval*Cultural Group*Gender	-1.82	0.51	-2.08	-1.35	-3.58	<.001
Observe*Approval*Gender	0.89	0.36	0.44	1.09	2.49	.013

[(1)] dummy coded with the Unobserved condition being the reference level

[(2)] dummy coded with the Indifference condition being the reference level

[(3)] dummy coded with the Ugandan sample being the reference level

[(4)] dummy coded with girls being the reference level

2.1.4 Additional Statistical Analyses Included in the Pre-registered Analysis Plan

2.1.4.1 Measures of Additional Pre-Registered Analyses

Binary coding of Sensitivity to Observation: In order to assess whether children were sensitive to the observation manipulation, in a similar fashion to previous research, the tokens children gave across both unobserved conditions were summed, and subtracted from the sum of the tokens children gave across both observed conditions. Positive scores were coded as 'Yes' (i.e. children showed sensitivity to observation), whereas negative scores or scores of zero were coded as 'No' (i.e. children did not show sensitivity to observation).

Binary coding of Sensitivity to Expectation: Following the suggestion that sensitivity to expectation should also be considered when understanding children's prosocial behaviour, sensitivity to expectation was also binary coded; the tokens children gave across both 'Indifference' conditions were summed, and subtracted from the sum of the tokens children gave across both 'Approval' conditions. Positive scores were coded as 'Yes' (i.e. children showed sensitivity to the expectation of the experimenter), whereas negative scores or scores of zero were coded as 'No' (i.e. children did not show sensitivity to the expectation of the experimenter).

Binary coding of Sensitivity to Normative Expectations: Lastly, to explore whether children showed sensitivity to normative expectations, one final score was calculated per participant. As previously mentioned, sensitivity to normative expectations was defined as having a positive effect of observation in the Approval condition, in comparison to the Indifference condition (i.e. children should be more sensitive to observation when they are being watched by someone who has an expectation for how they ought to behave). Therefore, a sensitivity to normative expectations (NE Score) was calculated for each participant by subtracting the difference between the Approval and Indifference Observed tokens given, from the difference between the tokens given in the Approval Observed and Indifference Unobserved conditions.

2.1.4.2 Statistical Approach of Additional Pre-Registered Analyses

<u>Models A1, A2 and A3: Is sensitivity to observation, approval or normative expectations common between 4 to 5 years and does it vary across cultural groups?</u>

Three separate models were run to explore children's i) sensitivity to observation (Model A1), ii) sensitivity to approval (Model A2) or iii) sensitivity to normative expectations (i.e. the integration of both sensitivity to observation and expectations; Model A3). To analyse how children's sensitivity varied with cultural group (UK and Uganda), a logistic Generalized Linear Mixed Model (Baayen,

2008) was used, with this factor and gender (both considered as categorical predictors) were included as fixed effects into the model. Random intercepts were the experimenters on a given trial, and the order of conditions. Random slopes were included if deemed to be theoretically identifiable, using a function written by Roger Mundry. By including all possible random slopes, we maximise the precision of fixed effect estimates and keep the type 1 error rate at 5% (Schielzeth and Forstmeier, 2009; Barr et al., 2013). We included 'gender' as a random slope within 'experimenters on trial', and included 'cultural group' as a random slope within the 'order of conditions'. The random slope parameters were manually dummy coded and centred. Initially this was a maximal model (Barr et al., 2013), including parameters for the correlations among random intercepts and slopes. However, as all correlation parameters were estimated to have a coefficient of exactly 1 in all 3 models, which indicated that they were likely not identifiable (Matuschek et al., 2017), we removed the correlation parameters of the random slopes from the models which led to an only minor reduction in model fit (the maximum difference was in Model A3; log-likelihoods, model with correlation parameters included: -78.47, df = 9; model without correlation parameters: -78.86, df = 7). To test the significance of the predictor variables, the fit of the full model was compared to the fit of a null model comprising only the control variable of 'gender' and the random effects using a likelihood ratio test (Dobson, 2002). This avoids cryptic multiple testing (Forstmeier and Schielzeth, 2011). For each model the respective sample consisted of 129 sensitivity scores from 129 participants.

2.1.4.3 Results of Additional Pre-Registered Analyses

Model A1: Is sensitivity to observation common between 4 to 5 years and does it vary across cultural groups?

Descriptively, 30/68 children in the UK and 17/61 children in Uganda showed sensitivity to observation, meaning they shared more when they were being observed by an experimenter, compared to when they were not being observed. Overall, Model A1 did not explain significantly more variation in sensitivity to observation scores than the null model (likelihood ratio test: $\chi 2 = 3.039$, df = 1, $p = .081$). This suggests that cultural group does not significantly impact how sensitive children are to observation.

Model A2: Is sensitivity to expectations common between 4 to 5 years and does it vary across cultural groups?

Descriptively, 31/68 children in the UK and 14/61 children in Uganda showed sensitivity to expectation, meaning they shared more when the experimenter had expressed an approval of generous sharing behaviour, compared to when the experimenter was indifferent to how children shared.

Overall, Model A2 explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi2 = 6.487$, df = 1, $p = .011$). The model showed a significant effect of 'cultural group' (model results can be seen in Table A.2.12), which suggested that children in the UK were more sensitive to approval than children in Uganda.

Table A. 2.12

Results of Model A2 exploring 4-year-old children's sensitivity to approval (estimates and standard errors, together with confidence limits, significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
intercept	-1.19	0.35	-2.16	-0.61	-3.38	.001
Cultural Group[(1)]	1.03	0.39	0.30	2.00	2.65	.008
Gender[(2)]	-0.04	0.38	-0.87	0.74	-0.09	.926

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

Model A3: Is sensitivity to normative expectations common between 4 to 5 years and does it vary across cultural groups?

Descriptively, 21/68 children in the UK and 18/61 children in Uganda showed sensitivity to normative expectations, meaning there was a more positive effect of observation when the experimenter approved of generosity, compared to the effect of observation when the experimenter was indifferent to generosity. Overall, Model A3 did not explain significantly more variation in sensitivity to observation scores than the null model (likelihood ratio test: $\chi2 = 0.03$, df = 1, $p = .863$). This suggests that cultural group does not significantly impact how sensitive children are to normative expectations.

2.2 Appendices for Study 2

2.2.1 Materials and Procedure

2.2.1.1 Deviations from Protocol

After incorrectly answering the first questions about the Generous Prime, two children were supposed to watch the Generous Prime video again and be asked the same questions to understand whether they comprehended what happened in the video. However, these children were asked the questions again without watching the video for a second time due to experimenter error. These participants were nonetheless included in the analyses given that the questions were designed to determine whether children needed further instruction, and not as exclusion criteria. However, in the additional analysis which only consisted of children who passed the Generous Prime on the first attempt (see Appendix 2.2.3), these children were excluded.

2.2.2 Statistical Approach and Analyses

2.2.2.1 Additional Details Regarding the Process of Reaching the Final Model: Model 4 (RQ4)

To analyse how the proportion of tokens given to the unfamiliar child in the Training DG varied with cultural group (UK or Uganda), gender (girls or boys) and age (covariate), a logistic Generalized Linear Model with binomial error structure and logit link function was used (McCullagh & Nelder, 1989), with these three predictors, and their interaction included as fixed effects into the model. The model stability parameters for the full model can be viewed in Table A.2.14.

The process of reaching the final model discussed in the Results section can be viewed in Table A.2.13. The full model did not show a significant three-way interaction between 'cultural group', 'age' and 'gender' (see Table A.2.15), so a reduced model was run just including the 2-way interactions (see Model B in Table A.2.13). The reduced model results suggested there was a significant interaction between 'cultural group' and 'age' (Table A.2.16), so this interaction was retained and a final reduced model was run without the non-significant interactions (see Model C in Table A.2.13), which is discussed in the Results Section (RQ4).

Table A. 2.13

Model reduction prior to the final Model 4 presented in the Results section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLM	Proportion of Tokens Given and Kept in Familiarity Trial	Cultural Group*Gender*Age	NA	NA	NA
B (reduced)	Binomial GLM	Proportion of Tokens Given and Kept in Familiarity Trial	Cultural Group*Gender + Cultural Group*Age + Gender*Age	NA	NA	NA
C (final)	Binomial GLM	Proportion of Tokens Given and Kept in Familiarity Trial	Cultural Group*Age + Gender	NA	NA	NA

Table A. 2.14

Model stability parameters for Model 4 (Model A in Table A.2.13), exploring 5-11-year-old children's baseline sharing in the Training DG in the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	-0.34	-0.38	-0.27
Cultural Group[(1)]	-0.07	-0.14	-0.02
Gender[(2)]	-0.08	-0.15	-0.03
Age[(3)]	0.01	-0.07	0.05
Cultural Group*Gender	0.03	-0.03	0.10
Cultural Group*Age	0.21	0.13	0.29
Age*Gender	-0.10	-0.18	-0.02
Cultural Group*Gender*Age	0.17	0.02	0.25

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

[(3)] measured in days; z-transformed

Table A. 2.15

Results of full Model 4 (Model A in Table A.2.13) exploring 5- to 11-year-old children's baseline sharing in the Training DG in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	-0.34	0.09	-0.52	-0.16	-3.69	<.001
Cultural Group[(1)]	-0.07	0.13	-0.33	0.19	-0.56	.579
Gender[(2)]	-0.08	0.13	-0.33	0.17	-0.62	.534
Age[(3)]	0.01	0.09	-0.16	0.19	0.13	.895
Cultural group*Gender	0.03	0.19	-0.35	0.49	0.14	.887
Cultural Group*Age	0.21	0.13	-0.04	0.47	1.63	.102
Gender*Age	-0.19	0.13	-0.35	0.14	-0.82	.415
Cultural Group*Gender*Age	0.17	0.19	-0.21	0.55	0.85	.394

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

[(3)] measured in days; z-transformed

Table A. 2.16

Results of reduced Model 4 (Model B in Table A.2.13) exploring 5- to 11-year-old children's baseline sharing in the Training DG in the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	-0.33	0.09	-0.51	-0.15	-3.64	<.001
Cultural Group[(1)]	-0.08	0.13	-0.34	0.18	-0.59	.555
Gender[(2)]	-0.08	0.13	-0.34	0.17	-0.65	.513
Age[(3)]	-0.02	0.08	-0.18	0.13	-0.30	.766
Cultural group*Gender	0.03	0.19	-0.34	0.40	0.15	.877
Cultural Group*Age	0.29	0.10	0.10	0.48	2.97	.003
Gender*Age	-0.03	0.10	-0.22	0.15	-0.34	.731

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

[(3)] measured in days; z-transformed

2.2.2.2 Additional Details Regarding the Process of Reaching the Final Model: Model 5 (RQ5)

To analyse how the proportion of tokens given to the unfamiliar child varied with approval (Approval or Indifference), observation (Observed or Not Observed), cultural group (UK and Uganda) and age, a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these factors (all considered as categorical predictors, except age which was considered as a continuous covariate), and their interactions were included as fixed effects into the model. Additionally, gender was controlled for by including it as a fixed effect in both full and null models. As all the correlations between the random slopes of 'condition order' were estimated to have coefficients of exactly 1, we removed the correlation parameters of these random slopes from the model which led to an only minor reduction in model fit (log-likelihoods, model with correlation parameters included: -1595.95, df = 54; model without correlation parameters: -1600.32, df = 39). Model stability parameters for the full model can be viewed in Table A.2.18.

The full model (Model A in Table A.2.17) revealed that the 4-way interaction between 'approval', 'observe', 'cultural group' and 'age' was non-significant (Table A.2.19), so a reduced model was run without this interaction (Model B in Table A.2.17). There were also no significant 3-way interactions observed (Table A.2.20), so these were removed from the model, leaving only the 2-way interactions (Model C in Table A.2.13). There was only a significant interaction between 'observe' and 'age' (Table A.2.21), so the remaining non-significant interactions were removed so that the final model only consisted of this significant 2-way interaction and the main effects (Model D in Table A.2.13), which is discussed in the Results Section (RQ5).

Table A. 2.17
Model reduction prior to the final Model 5 presented in the Results section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Control Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLMM	Tokens given|kept in test trials	Observe*Approval *Cultural Group*Age	Gender	Participant ID, Experimenter on Trial, Order of Conditions	Age + Gender + Observe + Approval|Experimenter on Trial Age + Gender + Cultural Group + Observe + Approval||Order of Conditions[1]
B (reduced)	Binomial GLMM	Tokens given|kept in test trials	Observe*Approval* *Cultural Group + Observe*Approval*Age + Approval*Cultural Group*Age + Observe*Cultural Group*Age	Gender	Participant ID, Experimenter on Trial, Order of Conditions	Age + Gender + Observe + Approval|Experimenter on Trial Age + Gender + Cultural Group + Observe + Approval||Order of Conditions[1]
C (reduced)	Binomial GLMM	Tokens given|kept in test trials	Approval*Observe + Approval*Cultural Group + Observe*Cultural Group + Approval*Age + Observe*Age + Cultural Group*Age	Gender	Participant ID, Experimenter on Trial, Order of Conditions	Age + Gender + Observe + Approval|Experimenter on Trial Age + Gender + Cultural Group + Observe + Approval||Order of Conditions[1]
D (final)	Binomial GLMM	Tokens given|kept in test trials	Observe*Age + Approval + Cultural Group	Gender	Participant ID, Experimenter on Trial, Order of Conditions	Age + Gender + Observe + Approval|Experimenter on Trial Age + Gender + Cultural Group + Observe + Approval||Order of Conditions[1]

[1]The correlations between random slopes and intercepts were unidentifiable and therefore excluded from the model

Table A. 2.18

Model stability parameters for Model 5 (Model A in Table A.2.17), exploring 5- to 11-year-old children's sensitivity to normative expectations, controlling for gender, in the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.84	0.62	1.17
Approval(1)	0.27	0.02	0.58
Observe(2)	0.07	-0.13	0.21
Cultural Group(3)	-1.15	-1.51	-0.94
Age(4)	0.26	0.06	0.42
Gender(5)	-0.16	-0.26	-0.01
Approval*Observe	0.10	0.04	0.16
Approval*Cultural Group	-0.06	-0.43	0.20
Observe*Cultural Group	0.25	0.11	0.47
Approval*Age	-0.04	-0.17	0.25
Observe*Age	-0.11	-0.18	-0.03
Cultural Group*Age	0.21	-0.07	0.45
Approval*Observe*Cultural Group	-0.14	-0.20	-0.05
Approval*Observe*Age	0.07	-0.05	0.15
Approval*Cultural Group*Age	0.07	-0.21	0.23
Observe*Cultural Group*Age	-0.16	-0.26	-0.08
Approval*Observe*Cultural Group*Age	0.17	0.06	0.31

Term	Original	Min	Max
Participant ID@Intercept@NA	0.99	0.89	1.05
Experimenter on Trial@Intercept@NA	0.14	0.00	0.18
Experimenter on Trial@Age@NA	0.02	0.00	0.15
Experimenter on Trial@Gender@NA	0.18	0.07	0.35
Experimenter on Trial@Approval@NA	0.23	0.02	0.29
Experimenter on Trial@Observe@NA	0.19	0.05	0.25
Experimenter on Trial@Intercept@Age	-0.03	-1.00	1.00
Experimenter on Trial@Intercept@Gender	-0.12	-1.00	0.72
Experimenter on Trial@Intercept@Approval	-0.87	-1.00	0.91
Experimenter on Trial@Intercept@Observe	-0.96	-1.00	-0.60
Experimenter on Trial@Age@Gender	-0.99	-1.00	1.00
Experimenter on Trial@Age@Approval	-0.47	-0.98	0.99
Experimenter on Trial@Age@Observe	-0.25	-1.00	1.00
Experimenter on Trial@Gender@Approval	0.60	-0.28	1.00
Experimenter on Trial@Gender@Observe	0.39	-0.64	1.00
Experimenter on Trial@Approval@Observe	0.97	-0.83	1.00
Condition Order@Intercept@NA	0.00	0.00	0.00
Condition Order@Age@NA	0.05	0.00	0.14
Condition Order@Gender@NA	0.00	0.00	0.01

Term	Original	Min	Max
Condition Order@Cultural Group@NA	0.00	0.00	0.19
Condition Order@Approval@NA	0.00	0.00	0.13
Condition Order@Observe@NA	0.36	0.27	0.41

(1) dummy coded with the Indifference condition being the reference level

(2) dummy coded with the Unobserved condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

Table A. 2.19

Results of full Model 5 (Model A in Table A.2.17) exploring 5- to 11-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.84	0.28	0.30	1.35	3.00	.003
Observe[(1)]	0.07	0.24	-0.37	0.54	0.29	.773
Approval[(2)]	0.27	0.19	-0.07	0.59	1.40	.163
Cultural Group[(3)]	-1.15	0.33	-1.77	-0.55	-3.50	<.001
Age[(4)]	0.26	0.13	0.03	0.52	1.98	.049
Gender[(5)]	-0.16	0.17	-0.49	0.15	-0.97	.335
Approval*Observe	0.10	0.15	-0.19	0.37	0.66	.511
Approval*Cultural Group	-0.06	0.25	-0.59	0.39	-0.22	.827
Observe*Cultural Group	0.25	0.22	-0.13	0.67	1.17	.240
Approval*Age	-0.04	0.11	-0.26	0.17	-0.34	.735
Observe*Age	-0.11	0.11	-0.31	0.09	-1.00	.315
Cultural Group*Age	0.21	0.19	-0.15	0.56	1.12	.263
Approval*Observe*Cultural Group	-0.14	0.21	-0.56	0.26	-0.68	.499
Approval*Observe*Age	0.07	0.15	-0.24	0.36	0.48	.633
Approval*Cultural Group*Age	0.07	0.16	-0.23	0.39	0.46	.649
Observe*Cultural Group*Age	-0.17	0.15	-0.48	0.15	-1.08	.278

Approval*Observe*Cultural Group*Age	0.17	0.21	-0.26	0.63	0.80	.427

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

Table A. 2.20

Results of reduced Model 5 (Model B in Table A.2.17) exploring 5- to 11-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.83	0.28	0.28	1.37	2.96	.003
Observe(1)	0.07	0.24	-0.36	0.56	0.31	.758
Approval(2)	0.27	0.19	-0.08	0.61	1.40	.161
Cultural Group(3)	-1.15	0.33	-1.75	-0.50	-3.45	.001
Age(4)	0.28	0.13	0.04	0.53	2.18	.029
Gender(5)	-0.16	0.17	-0.47	0.17	-0.96	.336
Approval*Observe	0.10	0.15	-0.21	0.41	0.66	.509
Observe*Cultural Group	0.25	0.22	-0.17	0.67	1.14	.256
Approval*Cultural Group	-0.06	0.26	-0.51	0.41	-0.23	.817
Observe*Age	-0.15	0.09	-0.32	0.02	-1.59	.113
Approval*Age	-0.08	0.10	-0.25	0.10	-0.84	.401
Cultural Group*Age	0.16	0.18	-0.19	0.52	0.92	.357

Approval*Observe*Cultural Group	-0.15	0.21	-0.58	0.29	-0.69	.489
Approval*Observe*Age	0.15	0.11	-0.08	0.36	1.44	.150
Observe*Cultural Group*Age	-0.08	0.11	-0.29	0.12	-0.74	.460
Approval*Cultural Group*Age	0.16	0.11	-0.06	0.39	1.42	.156

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

Table A. 2.21

Results of reduced Model 5 (Model C in Table A.2.17) exploring 5- to 11-year-old children's sensitivity to normative expectations, controlling for gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.83	0.27	0.35	1.34	3.07	.002
Observe[(1)]	0.10	0.23	-0.34	0.53	0.46	.648
Approval[(2)]	0.29	0.17	-0.02	0.61	1.67	.095
Cultural Group[(3)]	-1.13	0.31	-1.69	-0.54	-3.60	<.001
Age[(4)]	0.23	0.12	0.00	0.46	1.98	.047
Gender[(5)]	-0.16	0.17	-0.48	0.16	-0.97	.333
Observe*Approval	0.02	0.11	-0.18	0.22	0.17	.864
Observe*Cultural Group	0.19	0.19	-0.18	0.51	1/02	.307
Observe*Age	-0.11	0.05	-0.21	0.00	-2.05	.040
Approval*Cultural Group	-0.11	0.22	-0.49	0.26	-0.52	.604
Approval*Age	0.08	0.06	-0.02	0.20	1.49	.136
Cultural Group*Age	0.19	0.15	-0.11	0.50	1.26	.208

[(1)] dummy coded with the Unobserved condition being the reference level

[(2)] dummy coded with the Indifference condition being the reference level

[(3)] dummy coded with the Ugandan sample being the reference level

[(4)] measured in days; z-transformed

[(5)] dummy coded with girls being the reference level

2.2.2.3 Additional Details Regarding the Process of Reaching the Final Model 6 (RQ6)

To analyse how the proportion of tokens given to the unfamiliar child varied with condition (Approval or Indifference), observation (Observed or Not Observed), cultural group (UK and Uganda), age and gender (boy or girl), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these factors (all considered as categorical predictors, except age as a covariate), and their interactions were included as fixed effects into the model. As all the correlations between random slopes within 'order of condition' were estimated to have coefficients of exactly 1, we removed the correlation parameters of these random slopes from the model which led to an only minor reduction in model fit (log-likelihoods, model with correlation parameters included: -1584.97, df = 69; model without correlation parameters: 1589.12, df = 54). The process of reaching the final model can be viewed in Table A.2.22. Model stability parameters for the full model can be viewed in Table A.2.23.

The full model (Model A in Table A.2.22) revealed that the 5-way interaction between 'approval', 'observe', 'cultural group', 'age' and 'gender' was non-significant (Table A.2.24), so a reduced model was run without this interaction (Model B in Table A.2.22). The reduced model revealed a significant interaction between 'observe', 'cultural group', 'age' and 'gender', but no significant 4-way interactions with 'approval' (see Table A.2.25), so another reduced model was run without these non-significant interactions (Model C in Table A.2.22). This model further revealed no significant 3-way interactions with the 'approval' term (see Table A.2.26), so the model was further reduced (Model D in Table A.2.22). There were also no significant 2-way interactions including the 'approval' term (see Table A.2.27), so any interactions including 'approval' were removed from the model (Model E in Table A.2.22), which is discussed in the Results section (RQ6).

Table A. 2.22

Model reduction prior to the final Model 5 presented in the Results section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLMM	Tokens given|kept in test trials	Observe *Approval*Cultural Group*Age*Gender	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)||Experimenter on Trial[2] (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
B (reduced)	Binomial GLMM	Tokens given|kept in test trials	Observe*Approval*Cultural Group*Gender + Observe*Approval*Age*Gender + Observe*Cultural Group*Age*Gender + Approval*Cultural Group*Age*Gender + Observe*Approval*Cultural Group*Age	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)||Experimenter on Trial[2] (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
C (reduced)	Binomial GLMM	Tokens given|kept in test trials	Observe*Cultural Group*Age*Gender + Observe*Approval*Cultural Group + Observe*Approval*Gender + Approval*Cultural Group*Gender + Observe*Approval*Age + Approval*Cultural Group*Age + Approval*Gender*Age	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)||Experimenter on Trial[2] (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Random Intercepts	Random Slopes
D (reduced)	Binomial GLMM	Tokens given|kept in test trials	Observe*Cultural Group*Age*Gender + Observe*Approval + Approval*Cultural Group + Approval*Gender + Approval*Age	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)||Experimenter on Trial[2] (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
E (final)	Binomial GLMM	Tokens given|kept in test trials	Observe*Cultural Group*Age*Gender + Approval	Participant ID, Experimenter on Trial, Order of Conditions	(Age + Gender + Observe + Approval)||Experimenter on Trial[2] (Age + Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]

[1]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

Table A. 2.23

Model stability parameters for Model 6 (Model A in Table A.2.22) examining 5- to 11-year-old children's sensitivity to normative expectations, exploring gender, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.03	0.82	1.28
Approval(1)	0.05	-0.17	0.32
Observe(2)	0.12	-0.11	0.41
Cultural Group(3)	-1.47	-1.72	-1.25
Age(4)	0.15	-0.13	0.36
Gender(5)	-0.43	-0.82	-0.19
Approval*Observe	0.39	0.15	0.80
Approval*Cultural Group	0.20	0.12	0.41
Observe*Cultural Group	0.10	-0.19	0.32
Approval*Age	0.14	0.06	0.33
Observe*Age	-0.19	-0.31	-0.04
Cultural Group*Age	0.23	-0.04	0.52
Approval*Gender	0.25	0.12	1.10
Observe*Gender	-0.12	-0.46	0.35
Cultural Group*Gender	0.48	0.22	0.85
Age*Gender	0.16	0.11	0.50
Approval*Observe*Cultural Group	-0.33	-0.73	-0.03

Term	Original	Min	Max
Approval*Observe*Age	-0.14	-0.28	-0.03
Approval*Cultural Group*Age	-0.05	-0.24	0.09
Observe*Cultural Group*Age	0.07	-0.07	0.23
Approval*Observe*Gender	-0.51	-1.20	-0.07
Approval*Cultural Group*Gender	-0.29	-1.11	-0.20
Observe*Cultural Group*Gender	0.36	-0.11	0.68
Approval*Age*Gender	-0.29	-0.70	0.03
Observe*Age*Gender	0.17	-0.10	0.29
Cultural Group*Age*Gender	0.02	-0.33	0.47
Approval*Observe*Cultural Group*Age	0.24	0.06	0.37
Approval*Observe*Cultural Group*Gender	0.28	-0.13	0.99
Approval*Observe*Age*Gender	0.31	-0.06	0.65
Approval*Cultural Group*Age*Gender	0.16	-0.17	0.59
Observe*Cultural Group*Age*Gender	-0.51	-0.70	-0.24
Approval*Observe*Cultural Group*Age*Gender	0.02	-0.33	0.41
Participant ID@Intercept@NA	0.99	0.89	1.04
Experimenter on Trial@Intercept@NA	0.13	0.00	0.18
Experimenter on Trial@Age@NA	0.01	0.00	0.18

Term	Original	Min	Max
Experimenter on Trial@Gender@NA	0.11	0.01	0.24
Experimenter on Trial@Approval@NA	0.18	0.02	0.24
Experimenter on Trial@Observe@NA	0.18	0.06	0.25
Experimenter on Trial@Intercept@Age	1.00	-1.00	1.00
Experimenter on Trial@Intercept@Gender	0.25	-1.00	0.97
Experimenter on Trial@Intercept@Approval	-0.97	-1.00	0.98
Experimenter on Trial@Intercept@Observe	-1.00	-1.00	-0.54
Experimenter on Trial@Age@Gender	0.25	-1.00	0.99
Experimenter on Trial@Age@Approval	-0.97	-1.00	1.00
Experimenter on Trial@Age@Observe	-1.00	-1.00	1.00
Experimenter on Trial@Gender@Approval	-0.02	-0.97	1.00
Experimenter on Trial@Gender@Observe	-0.16	-0.98	1.00
Experimenter on Trial@Approval@Observe	0.99	-0.95	1.00
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order@Age@NA	0.00	0.00	0.11
Condition Order@Gender@NA	0.00	0.00	0.01
Condition Order@Cultural Group@NA	0.00	0.00	0.18
Condition Order@Approval@NA	0.00	0.00	0.12
Condition Order@Observe@NA	0.38	0.26	0.44

(1) dummy coded with the Indifference condition being the reference level

(2) dummy coded with the Unobserved condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

Table A. 2.24

Results of full Model 6 (Model A in Table A.2.22) examining 5- to 11-year-old children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	1.03	0.27	0.54	1.53	3.79	<.001
Observe[(1)]	0.12	0.26	-0.39	0.59	0.47	.642
Approval[(2)]	0.05	0.20	-0.29	0.41	0.26	.796
Cultural Group[(3)]	-1.47	0.34	-2.11	-0.81	-4.31	<.001
Age[(4)]	0.15	0.18	-0.18	0.49	0.85	.393
Gender[(5)]	-0.43	0.26	-0.92	0.05	-1.65	.099
Approval*Observe	0.39	0.22	-0.04	0.79	1.76	.079
Approval*Cultural Group	0.20	0.26	-0.27	0.69	0.77	.441
Observe*Cultural Group	0.10	0.25	-0.41	0.64	0.39	.699
Approval*Age	0.14	0.16	-0.15	0.45	0.90	.367
Observe*Age	-0.19	0.15	-0.50	0.09	-1.25	.212
Cultural Group*Age	0.23	0.25	-0.23	0.68	0.92	.357
Approval*Gender	0.25	0.22	-0.17	0.66	1.15	.250
Observe*Gender	-0.12	0.22	-0.54	0.32	-0.53	.594
Cultural Group*Gender	0.48	0.37	-0.22	1.18	1.30	.193
Age*Gender	0.16	0.24	-0.31	0.63	0.66	.506

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Approval*Observe*Cultural Group	-0.33	0.31	-0.93	0.24	-1.07	.284
Approval*Observe*Age	-0.14	0.22	-0.56	0.28	-0.63	.528
Approval*Cultural Group*Age	-0.05	0.21	-0.44	0.36	-0.24	.812
Observe*Cultural Group*Age	0.07	0.21	-0.32	0.46	0.31	.754
Approval*Observe*Gender	-0.51	0.31	-1.11	0.09	-1.66	.098
Approval*Cultural Group*Gender	-0.29	0.31	-0.89	0.28	-0.92	.356
Observe*Cultural Group*Gender	0.36	0.30	-0.27	0.96	1.19	.236
Approval*Age*Gender	-0.29	0.22	-0.72	0.09	-1.34	.180
Observe*Age*Gender	0.17	0.21	-0.23	0.59	0.79	.432
Cultural Group*Age*Gender	0.02	0.36	-0.67	0.67	0.05	.959
Approval*Observe*Cultural Group*Age	0.25	0.30	-0.33	0.81	0.82	.413
Approval*Observe*Cultural Group*Gender	0.28	0.43	-0.51	1.15	0.65	.515
Approval*Observe*Age*Gender	0.31	0.30	-0.24	0.93	1.04	.299
Approval*Cultural Group*Age*Gender	0.16	0.31	-0.44	0.79	0.52	.604
Observe*Cultural Group*Age*Gender	-0.51	0.31	-1.07	0.05	-1.68	.093
Approval*Observe*Cultural Group*Age*Gender	0.02	0.43	-0.85	0.86	0.05	.962

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

Table A. 2.25

Results of reduced Model 6 (Model B in Table A.2.22) examining 5- to 11-year-old children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
intercept	1.02	0.27	0.53	1.57	3.76	<.001
Observe[(1)]	0.12	0.26	-0.39	0.62	0.46	.643
Approval[(2)]	0.05	0.20	-0.34	0.42	0.26	.793
Cultural Group[(3)]	-1.46	0.34	-2.19	-0.85	-4.29	<.001
Gender[(4)]	-0.43	0.26	-0.93	0.08	-1.62	.104
Age[(5)]	0.15	0.18	-0.19	0.49	0.86	.388
Approval*Observe	0.39	0.22	-0.05	0.85	1.77	.077
Observe*Cultural Group	0.10	0.25	-0.38	0.61	0.39	.698
Approval*Cultural Group	0.20	0.26	-0.31	0.72	0.76	.445
Observe*Gender	-0.12	0.22	-0.58	0.27	-0.53	.595
Approval*Gender	0.25	0.22	-0.20	0.71	1.15	.251
Cultural Group*Gender	0.48	0.37	-0.23	1.21	1.29	.198
Observe*Age	-0.19	0.15	-0.47	0.10	-1.31	.189
Approval*Age	0.15	0.15	-0.15	0.43	0.97	.333
Gender*Age	0.16	0.24	-0.33	0.60	0.69	.492
Cultural Group*Age	0.23	0.24	-0.26	0.70	0.95	.342

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Observe*Approval*Cultural Group	-0.33	0.31	-0.93	0.27	-1.07	.283
Observe*Approval*Gender	-0.51	0.31	-1.07	0.12	-1.66	.098
Observe*Cultural Group*Gender	0.36	0.30	-0.21	0.95	1.18	.236
Approval*Cultural Group*Gender	-0.28	0.31	-0.89	0.34	-0.92	.357
Observe*Approval*Age	-0.14	0.19	-0.54	0.23	-0.76	.448
Observe*Gender*Age	0.16	0.19	-0.22	0.52	0.87	.384
Approval*Gender*Age	-0.29	0.19	-0.66	0.07	-1.56	.119
Observe*Cultural Group*Age	0.06	0.19	-0.32	0.42	0.33	.741
Cultural Group*Gender*Age	0.02	0.34	-0.62	0.67	0.04	.965
Approval*Cultural Group*Age	-0.06	0.19	-0.42	0.31	-0.29	.769
Observe*Approval*Cultural Group*Gender	0.28	0.43	-0.58	1.11	0.65	.515
Observe*Approval*Gender*Age	0.32	0.22	-0.11	0.77	1.49	.135
Observe*Cultural Group*Gender*Age	-0.50	0.22	-0.93	-0.08	-2.29	.022
Approval*Cultural Group*Gender*Age	0.17	0.22	-0.28	0.61	0.76	.449
Observe*Approval*Cultural Group*Age	0.25	0.22	-0.15	0.72	1.17	.242

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) dummy coded with girls being the reference level

(5) measured in days; z-transformed

Table A. 2.26

Results of reduced Model 6 (Model C in Table A.2.22) examining 5- to 11-year-old children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.98	0.28	0.47	1.53	3.54	<.001
Observe[(1)]	0.19	0.26	-0.31	0.66	0.71	.476
Cultural Group[(2)]	-1.41	0.35	-2.13	-0.79	-4.07	<.001
Age[(3)]	0.25	0.17	-0.08	0.59	1.50	.133
Gender[(4)]	-0.37	0.25	-0.87	0.12	-1.47	.142
Approval[(5)]	0.12	0.19	-0.26	0.47	0.64	.525
Observe*Cultural Group	0.01	0.24	-0.43	0.48	0.03	.974
Observe*Age	-0.33	0.12	-0.58	-0.09	-2.69	.007
Cultural Group*Age	0.12	0.23	-0.35	0.53	0.52	.604
Observe*Gender	-0.20	0.19	-0.57	0.14	-1.05	.294
Cultural Group*Gender	0.40	0.35	-0.28	1.11	1.12	.261
Age*Gender	0.04	0.22	-0.45	0.46	0.16	.871
Observe*Approval	0.28	0.19	-0.09	0.66	1.50	.135
Cultural Group*Approval	0.10	0.24	-0.36	0.61	0.42	.673
Gender*Approval	0.17	0.19	-0.21	0.52	0.91	.365
Age*Approval	-0.04	0.11	-0.26	0.20	-0.39	.700

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Observe*Cultural Group*Age	0.19	0.15	-0.11	0.49	1.25	.211
Observe*Cultural Group*Gender	0.50	0.22	0.07	0.91	2.28	.023
Observe*Age*Gender	0.32	0.15	0.03	0.62	2.09	.037
Cultural Group*Age*Gender	0.10	0.32	-0.49	0.73	0.32	.749
Observe*Cultural Group*Approval	-0.16	0.21	-0.62	0.26	-0.75	.452
Observe*Gender*Approval	-0.35	0.21	-0.76	0.05	-1.64	.102
Cultural Group*Gender*Approval	-0.15	0.22	-0.55	0.27	-0.67	.500
Observe*Age*Approval	0.14	0.11	-0.07	0.36	1.32	.186
Age*Gender*Approval	-0.05	0.11	-0.27	0.17	-0.44	.659
Cultural Group*Age*Approval	0.15	0.11	-0.06	0.37	1.35	.178
Observe*Cultural Group*Age*Gender	-0.50	0.22	-0.92	-0.09	-2.29	.022

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Ugandan sample being the reference level

(3) measured in days; z-transformed

(4) dummy coded with girls being the reference level

(5) dummy coded with the Indifference condition being the reference level

Table A. 2.27

Results of reduced Model 6 (Model D in Table A.2.22) examining 5- to 11-year-old children's sensitivity to normative expectations, exploring gender (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.92	0.26	0.39	1.43	3.48	.001
Observe[(1)]	0.30	0.24	-0.18	0.78	1.21	.225
Cultural Group[(2)]	-1.35	0.33	-1.97	-0.66	-4.13	<.001
Age[(3)]	0.19	0.16	-0.11	0.54	1.19	.234
Gender[(4)]	-0.26	0.24	-0.73	0.22	-1.08	.281
Approval[(5)]	0.27	0.15	-0.02	0.57	1.72	.086
Observe*Cultural Group	-0.05	0.20	-0.46	0.31	-0.27	.786
Observe*Age	-0.26	0.11	-0.50	-0.04	-2.35	.019
Cultural Group*Age	0.19	0.22	-0.23	0.62	0.89	.375
Observe*Gender	-0.37	0.16	-0.67	-0.09	-2.35	.019
Cultural Group*Gender	0.33	0.34	-0.36	0.94	0.97	.332
Age*Gender	0.01	0.22	-0.44	0.43	0.06	.951
Observe*Approval	0.02	0.11	-0.20	0.22	0.16	.874
Gender*Approval	-0.06	0.11	-0.27	0.15	-0.53	.595
Cultural Group*Approval	-0.04	0.17	-0.39	0.29	-0.24	.812
Age*Approval	0.08	0.06	-0.02	0.20	1.49	.136

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Observe*Cultural Group*Age	0.19	0.15	-0.12	0.50	1.23	.218
Observe*Cultural Group*Gender	0.50	0.22	0.13	0.95	2.28	.023
Observe*Age*Gender	0.32	0.15	0.01	0.64	2.10	.036
Cultural Group*Age*Gender	0.10	0.32	-0.55	0.73	0.32	.752
Observe*Cultural Group*Age*Gender	-0.50	0.22	-0.95	-0.08	-2.28	.022

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Ugandan sample being the reference level

(3) measured in days; z-transformed

(4) dummy coded with girls being the reference level

(5) dummy coded with the Indifference condition being the reference level

2.2.3 Additional Analyses with 5- to 11-Year-Old Participants who Passed the Generous Prime Comprehension

2.2.3.1 Excluding Participants Who Did Not Pass the Norm Comprehension

In order to check the robustness of the results, the analysis for Model 6 (i.e. the most complex model which included all other lower-order interactions, as well as the interaction between approval, observation, cultural group, age and gender) was rerun with only participants who passed the Generous Prime comprehension on the first attempt for both the Approval and Indifference conditions. This reduced the sample to 97 UK children (50 girls) and 97 Ugandan children (50 girls), meaning that the sample consisted of 776 proportions of tokens given and kept from 194 participants. The same statistical approach was used as in the main text.

2.2.3.2 Results of Model 6 Excluding Participants Who Did Not Pass the Norm Comprehension

Model 6: Is sensitivity to normative expectations common between 4 - 5 years, and does it vary with age across cultural groups and genders?

Overall, Model 6 explained significantly more variation in the proportion of tokens given away than the null model (likelihood ratio test: $\chi 2 = 61.64$, df = 31, $p = .001$). The 5-way interaction between 'observe', 'approval', 'cultural group', 'age' and 'gender' was non-significant (estimate = 0.2, SE = 0.45, $p = .649$), so a reduced model that only comprised the 4-way interactions was run. The significant interaction between 'observe', 'cultural group', 'gender' and 'age' remained, and when the model was reduced to exclude any non-significant interactions with approval, only the main effect of approval remained (Table A.2.28), suggesting the pattern of results remained the same, and as such, is unlikely to be an effect of children not being able to answer explicit questions about the Generous Prime.

Table A. 2.28

Results of final Model 6 of 5-11-year-old children's sensitivity to normative expectations exploring gender, only including participants who passed the Generous Prime comprehension on the first attempt (estimates and standard errors, together with confidence limits and significance tests are shown).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.84	0.25	0.33	1.35	3.42	.001
Observe[(1)]	0.27	0.23	-0.15	0.73	1.16	.247
Approval[(2)]	0.24	0.11	0.03	0.46	2.27	.023
Cultural Group[(3)]	-1.27	0.25	-1.81	-0.72	-5.10	<.001
Age[(4)]	0.35	0.17	0.00	0.67	2.05	.040
Gender[(5)]	-0.23	0.24	-0.71	0.28	-0.93	.352
Observe*Cultural Group	-0.01	0.18	-0.39	0.36	-0.07	.945
Observe*Age	-0.26	0.12	-0.49	-0.03	-2.15	.031
Cultural Group*Age	0.08	0.22	-0.38	0.55	0.36	.720
Observe*Gender	-0.38	0.17	-0.69	-0.06	-2.30	.021
Cultural Group*Gender	0.27	0.33	-0.41	0.96	0.80	.425
Age*Gender	-0.07	0.22	-0.53	0.38	-0.33	.739
Observe*Cultural Group*Age	0.19	0.16	-0.11	0.49	1.19	.234
Observe*Cultural Group*Gender	0.51	0.23	0.07	0.95	2.23	.026
Observe*Age*Gender	0.32	0.17	0.00	0.64	1.94	.053
Cultural Group*Age*Gender	0.20	0.32	-0.43	0.86	0.61	.543
Observe*Cultural Group*Age*Gender	-0.50	0.23	-0.95	-0.11	-2.21	.027

(1) dummy coded with the Unobserved condition being the reference level

(2) dummy coded with the Indifference condition being the reference level

(3) dummy coded with the Ugandan sample being the reference level

(4) measured in days; z-transformed

(5) dummy coded with girls being the reference level

2.2.4 Additional Statistical Analyses Included in the Pre-Registered Analysis Plan

2.2.4.1 Measures of Additional Pre-Registered Analyses

The measures remained the same to those described in Appendix 2.1.4.

2.2.4.2 Statistical Approach of Additional Pre-Registered Analyses

Models A4, A5 and A6: Is sensitivity to observation, expectation or normative expectations common between 5 to 11 years and does it vary with age across cultural groups?

Three separate models were run to explore children's i) sensitivity to observation (Model A4), ii) sensitivity to approval (Model A5) or iii) sensitivity to normative expectations (i.e. the integration of both sensitivity to observation and expectations; Model A6). To analyse how children's sensitivity varied with cultural group (UK and Uganda) and age (a covariate, a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these predictors and their interaction included in the model. Gender was also included as a fixed effect into the model. Random intercepts were the experimenters on a given trial, and the order of conditions. We included 'gender' as a random slope within 'experimenters on trial' and included 'cultural group' as a random slope within the 'order of conditions'. The random slope parameters were manually dummy coded and centred. Initially this was a maximal model (Barr et al., 2013), including parameters for the correlations among random intercepts and slopes. However, as all correlation parameters were estimated to have a coefficient of exactly 1 in all 3 models, which indicated that they were likely not identifiable (Matuschek et al., 2017), we removed the correlation parameters of the random slopes from the models which led to an only minor reduction in model fit (the maximum difference was in Model A4; log-likelihoods, model with correlation parameters included: -135.96, df = 21 model without correlation parameters: -137.43, df = 11). To test the significance of the predictor variables, the fit of the full model was compared to the fit of a null model comprising only the control variable of 'gender' and the random effects using a likelihood ratio test (Dobson, 2002). For each model the respective sample consisted of 211 sensitivity scores from 211 participants.

2.2.4.3 Results of Additional Pre-Registered Analyses

Model A4: Is sensitivity to observation evident between 5 to 11 years and does it vary with age across cultural groups?

Descriptively, 31/98 children in the UK and 51/113 children in Uganda showed sensitivity to observation, meaning they shared more when they were being observed by an experimenter, compared to when they were not being observed. Overall, Model A4 did not explain significantly more variation in sensitivity to observation scores than the null model (likelihood ratio test: $\chi2 = 0.13$, df = 2, $p = .528$). This suggests that cultural group or age do not significantly impact how sensitive children are to observation.

Model A5: Is sensitivity to approval evident between 5 to 11 years and does it vary with age across cultural groups?

Descriptively, 32/98 children in the UK and 45/113 children in Uganda showed sensitivity to approval, meaning they shared more when the experimenter had expressed an approval of generous sharing behaviour, compared to when the experimenter was indifferent to how children shared. Overall, Model A5 did not explain significantly more variation in sensitivity to approval scores than the null model (likelihood ratio test: $\chi2 = 4.68$, df = 3, $p = .197$). This suggests that cultural group or age do not significantly impact how sensitive children are to approval.

Model A6: Is sensitivity to normative expectations evident between 5 to 11 years and does it vary with age across cultural groups?

Descriptively, 19/98 children in the UK and 40/113 children in Uganda showed sensitivity to normative expectations, meaning there was a more positive effect of observation when the experimenter approved of generosity, compared to the effect of observation when the experimenter was indifferent to generosity. Overall, Model A6 did not explain significantly more variation in sensitivity to observation scores than the null model (likelihood ratio test: $\chi2 = 5.96$, df = 9, $p = .744$). This suggests that cultural group or age do not significantly impact how sensitive children are to normative expectations.

Appendix 3: Supplementary Materials for Chapter 3

3.1 Appendix for Part 1

3.1.1 Materials and Procedures

The experimenter introduced the game by saying: "I'm going to tell you some stories about my friends now and ask you some questions as we go along. You can answer the questions however you like!"

3.1.1.1 Diverse Desires

UK Protocol:

The experimenter shows the child a boy puppet and the sheet of paper with a carrot and a biscuit and says*: "This is Tom. Tom wants a snack to eat. Here are two different snacks: a carrot and a biscuit. Which snack do you like best?"* (Content/Reality Question)

If the child chooses the carrot, the experimenter says: "Well, that's a good choice, but Tom really likes biscuits. He doesn't like carrots. What he likes best are biscuits."

If the child chooses the biscuit, the experimenter says: "Well, that's a good choice, but Tom really likes carrots. He doesn't like biscuits. What he likes best are carrots."

The experimenter then asks: "Which snack will Tom eat? A carrot or a biscuit?" (Target Question)

Ugandan Protocol:

The experimenter shows the child a boy puppet and the sheet of paper with a banana and mandazi and says*: "This is Tom. Tom wants a snack to eat. Here are two different snacks: a banana and a mandazi. Which snack do you like best?"* (Content/Reality Question)

If the child chooses the banana, the experimenter says: "Well, that's a good choice, but Tom really likes mandazi. He doesn't like bananas. What he likes best are mandazi."

If the child chooses the mandazi, the experimenter says: **"**Well, that's a good choice, but Tom really likes bananas. He doesn't like mandazi. What he likes best are bananas."

The experimenter then asks: "Which snack will Tom eat? A banana or mandazi?" (Target Question)

Scoring:

To be scored correct, the child must answer the Target Question opposite from their answer to the Contrast/Reality Question.

3.1.1.2 Diverse Beliefs

UK Protocol:

The experimenter shows the child a girl puppet, and a sheet of paper with bushes and a garage.

The experimenter says: “This is Linda. Linda has lost her cat and is looking for her cat. Her cat might be hiding in the bushes or it might be hiding in the garage. Where do you think the cat is? In the bushes or in the garage?” (Contrast/Reality Question)

If the child chooses the bushes, the experimenter says: “Well, that’s a good idea, but Linda thinks her cat is in the garage. She thinks her cat is in the garage.”

If the child chooses the garage, the experimenter says: “Well, that’s a good idea, but Linda thinks her cat is in the bushes. She thinks her cat is in the bushes.”

The experimenter then asks: “So, where will Linda look for her cat? In the bushes or in the garage?” (Target Question)

Ugandan Protocol:

The experimenter shows the child a girl puppet, and a sheet of paper with maize and a goat house (i.e. structure built to keep goats).

The experimenter says: “This is Linda. Linda has lost her goat and is looking for her goat. Her goat might be hiding in the maize or it might be hiding in the goat house. Where do you think the goat is? In the maize or in the goat house?” (Contrast/Reality Question)

If the child chooses the maize, the experimenter says: “Well, that’s a good idea, but Linda thinks her goat is in the goat house. She thinks her goat is in the goat house.”

If the child chooses the house, the experimenter says: “Well, that’s a good idea, but Linda thinks her goat is in the maize. She thinks her goat is in the maize.”

The experimenter then asks: “So, where will Linda look for her goat? In the maize, or in the goat house? (Target Question)

Scoring:

To be scored correct, the child must answer the Target Question opposite from their answer to the Contrast/Reality Question.

3.1.1.3 Knowledge Access

UK Protocol:

The experimenter shows the child a cardboard box (with a dog figurine inside) and says: “Here’s a box. What do you think is inside the box?” (The child can give any answer he or she likes or indicate that he or she does not know.)

Next, the experimenter opens the box and show the contents of the box to the child, and says: *“It’s really a dog inside!”*

The experimenter puts the dog back inside, close the box and asks: *“What is inside the box?”*

This is repeated up to 3 times if the child does not answer or answers incorrectly. If the child still answers incorrectly after the third demonstration, the experimenter asks: *“Is it a dog inside or a chocolate bar?”*

Regardless of the answer, the experimenter then shows the child a girl doll and says: *“Polly has never ever seen inside the box. Does Polly know what is in the box?”* (Target Question)

The experimenter then asks: *“Did Polly see inside this box?”* (Contrast/Reality Question)

Ugandan Protocol:

The experimenter shows the child a cardboard box (with a goat figurine inside) and says: “Here’s a box. What do you think is inside the box?” (The child can give any answer he or she likes or indicate that he or she does not know.)

Next, the experimenter opens the box and show the contents of the box to the child, and says: *“It’s really a goat inside!”*

The experimenter puts the goat back inside, close the box and asks: *“What is inside the box?” (Prompt Question)*

This is repeated up to 3 times if the child does not answer or answers incorrectly. If the child still answers incorrectly after the third demonstration, the experimenter asks: *"Is it a dog inside or a chocolate bar?"*

Regardless of the answer, the experimenter then shows the child a girl doll and says: *"Alice has never ever seen inside the box. Does Alice know what is in the box?"* (Target Question)

The experimenter then asks: *"Did Alice see inside this box?"* (Contrast/Reality Question)

Scoring

To be scored correct, the child must answer the Target Question, "no" and the Contrast/Reality Question, "no".

3.1.1.4 Contents False Belief

UK Protocol:

Prior to testing, the mother of the child was asked whether she thought her child would be most familiar with raisins or smarties to ensure the container was recognisable to the child.

The experimenter shows the child the raisins box/smarties box (with a toy plastic pig inside) and says: *"Here's a raisins/smarties box. What do you think is inside the raisin/smarties box?"*

The experimenter encourages the child to answer. If the child does not answer or answers incorrectly (i.e., does not say "raisins/smarties"), the experimenter says: *"Can you see the raisins/smarties on the outside of the box? So, what do you think is inside?"*

If the child still fails to answer the question correctly, the experimenter says: "Do you think there is raisins/smarties inside the box or berries inside the box?"

Next, the experimenter opens the box and says: *"It's really a pig inside!"*

The experimenter puts the pig back inside, closes the box and asks: *"Ok, what is inside the X box?"*

This was repeated up to 3 times if the child does not answer or answers incorrectly. If the child still answers incorrectly after the third demonstration, the experimenter asks: *"Is there a pig inside, or raisins/smarties inside?"*

Then, the experimenter shows the child a boy doll and says: “Peter has never ever seen inside this raisins/smarties box. What does Peter think is in the box? Raisins/smarties or a pig?” (Target Question)

The experimenter then asks: *“Did Peter see inside this box?”* (Contrast/Reality Question)

Ugandan Protocol:

Prior to testing, the Ugandan research assistants advised that a bottle of Baby Junior (petroleum jelly used as body balm) would be recognisable to all children.

The experimenter shows the child the bottle of Baby Junior (with a toy plastic chicken inside) and says: *“Here’s a bottle of Baby Junior. What do you think is inside the bottle of Baby Junior?”*

The experimenter encourages the child to answer. If the child does not answer or answers incorrectly (i.e., does not say “jelly”), the experimenter says: *“Can you see the jelly on the outside of the bottle? So, what do you think is inside?”*

If the child still fails to answer the question correctly, the experimenter says: *“Do you think there is jelly inside the bottle or leaves inside the bottle?”*

Next, the experimenter opens the bottle and says: *“It’s really a chicken inside!”*

The experimenter puts the chicken back inside, closes the bottle and asks: *“Ok, what is inside the bottle of Baby Junior?”*

This was repeated up to 3 times if the child does not answer or answers incorrectly. If the child still answers incorrectly after the third demonstration, the experimenter asks: *“Is there a chicken inside, or jelly inside?”*

Then, the experimenter shows the child a boy doll and says: “Moses has never ever seen inside this bottle of Baby Junior. What does Moses think is in the bottle? Jelly or a chicken?” (Target Question)

The experimenter then asks: *“Did Moses see inside this bottle?”* (Contrast/Reality Question)

Scoring:

To be scored correct, the child must answer the Target Question, “no” and the Contrast/Reality Question, “no”.

3.1.1.5 Explicit False Belief

UK Protocol:

The experimenter shows the child a boy puppet and a sheet of paper with a bag and a box on and says: "This is Scott. Scott wants to find his hat. His hat might be in his bag, or it might be in the box. Really, Scott's hat is in his bag. But Scott thinks his hat is in the box."

The experimenter then asks: "So, where will Scott look for his hat? In his bag or in the box?" (Target Question)

"Where is Scott's hat really? In his bag, or in the box?" (Contrast/Reality Question)

Ugandan Protocol:

The experimenter shows the child a boy puppet and a sheet of paper with a bag and a box on and says: "This is Simon. Simon wants to find his hat. His hat might be in his bag, or it might be in the box. Really, Simon's hat is in his bag. But Simon thinks his hat is in the box."

The experimenter then asks: "So, where will Simon look for his hat? In his bag or in the box?" (Target Question)

"Where is Simon's hat really? In his bag, or in the box?" (Contrast/Reality Question)

Scoring

To be scored correct, the child must answer the Target Question, "box" and the Contrast/Reality Question, "bag".

3.1.1.6 Belief-Emotion

UK Protocol:

Prior to testing, the mother of the child was asked whether she thought her child would be most familiar with Cheerios, Cocopops, Rice Krispies or Shreddies, to ensure the container was recognisable to the child.

The experimenter shows the child a boy puppet and a cereal box (with rocks inside the closed box) and says: *"This is Michael. Here's a X box. What do you think is inside the X box?"* (X = whichever cereal box the mother thinks the child will most likely recognise)

If the child does not answer or answers incorrectly (i.e., does not say "X cereal"), the experimenter says: *"Can you see the X cereal on the outside of the box? So what do you think is inside?"*

If the child still fails to answer the question correctly, the experimenter says: *"Do you think there is X inside the box or grapes inside the box?"* Followed by, "That is good because X is Michael's favourite food. Now Michael will go play over here."

Michael is then put away out of sight. Next, the cereal box is opened and the contents are shown to the child, and the experimenter says: *"There are really stones inside and not X! There's nothing but stones."*

The experimenter puts the stones back inside the cereal box, closes it and then asks: *"What is Michael's favourite food?"* If the child does not answer or answers incorrectly, the experimenter says, *"Remember Michael's favourite food is X. So, what is Michael's favourite food?"*

If the child still fails to answer the question correctly, the experimenter says: *"Do you think Michael's favourite food is X, or is it banana?"*

The experimenter brings Michael back and says: "Michael has never ever seen inside this box. Michael is back and he wants to eat. Let's give Michael this box. How does Michael feel now that he has the box? Happy or sad?" (Target Question)

The experimenter opens the cereal box and lets the puppet look inside. The experimenter asks the child*: "How does Michael feel now he has looked inside the box? Happy or sad?"* (Contrast/Reality Question)

<u>Ugandan Protocol:</u>

Prior to testing, the Ugandan research assistants advised that a plastic food dish would be recognisable to all children as a container that held food.

The experimenter shows the child a boy puppet and a food dish (with rocks inside the closed dish) and says: *"This is Joseph. Here's a dish. What do you think is inside the dish?"*

If the child does not answer or answers incorrectly (i.e., does not say a food item) the experimenter says: *"Do you think there is food inside?"*

If the child still fails to answer the question correctly, the experimenter says: *"Do you think there is food inside the dish or a leaf inside the dish?"* Followed by, "That is good because Joseph is hungry."

Joseph is then put away out of sight. Next, the food dish is opened and the contents are shown to the child, and the experimenter says: *"There are really stones inside and not food! There's nothing but stones."*

The experimenter puts the stones back inside the food dish, closes it and then asks: *"Is Joseph hungry?"*

The experimenter brings Joseph back and says: "Joseph has never ever seen inside this dish. Joseph is back and he wants to eat. Let's give Joseph this dish. How does Joseph feel now that he has the dish? Happy or sad?" (Target Question)

The experimenter opens the cereal dish and lets the puppet look inside. The experimenter asks the child*: "How does Joseph feel now he has looked inside the dish? Happy or sad?"* (Contrast/Reality Question)

Scoring

To be scored correct, the child must answer the Target Question, 'happy' and answer the Contrast/Reality Question, 'sad'.

3.1.1.7 Real-Apparent Emotion

UK Protocol:

First, the experimenter shows the child the sheet of paper with three faces drawn on it – a happy, neutral and sad face. The experimenter checks the child knows these emotional expressions. First, by pointing to the happy face and say, "What is this one?" before repeating this with the sad and neutral faces. If the child does not know the 'Okay' face, the experimenter can say, "Shall we call this one okay?"

The experimenter then shows the picture of the boy from the back (face not visible) and says: "This is a story about a boy. I'm going to ask you about how the boy really feels inside and how he looks on his face. He might really feel one way inside but look a different way on his face. Or, he might really feel the same way inside as he looks on his face. I want you to tell me how he really feels inside and how he looks on his face.''

''This story is about Matt. Matt's friends were playing together and telling jokes. One of the older children, Rosie, told a mean joke about Matt and everyone laughed. Everyone thought it was very

funny, but not Matt. But, Matt didn't want the other children to see how he felt about the joke, because they would call him a baby. So, Matt tried to hide how he felt.''

The experimenter then probes the child's memory of the story by asking, *''What did the other children do when Rosie told a mean joke about Matt?''* (Laughed or thought it was funny.)

If the child answers incorrectly, or refuses to answer, the experimenter says: "Remember in the story, Rosie, told a mean joke about Matt and everyone laughed. Everyone thought it was very funny, so what did the other children do when Rosie told a mean joke about Matt?"

If the child still fails to answer the question correctly, the experimenter asks: "*Did the other children laugh, or did they not laugh?"*

The experimenter then proves the child's memory of the story again by asking: *''In the story, what would the other children do if they knew how Matt felt?''* (Call Matt a baby or tease him.)

If the child answers incorrectly, or refuses to answer, the experimenter says: "Remember in the story, Matt didn't want the other children to see how he felt about the joke, because they would call him a baby, so what would the other children do if they knew how Matt felt?"

If the child still fails to answer the question correctly, the experimenter says, "*Would they tease him and call him a baby, or would they not?"*

Regardless of the child's answers, the experimenter then continues with the questions.

Pointing to the three emotion pictures, the experimenter asks: *''So, how did Matt really feel, when everyone laughed? Did he feel happy, sad, or okay?''* (Contrast/Reality Question)

The experimenter then asks: ''How did Matt try to look on his face, when everyone laughed? Did he look happy, sad, or okay?" (Target Question).

<u>Ugandan Protocol</u>

The Ugandan version of this task was identical to the UK version, apart from the boy was named Richard (instead of Matt) and the girl was called Rose (instead if Rosie). However, this task was only conducted with 12/61 children due to children not being able to correctly identify the emotions represented by the faces. This story was therefore removed from analysis for both samples.

Scoring

To be scored correct, the child's answer to Contrast/Reality Question must be more negative than his or her answer to the Target Question (i.e., sad for Contrast/Reality and happy or okay for Target, or okay for Contrast/Reality and happy for Target).

3.1.2 Prompt Questions and Deviations from the Protocol

Table A.3.1 includes a summary of children's responses to the Prompt Questions in the first 6 stories. The stories related to Diverse Desires, Diverse Beliefs and Explicit False Belief did not include Prompt Questions, meaning these questions were only included in Knowledge Access, Contents False Belief and Belief-Emotion. Experimenter error meant that some Prompt Questions were omitted from the protocol, however, because these questions were never intended as exclusion criteria, these participants were included in the final analysis.

Table A. 3.1

Children's responses to the Prompt Questions in the Knowledge Access, Contents False Belief and Belief-Emotion stories.

		UK				Uganda			
Story	Prompt Question	Correct Response	Incorrect Response	No Response	Experimenter Error – Answer Unknown	Correct Response	Incorrect Response	No Response	Experimenter Error – Answer Unknown
Knowledge Access	What's inside the box?	66 (100%)	0	0	0	61 (100%)	0	0	0
Contents False Belief	What's do you think is inside the raisins box/smarties box/bottle of Baby Junior?	56 (84.85%)	3 (4.55%)	3 (4.55%)	4 (6.05%)	50 (81.97%)	2 (3.28%)	2 (3.28%)	7 (11.47%)
	What's inside the raisins box/smarties box/bottle of Baby Junior?	65 (98.48%)	1 (1.52%)	0	0	60 (98.36%)	1 (1.64%)	0	0
Belief-Emotion	What do you think is inside the cereal box/dish?	39 (59.09%)	24 (36.36%)	3 (4.55%)	0	37 (60.66%)	11 (18.03%)	6 (9.84%)	7 (11.47%)
	What's Michael's favourite food/Is Joseph hungry?	66 (100%)	0	0	0	53 (86.89%)	7 (11.47%)	0	1 (1.64%)

3.1.3 Interobserver Reliability for the Theory of Mind Items

To begin, the anchor coder developed a coding scheme and coded all videos. The anchor coder then trained a research assistant on the coding scheme, which gave the coder the opportunity to seek clarification and ask questions. During the training phase, 4 videos were coded by the research assistant. Upon feedback from the anchor coder, the research assistant then coded 28 videos that were randomly selected from the total dataset (22.05% of the data). Cohen's Kappa for the items used in the analysis is reported in Table A.3.2, along with the percentage of agreement between the two coders.

Table A. 3.2
The individual Cohen's kappa's and percentage of agreement for each Theory of Mind item's Target Question and Contrast/Reality Question for the lead author and research assistant coder.

Item	Question	Cohen's Kappa	Percentage agreement
Diverse Desires	Target Q	1.00	100%
	Contrast/Reality Q	1.00	100%
Diverse Beliefs	Target Q	1.00	100%
	Contrast/Reality Q	1.00	100%
Knowledge Access	Target Q	0.94	96.43%
	Contrast/Reality Q	0.91	96.43%
Explicit False Belief	Target Q	1.00	100%
	Contrast/Reality Q	1.00	100%

3.1.4 Statistical Approach and Analyses

3.1.4.1 Additional Details Regarding the Process of Reaching the Final Model: Model 1 (RQ1)

To analyse how ToM score varied with gender and cultural group, the proportion of correct and incorrect ToM items were included as a response variable in the model, with gender (girls or boys), cultural group (UK or Uganda) and the interaction included into the model as categorical predictors. The model stability parameters for Model 1 can be viewed in Table A.3.3. As the model did not required reduction, the first and final model is discussed in the Results section.

Table A. 3.3

Model stability parameters for Model 1, exploring whether gender or cultural group predict ToM score, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	-0.32	-0.39	-0.26
Gender(2)	0.28	0.21	0.36
Cultural Group(3)	1.57	1.51	1.64
Gender*Cultural Group	-0.84	-0.92	-0.73
Pariticpant ID@Intercept@NA	0.00	0.00	0.00

(1) dummy coded with girls being the reference level

(2) dummy coded with the Ugandan sample being the reference level

3.1.4.2 Additional Details Regarding the Process of Reaching the Final Model: Model 2 (RQ2)

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), Theory of Mind (continuous score ranging from 0-4), gender (girls or boys) or cultural group (UK or Uganda), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these four variables and their 4-way interaction included as fixed effects into the model.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating high instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -1048.03, df = 42; model without correlation parameters: -1049.72, df = 26). The process of reaching the final model is highlighted in Table A.3.4. The model stability parameters for Model 2 can be viewed in Table A.3.5.

The 4-way interaction was non-significant (see Table A.3.6), so a reduced model was run which consisted of the 3-way interactions (see Model B in Table A.3.4). The reduced model also did not show any significant 3-way interactions (see Table A.3.7), so a further reduced model was run which only comprised the 2-way interactions (Model C in Table A.3.4). This model only showed a significant interaction between 'AO condition' and 'gender' (Table A.3.8), so the model was reduced further to only include this significant interaction and the main effect (Model D in Table A.3.4), which is primarily discussed in the results section (Chapter 3; RQ2).

Table A. 3.4

Model Reduction Prior to the Final Model 2 Presented in the Chapter 3 Results Section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	ToM Score*AO Condition *Gender*Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO condition + ToM)||Experimenter on Trial[1] (Gender + Cultural Group + AO condition + ToM)||Order of Conditions[1]
B (reduced)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	ToM Score*AO Condition *Gender + ToM Score*AO Condition*Cultural Group + ToM Score*Gender*Cultural Group + AO Condition*Gender*Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO condition + ToM)||Experimenter on Trial[1] (Gender + Cultural Group + AO condition + ToM)||Order of Conditions[1]
C (reduced)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	ToM Score*AO Condition + ToM Score*Gender + ToM Score*Cultural Group + AO Condition*Gender + AO Condition*Cultural Group + Gender*Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO condition + ToM)||Experimenter on Trial[1] (Gender + Cultural Group + AO condition + ToM)||Order of Conditions[1]
D (final)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	AO Condition*Gender + Cultural Group + ToM Score	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO condition + ToM)||Experimenter on Trial[1] (Gender + Cultural Group + AO condition + ToM)||Order of Conditions[1]

[1]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

Table A. 3.5

Model stability parameters for Model 2, exploring whether ToM score predicts 4-year-old children's sensitivity to normative expectations in girls and boys in the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.76	0.43	1.09
AO[(1)]	-0.81	-1.79	0.06
ToM Score	0.12	-0.06	0.36
Gender[(2)]	0.44	-2.51	2.47
Cultural Group[(3)]	-1.89	-3.07	-0.02
AO*ToM Score	0.46	0.19	0.99
AO*Gender	1.72	0.02	2.57
ToM Score*Gender	-0.67	-2.03	0.56
AO*Cultural Group	1.20	-0.14	2.28
ToM Score*Cultural Group	0.10	-0.51	0.74
Gender*Cultural Group	-0.74	-2.51	1.51
AO*ToM Score*Gender	-0.63	-1.57	-0.06
AO*ToM Score*Cultural Group	-0.54	-1.04	-0.18
AO*Gender*Cultural Group	-0.94	-2.63	1.19
ToM Score*Gender*Cultural Group	0.49	-0.29	1.98
AO*ToM Score*Gender*Cultural Group	0.50	-0.15	1.38
Pariticpant ID@Intercept@NA	2.09	1.96	2.39

Term	Original	Min	Max
Experimenter on Trial@Intercept@NA	0.12	0.00	0.20
Experimenter on Trial@Gender@NA	0.00	0.00	0.49
Experimenter on Trial@AO@NA	0.38	0.01	0.55
Experimenter on Trial@ToM@NA	0.02	0.00	0.06
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order@Gender@NA	0.00	0.00	0.03
Condition Order@Cultural Group@NA	0.00	0.00	0.04
Condition Order@AO@NA	0.13	0.00	0.19
Condition Order@ToM@NA	0.00	0.00	0.00

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) dummy coded with girls being the reference level

(3) dummy coded with the Ugandan sample being the reference level

Table A. 3.6

Results of full Model 2 (Model A in Table A.3.4) examining whether 4-year-old children's ToM scores predict sensitivity to normative expectations in girls and boys from the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.76	0.73	-0.65	2.27	1.04	.299
AO[(1)]	-0.81	0.37	-1.57	0.01	-2.6	.031
ToM Score	0.12	0.37	-0.62	0.85	0.32	.745
Gender[(2)]	0.44	1.41	-2.12	3.13	0.31	.757
Cultural Group[(3)]	-1.89	1.99	-5.94	2.17	-0.95	.344
AO*ToM Score	0.46	0.17	0.11	0.87	2.76	.006
AO*Gender	1.72	0.83	0.37	3.23	2.08	.038
ToM Score*Gender	-0.67	0.68	-2.01	0.58	-0.98	.329
AO*Cultural Group	1.20	0.96	-0.89	3.27	1.25	.212
ToM Score*Cultural Group	0.10	0.69	-1.25	1.48	0.15	.883
Gender*Cultural Group	-0.74	2.60	-5.74	3.99	-0.28	.777
AO*ToM Score*Gender	-0.63	0.38	-1.36	0.01	-1.66	.097
AO*ToM Score*Cultural Group	-0.54	0.32	-1.26	0.15	-1.68	.092
AO*Gender*Cultural Group	-0.94	1.30	-3.29	1.73	-0.72	.471
ToM Score*Gender*Cultural Group	0.49	0.98	-1.35	2.29	0.50	.619
AO*ToM Score*Gender*Cultural Group	0.50	0.50	-0.47	1.50	0.99	.324

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) dummy coded with girls being the reference level

(3) dummy coded with the Ugandan sample being the reference level

Table A. 3.7

Results of reduced Model 2 (Model B in Table A.3.4) examining whether 4-year-old children's ToM scores predict sensitivity to normative expectations in girls and boys from the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.73	0.73	-0.74	2.21	1.01	.313
AO[(1)]	-0.75	0.36	-1.53	0.07	-2.09	.036
ToM Score	0.13	0.37	-0.59	0.84	0.35	.724
Gender[(2)]	0.54	1.41	-2.22	3.07	0.38	.702
Cultural Group[(3)]	-1.73	1.99	-6.02	2.27	-0.87	.385
AO*ToM Score	0.43	0.16	0.05	0.80	2.72	.007
AO*Gender	1.19	0.58	0.16	2.24	2.03	.042
ToM Score*Gender	-0.72	0.68	-2.03	0.58	-1.05	.293
AO*Cultural Group	0.58	0.79	-1.00	2.23	0.73	.465
ToM Score*Cultural Group	0.05	0.69	-1.28	1.48	0.07	.941
Gender*Cultural Group	-1.00	2.59	-6.14	3.71	-0.38	.701
AO*ToM Score*Gender	-0.37	0.25	-0.87	0.11	-1.48	.139
AO*ToM Score*Cultural Group	-0.34	0.25	-0.86	0.15	-1.33	.184
AO*Gender*Cultural Group	0.59	0.98	-1.27	2.50	0.60	.549
ToM Score*Gender*Cultural Group	0.30	0.45	-0.64	1.27	0.67	.500

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with girls being the reference level

[(3)] dummy coded with the Ugandan sample being the reference level

Table A. 3.8

Results of reduced Model 2 (Model C in Table A.3.4) examining whether 4-year-old children's ToM scores predict sensitivity to normative expectations in girls and boys from the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.75	0.69	-0.54	2.08	1.09	.277
AO[(1)]	-0.26	0.30	-0.91	0.34	-0.87	.385
ToM Score	0.12	0.34	-0.52	0.75	0.36	.717
Gender[(2)]	0.18	1.09	-1.82	2.32	0.16	.869
Cultural Group[(3)]	-2.35	1.49	-5.18	0.48	-1.57	.115
AO*ToM Score	0.13	0.10	-0.08	0.35	1.25	.213
ToM Score*Gender	-0.53	0.49	-1.47	0.41	-1.09	.276
ToM Score*Cultural Group	0.25	0.49	-0.69	1.22	0.51	.608
AO*Gender	0.53	0.18	0.17	0.89	2.91	.004
Gender*Cultural Group	0.52	0.95	-1.35	2.23	0.55	.582
AO*Cultural Group	-0.04	0.29	-0.63	0.58	-0.13	.893

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with girls being the reference level

[(3)] dummy coded with the Ugandan sample being the reference level

3.1.4.3 Additional Details Regarding the Process of Reaching the Final Model: Model 3 (RQ2)

To check that a relationship between ToM and sensitivity to normative expectations was not observed in Model 2 due to a lack of statistical power, we ran an additional analysis which only tested the interaction between 'Approval Observe condition' and 'ToM Score'. Therefore, to analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions) and Theory of Mind (continuous score ranging from 0-4), these categorical predictors and their interaction were entered into a logistic Generalized Linear Mixed Model (Baayen, 2008). Gender (girls or boys) and cultural group (UK or Uganda) were included as control variables, and as such, were included in both the full and null models.
As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating high instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -1057.81, df = 32; model without correlation parameters: -1058.83, df = 16). The model stability parameters for Model 3 can be viewed in Table A.3.9.

As the full-null model comparison was non-significant, as described in the Statistical Approach section, we can assume that Approval Observe condition, ToM or the interaction of them two variables do not significantly affect the proportion of tokens children give.

Table A. 3.9

Model stability parameters for Model 3, exploring whether ToM score predicts 4-year-old children's sensitivity to normative expectations, controlling for gender and cultural group, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.91	0.49	1.36
AO[(1)]	-0.01	-0.36	0.44
ToM Score	0.00	-0.15	0.17
Gender[(2)]	-0.72	-0.92	-0.51
Cultural Group[(3)]	-1.41	-1.59	-1.21
AO*ToM Score	0.11	-0.04	0.25
Pariticpant ID@Intercept@NA	2.08	1.99	2.42
Experimenter on Trial@Intercept@NA	0.09	0.00	0.21
Experimenter on Trial@Gender@NA	0.00	0.00	0.49
Experimenter on Trial@AO@NA	0.31	0.19	0.44
Experimenter on Trial@ToM@NA	0.04	0.00	0.06
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order@Gender@NA	0.00	0.00	0.04
Condition Order@Cultural Group@NA	0.00	0.00	0.02
Condition Order@AO@NA	0.10	0.00	0.15
Condition Order@ToM@NA	0.00	0.00	0.01

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) dummy coded with girls being the reference level

(3) dummy coded with the Ugandan sample being the reference level

3.2 Appendix for Part 2

3.2.1 Statistical Approach and Analyses

3.2.1.1 Additional Details Regarding the Process of Reaching the Final Model: Model 4 (RQ3)

To analyse whether DCCS performance varied with gender or cultural group, a logistical Generalized Linear Model was used. Children's performance on the DCCS (pass/fail) was used as an outcome variable, and gender (girls or boys), cultural group (UK or Uganda) and the interaction were added as categorical predictors to the model. The model stability parameters for the full Model 5 can be viewed in Table A.3.10.

There was no significant interaction between cultural group and gender (Table A.3.11), so a reduced model with just the main effects was run (described in the Results section).

Table A. 3.10

Model stability parameters for Model 4, exploring whether gender or cultural group predict children's DCCS performance, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	-1.57	-1.63	-1.39
Gender[(1)]	-0.22	-0.40	0.06
Cultural Group[(2)]	1.93	1.75	1.99
Gender*Cultural Group	0.69	0.41	0.87

[(1)] dummy coded with girls being the reference level

[(2)] dummy coded with the Ugandan sample being the reference level

Table A. 3.11

Results of full Model 4 examining whether 4-year-old children's gender and cultural group predict DCCS performance (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	-1.57	0.49	-2.66	-0.69	-3.19	.001
Gender[(1)]	-0.22	0.79	-1.91	1.31	-0.28	.779
Cultural Group[(2)]	1.93	0.60	0.80	3.20	3.20	.001
Gender*Cultural Group	0.69	0.98	-1.20	2.70	0.71	.479

[(1)] dummy coded with girls being the reference level

[(2)] dummy coded with the Ugandan sample being the reference level

3.2.1.2 Additional Details Regarding the Process of Reaching the Final Model: Model 5 (RQ4)

To analyse how the proportion of tokens given to the unfamiliar child varied with the Approval Observe (Approval Observe or Other Conditions), DCCS (Pass/Fail), gender (girls or boys) or cultural group (UK or Uganda), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these four variables (all considered categorical predictors) and their 4-way interaction included as fixed effects into the model. As the estimates of the correlations between the random slopes of 'order of conditions' ranged from -1 to 1, indicating high instability, we removed these correlation parameters from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -858.57, df = 42; model without correlation parameters: -862.96, df = 26). The model stability parameters for Model 5 can be viewed in Table A.3.13.
The process of reaching the final model discussed in the Results section can be viewed in Table A.3.12. The 4-way interaction between 'Approval Observe condition', 'DCCS', 'gender' and 'cultural group' was non-significant (Table A.3.14), so a reduced model was run without this interaction (Model B in Table A.3.12). None of the 3-way interactions were significant (Table A.3.15), so a further reduced model was run without these interactions (Model C in Table A.3.12). This model revealed a significant interaction between 'DCCS' and 'cultural group' (Table A.3.16), so other non-significant 2-way interactions were removed (Model D in Table A.3.12), leaving this interaction and

the main effects of ‘Approval Observe condition’ and ‘gender’, which is discussed in the Results section.

Table A. 3.12

Model Reduction Prior to the Final Model 5 Presented in the Results Section.

Model Version	Model Type	Outcome Variable	Predictor Variable(s)	Random Intercepts	Random Slopes
A (full)	Binomial GLMM	Tokens given|kept in test trials	AO*DCCS *Gender *Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + AO + DCCS)|Experimenter on Trial[1] (Gender + Cultural Group + AO + DCCS)||Order of Conditions[1]
B (reduced)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	AO*DCCS*Gender + AO*DCCS*Cultural Group + AO*Gender*Cultural Group + DCCS*Gender*Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
C (reduced)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	AO*DCCS + AO*Gender + AO*Cultural Group + DCCS*Gender + DCCS*Cultural Group + Gender*Cultural Group	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]
D (final)	Binomial GLMM	Proportion of Tokens Given and Kept in Test Trials	DCCS*Cultural Group + Gender + AO	Participant ID, Experimenter on Trial, Order of Conditions	(Gender + Observe + Approval)||Experimenter on Trial[1] (Gender + Cultural Group + Observe + Approval)||Order of Conditions[1]

[1]The correlations between random slopes and intercepts was unidentifiable and therefore excluded from the model

Table A. 3.13

Model stability parameters for Model 5, exploring whether DCCS score predicts 4-year-old children's sensitivity to normative expectations in girls and boys from the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.23	0.97	1.49
AO(1)	0.12	-0.05	0.32
DCCS(2)	-0.37	-0.83	-0.08
Gender(3)	-1.32	-1.77	-0.94
Cultural Group(4)	-1.43	-1.89	-1.23
AO*DCCS	0.04	-0.15	0.40
AO*Gender	0.19	-0.08	0.58
DCCS*Gender	0.83	0.43	1.31
AO*DCCS*Gender	0.16	-0.19	0.42
Pariticpant ID@Intercept@NA	1.88	1.76	2.19
Experimenter on Trial@Intercept@NA	0.26	0.06	0.44
Experimenter on Trial@Gender@NA	0.40	0.03	0.97
Experimenter on Trial@AO@NA	0.24	0.09	0.46
Experimenter on Trial@DCCS@NA	0.57	0.40	0.83
Experimenter on Trial@Intercept@Gender	0.48	-0.94	0.87
Experimenter on Trial@Intercept@AO	-0.22	-0.69	0.70
Experimenter on Trial@Intercept@DCCS	-0.66	-0.90	0.14

Term	Original	Min	Max
Experimenter on Trial@Gender@AO	-0.32	-1.00	0.53
Experimenter on Trial@Gender@DCCS	0.23	-0.55	0.76
Experimenter on Trial@AO@DCCS	-0.40	-0.75	0.90
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order @Gender@NA	0.00	0.00	0.04
Condition Order @Cultural Group@NA	0.00	0.00	0.03
Condition Order @AO@NA	0.10	0.00	0.20
Condition Order@DCCS@NA	0.00	0.00	0.03

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) dummy coded with Fail being the reference level

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

Table A. 3.14

Results of full Model 5 (Model A in Table A.3.12) examining whether 4-year-old children's DCCS performance predicts sensitivity to normative expectations in girls and boys from the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.86	0.43	0.09	1.65	1.99	.046
AO[(1)]	0.09	0.22	-0.35	0.55	0.41	.681
DCCS[(2)]	0.40	0.99	-1.54	2.34	0.40	.689
Gender[(3)]	-1.25	0.64	-2.39	-0.07	-1.96	.050
Cultural Group[(4)]	-0.69	0.68	-1.97	0.50	-1.03	.304
AO*DCCS	-0.16	0.56	-1.25	1.13	-0.29	.775
AO*Gender	0.07	0.34	-0.52	0.69	0.22	.830
DCCS*Gender	2.86	1.59	-0.10	8.86	1.80	.071
AO*Cultural Group	0.01	0.35	-0.71	0.72	0.02	.984
DCCS*Cultural Group	-1.28	1.20	-3.71	1.05	-1.06	.288
Gender*Cultural Group	0.16	1.10	-1.93	2.40	0.14	.885
AO*DCCS*Gender	1.21	1.22	-1.23	9.27	0.99	.323
AO*DCCS*Cultural Group	0.29	0.66	-1.13	1.65	0.45	.654
AO*Gender*Cultural Group	0.29	0.57	-0.90	1.42	0.50	.617
DCCS*Gender*Cultural Group	-2.60	1.93	-8.86	1.25	-1.35	.178
AO*DCCS*Gender*Cultural Group	-1.33	1.36	-9.79	1.42	-0.98	.328

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) dummy coded with Fail being the reference level

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

Table A. 3.15

Results of reduced Model 5 (Model B in Table A.3.12) examining whether 4-year-old children's DCCS performance predicts sensitivity to normative expectations in girls and boys from the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.87	0.43	0.08	1.67	2.04	.042
AO[(1)]	0.06	0.23	-0.44	0.56	0.25	.805
DCCS[(2)]	0.34	0.99	-1.50	2.23	0.35	.729
Gender[(3)]	-1.29	0.63	-2.58	-0.09	-2.03	.043
Cultural Group[(4)]	-0.73	0.67	-2.00	0.53	-1.08	.278
AO*DCCS	0.04	0.52	-0.93	1.19	0.07	.943
AO*Gender	0.18	0.33	-0.46	0.84	0.54	.591
DCCS*Gender	3.07	1.57	0.04	7.63	1.95	.051
AO*Cultural Group	0.11	0.34	-0.56	0.80	0.32	.749
DCCS*Cultural Group	-1.19	1.19	-3.41	1.12	-1.00	.319
Gender*Cultural Group	0.24	1.09	-2.06	2.34	0.22	.824
AO*DCCS*Gender	0.15	0.51	-0.92	1.23	0.30	.767
AO*DCCS*Cultural Group	-0.01	0.57	-1.27	1.06	-0.02	.986
AO*Gender*Cultural Group	0.02	0.50	-1.08	1.05	0.04	.965
DCCS*Gender*Group	-2.88	1.90	-7.49	0.70	-1.51	.130

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with Fail being the reference level

[(3)] dummy coded with girls being the reference level

[(4)] dummy coded with the Ugandan sample being the reference level

Table A. 3.16

Results of reduced Model 5 (Model C in Table A.3.12) examining whether 4-year-old children's DCCS performance predicts sensitivity to normative expectations in girls and boys from the UK and Uganda (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.75	0.43	-0.07	1.59	1.77	.077
AO[(1)]	0.03	0.20	-0.38	0.40	0.17	.867
DCCS[(2)]	1.06	0.87	-0.51	3.04	1.21	.225
Gender[(3)]	-1.02	0.61	-2.26	0.22	-1.66	.096
Cultural Group[(4)]	-0.41	0.64	-1.60	0.86	-0.63	.527
AO*DCCS	0.08	0.23	-0.45	0.58	0.37	.714
AO*Gender	0.25	0.20	-0.16	0.66	1.22	.221
DCCS*Gender	0.11	0.26	-0.39	0.65	0.42	.673
AO*Cultural Group	1.13	0.89	-0.63	2.92	1.27	.204
DCCS*Cultural Group	-2.25	0.98	-4.41	-0.52	-2.31	.021
Gender*Cultural Group	-0.64	0.92	-2.55	1.21	-0.69	.490

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with Fail being the reference level

[(3)] dummy coded with girls being the reference level

[(4)] dummy coded with the Ugandan sample being the reference level

3.2.1.3 Additional Details Regarding the Process of Reaching the Final Model: Model 6 (RQ4)

To check that a relationship between DCCS performance and sensitivity to normative expectations was not observed in Model 5 due to a lack of statistical power, we ran an additional analysis which only tested the interaction between 'Approval Observe condition' and 'DCCS'. Therefore, to analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions) and DCCS (Fail or Pass), these categorical predictors and their interaction were entered into a logistic Generalized Linear Mixed Model (Baayen, 2008). Gender (girls or boys) and cultural group (UK or Uganda) were included as control variables, and as such, were included in both the full and null models.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating high instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -858.57, df = 42; model without correlation parameters: -862.96, df = 26). The model stability parameters for Model 6 can be viewed in Table A.3.17.

As the full-null model comparison was non-significant (see Results), as described in the Statistical Approach section, we can assume that the Approval Observe condition, performance on the DCCS, or the interaction of them two variables do not significantly affect the proportion of tokens children give.

Table A. 3.17

Model stability parameters for Model 6, exploring whether DCCS score predicts 4-year-old children's sensitivity to normative expectations, controlling for gender and cultural group, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.98	0.79	1.26
AO[(1)]	0.16	0.00	0.26
DCCS[(2)]	-0.09	-0.42	0.28
Gender[(3)]	-0.89	-1.29	-0.68
Cultural Group[(4)]	-1.29	-1.75	-1.11
AO*DCCS	0.15	0.02	0.47
Pariticpant ID@Intercept@NA	1.88	1.72	2.21
Experimenter on Trial@Intercept@NA	0.21	0.00	0.34
Experimenter on Trial@Gender@NA	0.45	0.00	0.88
Experimenter on Trial@AO@NA	0.21	0.00	0.34
Experimenter on Trial@DCCS@NA	0.59	0.36	0.76
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order @Gender@NA	0.00	0.00	0.04
Condition Order @Cultural Group@NA	0.00	0.00	0.03
Condition Order @AO@NA	0.06	0.00	0.15
Condition Order@DCCS@NA	0.00	0.00	0.03

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with Fail being the reference level

[(3)] dummy coded with girls being the reference level

[(4)] dummy coded with the Ugandan sample being the reference level

3.2.2 Additional Analyses with Only UK Data

3.2.2.1 Exploring Normative Sharing and DCCS Performance in the UK Participants

In order to check the robustness of the results for the UK sample, the analyses for Model 5 and 6 were rerun with only UK participants, to check that an absence of a relationship between normative sharing and performance on the DCCS in the UK children was not being unduly influenced by noise in the Ugandan DCCS data. The sample included the 57 UK children (34 girls), meaning the sample consisted of 228 proportions of tokens given and kept from 57 participants. The same statistical approach was used as in the main text.

3.2.2.2 Results of Model 5 with Only UK Participants

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), DCCS (Fail or Pass) and gender (girls or boys), these categorical predictors and their interaction were entered into a logistic Generalized Linear Mixed Model (Baayen, 2008).

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating high instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -464.57, df = 29; model without correlation parameters: -468.00, df = 17). The model stability parameters for Model 5 can be viewed in Table A.3.18.

Overall, the full model did not explain significantly more variation in the proportion of tokens given to the other child than the null model (likelihood ratio test: $\chi 2 = 12.17$, df = 7, p = .095). This further suggests that, even when removing noise from the Ugandan participants, UK children's cognitive flexibility, as measured by performance on the DCCS task, is not related to their ability to navigate normative expectations.

Table A. 3.18

Model stability parameters for Model 5, examining whether DCCS score predicts 4-year-old girl's and boy's sensitivity to normative expectations in the UK sample only, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.14	-0.21	0.72
AO[(1)]	0.02	-0.21	0.28
DCCS[(2)]	-0.81	-1.44	-0.39
Gender[(3)]	-0.93	-1.68	-0.16
AO*DCCS	0.16	-0.05	0.44
AO*Gender	0.37	0.06	0.90
DCCS*Gender	0.23	-0.43	0.98
AO*DCCS*Gender	-0.13	-0.51	0.15
Pariticpant ID@Intercept@NA	1.06	0.94	1.18
Experimenter on Trial@Intercept@NA	0.18	0.00	0.28
Experimenter on Trial@Gender@NA	0.27	0.00	0.68
Experimenter on Trial@AO@NA	0.00	0.00	0.00
Experimenter on Trial@DCCS@NA	0.40	0.00	0.66
Condition Order@Intercept@NA	0.00	0.00	0.17
Condition Order @Gender@NA	0.13	0.00	0.52
Condition Order @AO@NA	0.00	0.00	0.20
Condition Order@DCCS@NA	0.00	0.00	0.34

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with Fail being the reference level

[(3)] dummy coded with girls being the reference level

3.2.2.3 Results of Model 6 with Only UK Participants

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions) and DCCS (Fail or Pass), these categorical predictors and their interaction were entered into a logistic Generalized Linear Mixed Model (Baayen, 2008). Gender (girls or boys) was included as a control variable, and as such, were included in both the full and null models.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating high instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -465.29, df = 26; model without correlation parameters: -468.75, df = 14). The model stability parameters for Model 6 can be viewed in Table A.3.19.

Overall, the full model explained significantly more variation in the proportion of tokens given to the other child than the null model, which contained 'gender' (likelihood ratio test: $\chi 2 = 8.06$, df = 3, $p = .045$). In the process of generating the final model (i.e. dropping non-significant interaction terms), the 2-way interaction between 'Approval Observe condition' and 'DCCS performance' was non-significant (Table A.3.20). Therefore, a final model was constructed with only the main effects. The model revealed a significant effect of 'gender' (Table A.3.21), which suggested the girls in the UK were more generous than boys. There was also a significant effect of 'Approval Observe condition' (Table A.3.21), replicating the previous finding of 4-year-old children showing sensitivity to normative expectations in their sharing behaviour.

Table A. 3.19

Model stability parameters for Model 6, examining whether DCCS score predicts 4-year-old children's sensitivity to normative expectations in the UK sample only, controlling for gender, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.08	-0.13	0.46
AO[(1)]	0.19	0.04	0.33
DCCS[(2)]	-0.74	-1.09	-0.45
Gender[(3)]	-0.71	-0.93	-0.38
AO*DCCS	0.13	-0.06	0.33
Pariticpant ID@Intercept@NA	1.06	0.94	1.18
Experimenter on Trial@Intercept@NA	0.18	0.00	0.26
Experimenter on Trial@Gender@NA	0.31	0.00	0.64
Experimenter on Trial@AO@NA	0.00	0.00	0.02
Experimenter on Trial@DCCS@NA	0.41	0.00	0.64
Condition Order@Intercept@NA	0.00	0.00	0.17
Condition Order @Gender@NA	0.12	0.00	0.51
Condition Order @AO@NA	0.00	0.00	0.20
Condition Order@DCCS@NA	0.00	0.00	0.33

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] dummy coded with Fail being the reference level

[(3)] dummy coded with girls being the reference level

Table A. 3.20

Results of full Model 6 examining whether 4-year-old children's DCCS performance predicts sensitivity to normative expectations, controlling for gender, in just the UK sample (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.08	0.32	-0.52	0.69	0.25	.804
AO[(1)]	0.19	0.19	-0.21	0.59	1.00	.315
DCCS[(2)]	-0.74	0.39	-1.44	-0.04	-1.97	.049
Gender[(3)]	-0.71	0.34	-1.40	-0.02	-2.09	.037
AO*DCCS	0.13	0.24	-0.37	0.64	0.54	.588

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level
[(2)] dummy coded with Fail being the reference level
[(3)] dummy coded with girls being the reference level

Table A. 3.21

Results of reduced Model 6 examining whether 4-year-old children's DCCS performance predicts sensitivity to normative expectations, controlling for gender, in just the UK sample (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.06	0.31	-0.54	0.70	0.18	.856
AO[(1)]	0.27	0.12	0.03	0.50	2.32	.020
DCCS[(2)]	-0.71	0.37	-1.49	0.01	-1.91	.056
Gender[(3)]	-0.71	0.34	-1.42	-0.04	-2.09	.037

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level
[(2)] dummy coded with Fail being the reference level
[(3)] dummy coded with girls being the reference level

Appendix 4: Supplementary Materials for Chapter 4

4.1 Appendix for Part 1

4.1.1 Materials and Procedures

4.1.1.1 Auditory Joint Attention Stimuli

The sources of the Auditory Joint Attention stimuli can be viewed in Table A.4.1. These were sounds made by (i) a peacock, (ii) an orca or (iii) a zebra.

Table A. 4.1
Sources of sound files for the Auditory Joint Attention task.

Animal	Sources Used
Peacock	http://www.wavsource.com/animals/animals2.htm#
	www.soundrangers.com
	seh-audio.s3.amazonaws.com/peacock_sounds_long.mp3
	http://soundbible.com/tags-animal.html
	http://audiosoundclips.com/wp-content/uploads/2011/12/Peacock.mp3
Orca	http://www.bcu.ubc.ca/~deecke.aiff
	http://dobieaquaticscience.weebly.com/uploads/2/4/8/1/24814097/sound_7.mp3
	http://thestablesbb.powweb.com/WAV%20FILES/orca_1.wav
	http://www24.big.or.jp/~kyusoku/sound/int_orca.wav
	http://www.scuba.lu/sound/ngsorca03.wav
	http://soundbible.com/228-Orca-Whales-Talking.html
	https://freesound.org/people/mjudo12/sounds/74908/
Zebra	http://jansjoyousjungle.com/zebra.wav
	https://www.wildlifelands.com/wxs/Zebra.mp3
	http://www.billybear4kids.com/animal/sounds/zebra.wav
	www.soundrangers.com
	http://bbcsfx.acropolis.org.uk/

4.1.1.2 Coding Scheme

The full coding scheme for the Auditory Joint Attention task can be viewed in Table A.4.2.

Table A. 4.2
Full Coding Scheme for the Auditory Joint Attention Task.

Variable Coded	Options	Definition
Experimental Period	Not Coded	Default state. Also turned on if there were any distractions (e.g. infant is distressed, infant leaves testing area, third party interacting with one or both of the dyad) and at the end of the video (i.e. 30 seconds after the last sound has played or when the experimenter says they are finished)
	Control	Period before the first stimulus has been presented, but after the video has been introduced and the mother is no longer paying attention to any experimenter
	Experiment	From the start of the first sound stimulus. This is then left on until the end of the trial, unless a Not Coded period or Experiment - Loud Background Noise is necessary, in which case, this shall be resumed from the beginning of the next stimulus.
	Experiment - Loud Background Noise	Used in cases there are loud noises playing over the sounds, as long as the stimuli can be heard from both camera angles (and so we assume the participants also could hear the stimuli). If used, code from when the stimulus starts to when it ends. If the sound cannot be heard from both camera angles, code the period as Not coded (as it could be possible the participants also didn't hear the sound)

Variable Coded	Options	Definition
Stimulus Presence	Stimulus not present	Default state. Indicates no sound is being played
	Stimulus present	Sound is being played. Modifier includes the stimulus number
Gaze Direction (Infant)	Elsewhere	Default state. Eyes looking anywhere but stimulus, partner's face, experimenter or other individual. Can be coded even if you cannot see eyes, as long as there is no possibility the infant can be looking at anything of interest. Only code if >1 second
	Stimulus	Eyes looking towards the speaker or in the direction of sound. If the speaker cannot be seen, or there is no commentary about the speaker location, do not use this option during coding
	Partner	Eyes looking at partner's face
	Experimenter	Eyes looking at the experimenter or directly at, or slightly above the camera (which has an experimenter behind)
	Other individual	Eyes looking at individual that is not the partner or an experimenter
	Unclear	The focal individual's eyes are not visible, so unable to see where they're looking and their head direction means they might be looking at stimulus, partner's face, experimenter or other individual. Only code when this is the case for >3s, except

Variable Coded	Options	Definition
		when there is a possibility the infant is looking towards the partner's face. Also use in cases where is it unclear where
Gaze Direction (Partner)	Identical to Gaze Direction (Infant)	
Mutual Gaze	No mutual gaze	Default state.
	Mutual gaze	Mother's eyes and infant's eyes are looking at each other's faces. Only coded during mutual gaze codable periods (i.e. duration of stimulus and 3 seconds after its offset)
Infant Communication	Communication not coded	Default state. All cases outside the communication codable period.
	No infant communication (not all modalities available)	Within a communication period, the infant does not produce a communicative behaviour (vocalisation, gesture, facial expression), but at least one modality is not available for coding of communication (the infant's voice is not audible, and/or their face and/or hands are not visible)
	No infant communication (all modalities available)	Within a communication period, the infant does not produce a communicative behaviour (vocalisation, gesture, facial expression) and all modalities are available

Variable Coded	Options	Definition
		for coding of communication (the infant's voice is audible, and their face and hands are visible)
	Infant Vocal Communication	Any audible signal that comes out of the infant's mouth that is not a non-volitional signal (non-volitional = coughs, sneezes, burps, yawns). We are only really interested in signals directed at the partner – do not code signals that are clearly directed to a non-partner (these are flagged, see below). To be directed at a partner, they need to have looked to their partner (face direction or gaze) in the 3s before, during or after the signal. Code all vocalisations no matter how short but use a 3 second rule on breaks in expressions i.e. if an individual's vocalising stops for less than 3 seconds then code it as continuous
	Infant Gesture Pointing	Pointing (with arm and/or index finger extended) at some interesting object or event. If child touches object (poking) while directing attention to, that can also count as pointing (but not when they are touching it to explore the object itself or are going to grab it)

Variable Coded	Options		Definition
		Reaching for something out of reach	Arm extended, hand open, palm facing down OR hand opening and closing, directed towards person or object
		Requesting reach	Arm extended, palm open, upwards, requesting object
		Give/offer as a request	Reaches out and gives a toy or some object that he/she is holding. Eye contact with partner and partner takes it (not just dropping an object on some one or pushing object towards someone). Infant hands object to another. There is movement of arms and object changes hands.
		Raising arms to be carried; “Up”	Extends arm upwards to signal wish to be picked up. One or both arms are raised toward another or merely moved away from the body to allow room for another’s hands to pick the infant up. (The latter is usually in response to another’s motion toward the infant with hands out.)
		Show	Extend arm to show object in hand
		Nod	Affirmative head nod
		Head shake	Shakes head “no”

Variable Coded	Options		Definition
		Clap hands	Clap hands
		Shrugs	Raises and lowers shoulders to indicate “all gone” or “where did it go”
		Raising palm upwards	To indicate “all gone” or emphasize a question. Can be one or both hands
		Waving	Either with open/closed hand alternating or waving open hand from side to side
		Blow kisses	Blow kisses
		Tactile	Includes hitting, slapping or grab-pulling (tug) which should be delivered with some force (must be salient) or occur repeatedly. Signaller should have eye-contact with the recipient when producing the gesture. Also includes embrace (tactile; hugging other with both arms), nuzzle (rub head/face on other in affectionate way) or kiss
	Infant Facial Expression	Smiling	No coding of closed mouth smiling, only open mouth smiling.
		Surprise	Only eyebrows up, open smile or ‘oo’ mouth
		Sad	Lower lip out or frown

Variable Coded	Options		Definition
		Blend of joy and surprise	Often associated with gasping - involves open mouth, smile and eyebrows raised
		Pouting	Used as if to direct attention
	Infant Combination of Signals		If there is a mix of the above three categories of communicative behaviours (vocalisation, gesture or facial expression), regardless of whether they overlap or are in sequence, apply this code
Partner Communication	Partner Vocal communication		Identical to Infant Vocal Communication, except with the inclusion of verbal utterances
	Partner Gesture	Point	Finger extended towards an object or event
		Indicate	Object or event was singled out through movements of the head or hand towards the object or event, or direct contact with (i.e. tapping or touching) the object or location of the event
		Beckoning	Beckoning with either the whole hand or the index finger
		Raising palm upwards	To indicate “all gone” or emphasise a question. Can be one or both hands
		Request reach	Requesting an object. Reaching out with their hand

Variable Coded	Options	Definition
	Shaking head	Shaking the head from side to side to indicate ‘no’
	Nodding the head	Nodding the head up and down to indicate ‘yes’
	Clap hands	Clap hands
	Shrugs	Raises and lowers shoulders to indicate “all gone” or “where did it go”
	Waving	Either with open/closed hand alternating or waving open hand from side to side
	Blow kisses	Blow kisses
	Tactile	Includes hitting, slapping or grab-pulling (tug) which should be delivered with some force (must be salient) or occur repeatedly. Signaller should have eye-contact with the recipient when producing the gesture. Also includes embrace (tactile; hugging other with both arms), nuzzle (rub head/face on other in affectionate way) or kiss
	Smiling	No coding of closed mouth smiling, only open mouth smiling.
	Surprise	Only eyebrows up, open smile or ‘oo’ mouth

Variable Coded	Options		Definition
	Partner Facial Expression	Sad	Lower lip out or frown

4.1.1.3 Interobserver Reliability for the Joint Attention Measure

To begin, the anchor coder developed a coding scheme and trained three research assistants on the coding scheme using 4 videos. The lead author then coded 20.00% of videos (N = 18 / 90). These videos were double coded by three research assistants (6 videos each). Cohen's Kappa's for the items used in the analysis are reported in Table A.4.3, along with the percentage of agreement between the coders. In terms of how the IOR was calculated, frequency / sequence refers to whether the coded variables were coded in the same sequence between two coders. Duration / sequence refers to whether the durations of coded variables were coded in the same sequence between two coders. Cohen's Kappa's ranged from 0.72-1.00 across coders in comparison to the anchor coder, indicating excellent reliability.

Table A. 4.3

The Cohen's Kappa's and percentage of agreement for each coder when compared to the anchor coder's coding of coded variables.

Coded Variable	IOR Calculation	Coder 1		Coder 2		Coder 3	
		Cohen's Kappa	Percentage Agreement	Cohen's Kappa	Percentage Agreement	Cohen's Kappa	Percentage Agreement
Experimental Period	Frequency / Sequence	1.00	100%	0.90	94.00%	0.92	94.00%
	Duration / Sequence	0.98	99.21%	0.93	96.93%	0.93	97.05%
Stimulus Presence & Infant Gaze to Partner's Face	Frequency / Sequence	0.85	88.85%	0.88	91.32%	0.91	93.26%
	Duration / Sequence	0.97	98.26%	0.95	97.12%	0.97	98.28%
Stimulus Presence & Partner Gaze to Infant's Face	Frequency / Sequence	0.83	87.53%	0.72	79.49%	0.75	81.24%
	Duration / Sequence	0.94	96.11%	0.89	92.11%	0.94	96.02%

4.1.2 Statistical Approach and Analyses

4.1.2.1 Additional Details Regarding the Process of Reaching the Final Model: Model 1 (RQ1)

To analyse how joint attention events at 15 months varied with gender and cultural group, the number of joint attention events coded were included as a response in the model, with gender (girls or boys), cultural group (UK or Uganda) and the interaction included into the model as categorical predictors. The total time in which joint attention could have occurred was included as an offset term into the model (which was a continuous variable). The model stability parameters for Model 1 can be viewed in Table A.4.4.

Table A. 4.4

Model stability parameters for Model 1, exploring whether gender or cultural group predict joint attention at 15 months, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	-3.98	-4.39	-3.91
Gender[(2)]	-0.64	-0.86	-0.33
Cultural Group[(3)]	0.27	0.07	0.59
Gender*Cultural Group	0.55	0.24	0.78

[(1)] dummy coded with girls being the reference level

[(2)] dummy coded with the Ugandan sample being the reference level

4.1.2.2 Additional Details Regarding the Process of Reaching the Final Model: Model 2 (RQ2)

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), gender (girls or boys) and the rate of joint attention events, (as measured by the 'shared attention' variable). This was derived from dividing the number of shared attention events (triadic attention with both partners communicating) by the time in seconds in which shared attention could have occurred. A logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these three variables and their 3-way interaction included as fixed effects into the model. Cultural group was included as a fixed control effect (and as such, was also included in the null model).

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -552.80, df = 35; model without correlation parameters: -560.92, df = 19). The model stability parameters for Model 2 can be viewed in Table A.4.5.

The non-significant full-null comparison indicates that none of the predictor variables (nor their interactions) are having a significant influence on the proportion of tokens given. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.5

Model stability parameters for Model 2, exploring whether the rate of shared attention events at 15 months predicts 4-year-old children's sensitivity to normative expectations in girls and boys, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.75	0.74	2.73
AO[(1)]	0.20	0.06	0.49
Shared Attention[(2)]	-0.15	-0.29	-0.01
Gender[(3)]	-1.18	-1.60	-0.41
Cultural Group[(4)]	-1.61	-2.21	-0.98
AO*Shared Attention	-0.16	-0.30	0.00
AO*Gender	0.34	-0.02	0.55
Shared Attention*Gender	-0.15	-0.50	0.20
AO*Shared Attention*Gender	0.09	-0.06	0.41
Pariticpant ID@Intercept@NA	2.27	2.00	2.87
Experimenter on Trial@Intercept@NA	0.00	0.00	0.25
Experimenter on Trial@AO@NA	0.12	0.00	0.63
Experimenter on Trial@Shared Attention@NA	0.25	0.17	0.36
Experimenter on Trial@Gender@NA	0.50	0.00	0.82
Condition Order@Intercept@NA	0.00	0.00	0.04
Condition Order@AO@NA	0.41	0.00	0.57

Term	Original	Min	Max
Condition Order@Shared Attention@NA	0.00	0.00	0.02
Condition Order@Gender@NA	0.50	0.00	0.93
Condition Order@Cultural Group@NA	0.00	0.00	0.91

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) rate of shared attention at 15 months (both partners communicating)

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

4.1.2.3 Additional Details Regarding the Process of Reaching the Final Model: Model 3 (RQ2)

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), gender (girls or boys) and the rate of joint attention events, (as measured by the 'triadic attention with communication' variable). This was derived from dividing the number of triadic attention events (in which one partner communicated) by the time in seconds in which triadic attention with communication could have occurred. A logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these three variables and their 3-way interaction included as fixed effects into the model. Cultural group was included as a fixed control effect (and as such, was also included in the null model).

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -555.37, df = 35; model without correlation parameters: -561.49, df = 19). The model stability parameters for Model 3 can be viewed in Table A.4.6.

The non-significant full-null comparison indicates that none of the predictor variables (nor their interactions) are having a significant influence on the proportion of tokens given. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.6

Model stability parameters for Model 3, exploring whether the rate of triadic attention with communication events at 15 months predicts 4-year-old children's sensitivity to normative expectations in girls and boys, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.76	0.87	2.61
AO[(1)]	0.22	0.02	0.55
Triadic Attention w/Comm[(2)]	-0.06	-0.15	-0.01
Gender[(3)]	-1.28	-1.72	-0.69
Cultural Group[(4)]	-1.67	-2.18	-1.09
AO* Triadic Attention w/Comm	-0.09	-0.18	0.00
AO*Gender	0.40	-0.06	0.62
Triadic Attention w/Comm *Gender	-0.02	-0.15	0.17
AO* Triadic Attention w/Comm *Gender	0.04	-0.05	0.24
Pariticpant ID@Intercept@NA	2.27	1.99	2.96
Experimenter on Trial@Intercept@NA	0.00	0.00	0.25
Experimenter on Trial@AO@NA	0.05	0.00	0.51
Experimenter on Trial@ Triadic Attention w/Comm@NA	0.13	0.08	0.18
Experimenter on Trial@Gender@NA	0.39	0.00	0.74
Condition Order@Intercept@NA	0.00	0.00	0.06
Condition Order@AO@NA	0.42	0.00	0.58
Condition Order@ Triadic Attention w/Comm@NA	0.00	0.00	0.02

Term	Original	Min	Max
Condition Order@Gender@NA	0.52	0.03	0.90
Condition Order@Cultural Group@NA	0.00	0.00	0.57

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) rate of triadic attention with communication at 15 months (one partner communicating)

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

4.1.2.4 Additional Details Regarding the Process of Reaching the Final Model: Model 4 (RQ2)

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), gender (girls or boys) and the rate of joint attention events, (as measured by the 'triadic attention' variable). This was derived from dividing the number of triadic attention events (in which no partners communicated) by the time in seconds in which triadic attention could have occurred. A logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these three variables and their 3-way interaction included as fixed effects into the model. Cultural group was included as a fixed control effect (and as such, was also included in the null model).

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -555.31, df = 35; model without correlation parameters: -561.97, df = 19). The model stability parameters for Model 4 can be viewed in Table A.4.7.

The non-significant full-null comparison indicates that none of the predictor variables (nor their interactions) are having a significant influence on the proportion of tokens given. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.7

Model stability parameters for Model 4, exploring whether the rate of triadic attention events at 15 months predicts 4-year-old children's sensitivity to normative expectations in girls and boys, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.84	0.92	2.61
AO[(1)]	0.22	0.04	0.63
Triadic Attention[(2)]	-0.07	-0.14	-0.02
Gender[(3)]	-1.30	-1.80	-0.68
Cultural Group[(4)]	-1.72	-2.20	-1.11
AO*Triadic Attention	-0.08	-0.14	0.00
AO*Gender	0.37	-0.09	0.60
Triadic Attention*Gender	-0.01	-0.20	0.17
AO*Triadic Attention*Gender	0.04	-0.03	0.24
Pariticpant ID@Intercept@NA	2.27	1.99	2.95
Experimenter on Trial@Intercept@NA	0.00	0.00	0.22
Experimenter on Trial@AO@NA	0.09	0.00	0.52
Experimenter on Trial@ Triadic Attention@NA	0.11	0.07	0.16
Experimenter on Trial@Gender@NA	0.41	0.00	0.70
Condition Order@Intercept@NA	0.00	0.00	0.10
Condition Order@AO@NA	0.45	0.00	0.60
Condition Order@ Triadic Attention@NA	0.00	0.00	0.02

Term	Original	Min	Max
Condition Order@Gender@NA	0.52	0.00	0.97
Condition Order@Cultural Group@NA	0.00	0.00	0.67

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) rate of triadic attention at 15 months (no partner communicating)

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

4.1.2.5 Additional Details Regarding the Process of Reaching the Final Model: Model 5 (RQ2)

To check that a relationship between the rate of joint attention events and sensitivity to normative expectations was not observed in Model 2 due to a lack of statistical power, we ran an additional analysis which only tested the interaction between 'Approval Observe condition' and 'shared attention'. Therefore, to analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions) and the rate of joint attention events (as measured by the 'shared attention' variable), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these two variables and their interaction included as fixed effects into the model. Gender and cultural group were included as fixed control factors and as such, were included in the null model.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -554.40, df = 32; model without correlation parameters: -562.44, df = 16). The model stability parameters for Model 4 can be viewed in Table A.4.8.

The non-significant full-null comparison indicates that neither the 'Approval Observe condition' or 'shared attention' (nor their interactions) are having a significant influence on the proportion of tokens given. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.8

Model stability parameters for Model 5, exploring whether the rate of shared attention events at 15 months predicts 4-year-old children's sensitivity to normative expectations, controlling for gender and cultural group, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.76	0.81	3.01
AO[(1)]	0.35	0.24	0.50
Shared Attention[(2)]	-0.19	-0.42	-0.10
Gender[(3)]	-1.22	-1.99	-0.53
Cultural Group[(4)]	-1.63	-2.32	-1.03
AO*Shared Attention	-0.15	-0.23	-0.03
Pariticpant ID@Intercept@NA	2.26	2.01	2.88
Experimenter on Trial@Intercept@NA	0.00	0.00	0.21
Experimenter on Trial@AO@NA	0.00	0.00	0.60
Experimenter on Trial@Shared Attention@NA	0.26	0.18	0.39
Experimenter on Trial@Gender@NA	0.52	0.00	0.84
Condition Order@Intercept@NA	0.00	0.00	0.14
Condition Order@AO@NA	0.46	0.00	0.61
Condition Order@ Shared Attention@NA	0.00	0.00	0.02
Condition Order@Gender@NA	0.55	0.00	0.96
Condition Order@Cultural Group@NA	0.00	0.00	0.81

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) rate of shared attention at 15 months (both partners communicating)

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

4.2 Appendix for Part 2

4.2.1 Statistical Analyses

4.2.1.1 Additional Details Regarding the Process of Reaching the Final Model: Model 6 (RQ3)

To analyse how overimitation at 18 months varied with gender and cultural group, whether overimitation was present (Yes or No) was included as a response in the model, with gender (girls or boys), cultural group (UK or Uganda) and the interaction included into the model as categorical predictors. The model stability parameters for Model 6 can be viewed in Table A.4.9. There was no significant interaction between cultural group and gender (Table A.4.10), so a reduced model with just the main effects was run (described in the Results section).

Table A. 4.9

Model stability parameters for Model 6, exploring whether gender or cultural group predict overimitation at 18 months, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	-2.77	-2.87	-2.14
Cultural Group(1)	3.09	2.46	3.21
Gender(2)	0.83	0.19	1.24
Gender*Cultural Group	-0.86	-1.27	-0.23

(1) dummy coded with the Ugandan sample being the reference level

(2) dummy coded with girls being the reference level

Table A. 4.10

Results of full Model 6 examining whether gender and cultural group predict overimitation at 18 months (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	-2.77	1.03	-5.67	-1.18	-2.69	.007
Cultural Group[(1)]	3.09	1.13	1.22	6.09	2.73	.006
Gender[(2)]	0.83	1.28	-1.62	3.96	0.65	.518
Gender*Cultural Group	-0.86	1.43	-4.18	1.88	-0.60	.549

[(1)] dummy coded with the Ugandan sample being the reference level

[(2)] dummy coded with girls being the reference level

4.2.1.2 Additional Details Regarding the Process of Reaching the Final Model: Model 7 (RQ4)

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), gender (girls or boys), overimitation presence ('Yes' or 'No') and cultural group (UK or Uganda), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these four variables and their 4-way interaction included as fixed effects into the model.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -562.36, df = 342; model without correlation parameters: -566.53, df = 26). The model stability parameters for Model 7 can be viewed in Table A.4.11.

The non-significant full-null comparison indicates that none of the predictor variables (nor their interactions) are having a significant influence on the proportion of tokens given. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.11

Model stability parameters for Model 7, exploring whether the presence of overimitation at 18 months predicts 4-year-old children's sensitivity to normative expectations in girls and boys from the UK and Uganda, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.35	0.69	2.05
AO[(1)]	-0.15	-0.76	0.61
Overimitation[(2)]	-1.90	-2.27	0.14
Gender[(3)]	-1.32	-1.86	-0.33
Cultural Group[(4)]	-1.55	-2.42	-0.60
AO*Overimitation	-0.35	-3.15	0.43
AO*Gender	0.82	-0.36	1.58
Overimitation*Gender	3.67	1.66	5.20
AO*Cultural Group	0.54	-0.48	1.35
Overimitation*Cultural Group	1.93	-0.13	2.66
Gender*Cultural Group	-0.89	-1.98	0.26
AO*Overimitation*Gender	0.18	-0.70	0.80
AO*Overimitation*Cultural Group	-0.23	-0.87	2.69
AO*Gender*Cultural Group	-0.33	-1.23	1.05
Overimitation*Gender*Cultural Group	-2.02	-3.65	-0.92
AO*Overimitation*Gender*Cultural Group	-0.18	-1.11	0.63
Participant ID@Intercept@NA	2.13	1.94	2.64

Term	Original	Min	Max
Experimenter on Trial@Intercept@NA	0.10	0.00	0.34
Experimenter on Trial@sex.M@NA	0.29	0.00	0.69
Experimenter on Trial@AO@NA	0.00	0.00	0.27
Experimenter on Trial@Overimitation@NA	0.21	0.01	0.41
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order@Gender@NA	0.45	0.00	1.15
Condition Order@Cultural Group@NA	0.00	0.00	0.03
Condition Order@AO@NA	0.47	0.00	0.57
Condition Order@Overimitation@NA	0.00	0.00	0.02

(1) Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

(2) the presence of overimitation, dummy coded with No being the reference level

(3) dummy coded with girls being the reference level

(4) dummy coded with the Ugandan sample being the reference level

4.2.1.3 Additional Details Regarding the Process of Reaching the Final Model: Model 8 (RQ4)

To check that a relationship between overimitation and sensitivity to normative expectations was not observed in Model 7 due to a lack of statistical power, we ran an additional analysis which only tested the interaction between 'Approval Observe condition' and 'overimitation'. Therefore, to analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions) and the presence of Overimitation (Yes or No), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these two variables and their interaction included as fixed effects into the model. Gender and cultural group were included as fixed control factors and as such, were included in the null model.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -568.47, df = 32; model without correlation parameters: -570.75, df = 16). The model stability parameters for Model 8 can be viewed in Table A.4.12.

The non-significant full-null comparison indicates that neither the 'Approval Observe condition' or 'overimitation' (nor their interactions) are having a significant influence on the proportion of tokens given. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.12

Model stability parameters for Model 8, exploring whether overimitation at 18 months predicts 4-year-old children's sensitivity to normative expectations, controlling for gender and cultural group, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	1.21	0.73	1.93
AO[(1)]	0.24	0.02	0.50
Overimitation[(2)]	0.71	0.13	1.05
Gender[(3)]	-0.96	-1.29	-0.57
Cultural Group[(4)]	-1.89	-2.65	-1.50
AO*Overimitation	-0.18	-0.41	0.16
Pariticpant ID@Intercept@NA	2.19	2.01	2.84
Experimenter on Trial@Intercept@NA	0.00	0.00	0.19
Experimenter on Trial@Gender@NA	0.60	0.00	1.00
Experimenter on Trial@AO@NA	0.00	0.00	0.53
Experimenter on Trial@Overimitation@NA	0.23	0.00	0.35
Condition Order@Intercept@NA	0.00	0.00	0.02
Condition Order@Gender@NA	0.00	0.00	0.27
Condition Order@Cultural Group@NA	0.00	0.00	0.06
Condition Order@AO@NA	0.40	0.00	0.51
Condition Order@Overimitation@NA	0.00	0.00	0.03

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[(2)] the presence of overimitation, dummy coded with No being the reference level

[(3)] dummy coded with girls being the reference level

[(4)] dummy coded with the Ugandan sample being the reference level

4.3 Appendix for Part 3

4.3.1 Statistical Analyses

4.3.1.1 Additional Details Regarding the Process of Reaching the Final Model: Model 9 (RQ5)

To analyse how cognition at 11 months varied with gender and cultural group, the Bayley-III Cognitive Score was included as a response in the model, with gender (girls or boys), cultural group (UK or Uganda) and the interaction included into the model as categorical predictors. The model stability parameters for Model 9 can be viewed in Table A.4.13.

The non-significant full-null comparison indicates that neither 'gender' or 'cultural group' (nor their interactions) are having a significant influence on cognitive score at 11 months. Therefore, no further exploration of the coefficients is warranted.

Table A. 4.13
Model stability parameters for Model 9, exploring whether gender or cultural group predict cognition at 11 months, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	-0.37	-0.41	-0.34
Gender[(2)]	0.09	0.06	0.12
Cultural Group[(3)]	0.05	0.01	0.08
Gender*Cultural Group	-0.09	-0.13	-0.05

(1) dummy coded with girls being the reference level

(2) dummy coded with the Ugandan sample being the reference level

4.3.1.2 Additional Details Regarding the Process of Reaching the Final Model: Model 10 (RQ6)

To analyse how the proportion of tokens given to the unfamiliar child varied with Approval Observe condition (Approval Observe or Other Conditions), gender (girls or boys) or cognition (continuous Bayley-III Cognitive Score), a logistic Generalized Linear Mixed Model (Baayen, 2008) was used, with these three variables and their 3-way interaction included as fixed effects into the model. Cognitive score was z-scored (scaling to a mean of zero and standard deviation of 1) to ease interpretability of the coefficients.

As the estimates of the correlations between the random slopes ranged from -1 to 1, indicating instability, we removed the correlation parameters of the random slopes of 'experimenter on trial' and 'order of conditions' from the model. This only led to a minor reduction in model fit (log-likelihoods, model with correlation parameters included: -482.54, df = 35; model without correlation parameters: -484.97, df = 19). The model stability parameters for Model 10 can be viewed in Table A.4.14.

The 3-way interaction between the 'Approval Observe condition', 'gender' and 'cognitive score' was non-significant (see Table A.4.15), so a reduced model was run without this interaction (discussed in the Results section).

Table A. 4.14

Model stability parameters for Model 10, exploring whether cognition at 11 months predicts 4-year-old children's sensitivity to normative expectations in girls and boys, retrieved from dropping each level of the grouping factor one at a time.

Term	Original	Min	Max
Intercept	0.88	0.16	1.43
AO[(1)]	0.00	-0.17	0.23
Cognitive Score[(2)]	0.28	0.12	0.52
Gender[(3)]	-1.16	-1.72	-0.50
Cultural Group[(4)]	-1.15	-1.56	-0.58
AO*Cognitive Score	-0.67	-0.83	-0.26
AO*Gender	0.69	0.42	0.86
Cognitive Score*Gender	-0.43	-1.30	0.82
AO*Cognitive Score*Gender	0.49	0.15	0.66
Participant ID@Intercept@NA	2.30	2.10	2.88
Experimenter on Trial@Intercept@NA	0.19	0.00	0.34
Experimenter on Trial@AO@NA	0.14	0.00	0.54
Experimenter on Trial@Gender@NA	0.00	0.00	0.59
Experimenter on Trial@Cognitive Score@NA	0.46	0.20	0.67
Condition Order@Intercept@NA	0.00	0.00	0.01
Condition Order@AO@NA	0.17	0.00	0.36
Condition Order@Gender@NA	0.00	0.00	0.68
Condition Order@Cultural Group@NA	0.00	0.01	0.03
Condition Order@Cognitive Score@NA	0.37	0.00	0.60

[(1)] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[2] z-scored Bayley-III Cognitive Score

[3] dummy coded with girls being the reference level

[4] dummy coded with the Ugandan sample being the reference level

Table A. 4.15

Results of full Model 10 examining whether cognition at 11 months predict sensitivity to normative expectations in 4-year-old girls and boys (estimates and standard errors, together with confidence limits and significance tests).

Term	Estimate	SE	Lower CI	Upper CI	z-value	*p*
Intercept	0.88	0.62	-0.33	2.09	1.43	.153
AO[1]	0.00	0.20	-0.46	0.48	0.01	.994
Cognitive Score[2]	0.28	0.59	-1.04	1.51	0.47	.641
Gender[3]	-1.16	0.62	-2.36	0.06	-1.88	.060
Cultural Group[4]	-1.15	0.64	-2.36	0.15	-1.78	.075
AO*Cognitive Score	-0.67	0.22	-1.25	-0.12	-3.06	.002
AO*Gender	0.69	0.26	0.11	1.21	2.63	.009
Cognitive Score*Gender	-0.43	0.71	-1.85	0.92	-0.61	.542
AO*Cognitive Score*Gender	0.49	0.32	-0.28	1.13	1.53	.125

[1] Approval Observe condition; dummy coded with the other conditions (Approval Unobserve, Indifference Observe & Indifference Unobserve) being the reference level

[2] z-scored Bayley-III Cognitive Score

[3] dummy coded with girls being the reference level

[4] dummy coded with the Ugandan sample being the reference level

References

Allen, J. W., & Andrews, K. (2024). How not to find over-imitation in animals. *Human Development*, *68*(2), 77-89. https://doi.org/10.1159/000537938

Allen, J. W., Sümer, C., & Ilgaz, H. (2021). Cultural affordances: Does model reliability affect over-imitation in preschoolers. *Cognitive Development*, *57*, 100999. https://doi.org/10.1016/j.cogdev.2020.100999

Aloise-Young, P. A. (1993). The development of self-presentation: Self-promotion in 6-to 10-year-old children. *Social cognition, 11*(2), 201-222. https://doi.org/10.1521/soco.1993.11.2.201

Alpízar, F., & Martinsson, P. (2013). Does it matter if you are observed by others? Evidence from donations in the field. *The Scandinavian Journal of Economics, 115*(1), 74-83. https://doi.org/10.1111/j.1467-9442.2012.01744.x

Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics, 92*(5-6), 1047-1060. https://doi.org/10.1016/j.jpubeco.2007.11.004

Amir, D., Ahl, R., Jordan, M., Bolotin, H., Bogese, M., González, G. T., ... & McAuliffe, K. (2024). The emergence of cooperative behaviors, norms, and strategies across five diverse societies.

Andreoni, J., & Petrie, R. (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of public Economics, 88*(7-8), 1605-1623. https://doi.org/10.1016/S0047-2727(03)00040-9

Andrews, K. (2009). Understanding Norms Without a Theory of Mind. *Inquiry*, *52*(5), 433–448. https://doi.org/10.1080/00201740903302584

Astington, J. W. (2001). The future of theory-of-mind research: Understanding motivational states, the role of language, and real-world consequences. *Child development*, *72*(3), 685-687. https://doi.org/10.1111/1467-8624.00305

Astor, K., Lindskog, M., Juvrud, J., Wangchuk, Namgyel, S. C., Wangmo, T., Tshering, K., & Gredebäck, G. (2022). Maternal postpartum depression impacts infants' joint attention differentially across cultures. Developmental Psychology, 58(12), 2230-2238. https://doi.org/10.1037/dev0001413

Avis, J., & Harris, P. L. (1991). Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. *Child development*, *62*(3), 460-467. https://doi.org/10.1111/j.1467-8624.1991.tb01544.x

Baayen, R. H. (2008). *Analyzing Linguistic Data.* Cambridge University. Press, Cambridge

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 131-158). Lawrence Erlbaum Associates, Inc.

Banerjee, R. (2002). Audience effects on self-presentation in childhood. *Social Development, 11*(4), 487-507. https://doi.org/10.1111/1467-9507.00212

Barbarroja, N., Ordin, M., Walker, M. J., Posada-Salazar, S., & Manrique, H. M. (2024). Imitation fidelity increases with age in boys, but not in girls: An intriguing finding in a cohort of children aged 3 to 6 years. Journal of Experimental Child Psychology, 239, 105826. https://doi.org/10.1016/j.jecp.2023.105826

Bard, K. A., Keller, H., Ross, K. M., Hewlett, B., Butler, L., Boysen, S. T., & Matsuzawa, T. (2021). Joint attention in human and chimpanzee infants in varied socio-ecological contexts. Monographs of the Society for Research in Child Development, 86(4), 7-217. https://doi.org/10.1111/mono.12435

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Baron-Cohen, S., O'riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of autism and developmental disorders*, *29*, 407-418. https://doi.org/10.1023/A:1023035012436

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., ... & Laurence, S. (2013). Early false-belief understanding in traditional non-Western societies. Proceedings of the Royal Society B: Biological Sciences, 280(1755), 20122654. https://doi.org/10.1098/rspb.2012.2654

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Battich, L., Fairhurst, M., & Deroy, O. (2020). Coordinating attention requires coordinated senses. Psychonomic bulletin & review, 27, 1126-1138. https://doi.org/10.3758/s13423-020-01766-z

Bayanova, L., Popova, R., Veraksa, A., & Bukhalenkova, D. (2022). Executive functions of preschoolers with different levels of cultural congruence. *International Journal of Early Years Education*, *30*(2), 249-260. https://doi.org/10.1080/09669760.2020.1779040

Bayley, N. (2006). Bayley Scales of Infant Development (Third Edition). TX:Harcourt Assessment.

Bekkering, H., WohlschlaÈger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology Section A*, *53*(1), 153-164. https://doi.org/10.1080/713755872

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and human Behavior*, *28*(3), 168-175. https://doi.org/10.1016/j.evolhumbehav.2006.10.003

Bennett, M., & Yeeles, C. (1990). Children's understanding of showing off. *The Journal of Social Psychology, 130*(5), 591-596.https://doi.org/10.1080/00224545.1990.9922950

Bereczkei, T., Birkas, B., & Kerekes, Z. (2007). Public charity offer as a proximate factor of evolved reputation-building strategy: an experimental analysis of a real-life situation. *Evolution and Human Behavior, 28*(4), 277-284. https://doi.org/10.1016/j.evolhumbehav.2007.04.002

Berl, R. E., & Hewlett, B. S. (2015). Cultural variation in the use of overimitation by the Aka and Ngandu of the Congo Basin. PloS one, 10(3), e0120180. https://doi.org/10.1371/journal.pone.0120180

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, *81*(6), 1641-1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Białek, A., Białecka-Pikul, M., & Stępień-Nycz, M. (2013). Responding to joint attention predicts joint action. In Proceedings of the 17th Workshop on the Semantics and Pragmatics of Dialogue (DialDam).

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge University Press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms.* Oxford University Press.

Bicchieri, C., & Muldoon, R. (2018). *Social Norms.* The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/social-norms/#Bib

Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191-208. https://doi.org/10.1002/bdm.621

Black, M. M., Walker, S. P., Fernald, L. C., 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.

Blair, C. (2016). Executive function and early childhood education. Current opinion in behavioral sciences, 10, 102-107. https://doi.org/10.1016/j.cobeha.2016.05.009

Blair, K. A., Denham, S. A., Kochanoff, A., & Whipple, B. (2004). Playing it cool: Temperament, emotion regulation, and social behavior in preschoolers. *Journal of School Psychology, 42*(6), 419–443. https://doi.org/10.1016/j.jsp.2004.10.002

Blake, P. R., & Rand, D. G. (2010). Currency value moderates equity preference among young children. *Evolution and human behavior*, *31*(3), 210-218. https://doi.org/10.1016/j.evolhumbehav.2009.06.012

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K.L., Ross, E., Vongsachang, H., Wrangham, R., & Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature, 528*(7581), 258-261. https://doi.org/10.1038/nature15703

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25-B31. https://doi.org/10.1016/S0010-0277(00)00096-2

Bohnet, I., & Frey, B. S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of economic behavior & organization, 38*(1), 43-57. https://doi.org/10.1016/S0167-2681(98)00121-8

Bosacki, S. (2021). Theory of mind and peer relationships in middle childhood and adolescence. In *Theory of mind in middle childhood and adolescence* (pp. 142-168). Routledge.

Bosacki, S., & Wilde Astington, J. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social development*, *8*(2), 237-255. https://doi.org/10.1111/1467-9507.00093

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. Philosophical transactions of the Royal Society of London. Series B, *Biological sciences, 364*(1533), 3281–3288. https://doi.org/10.1098/rstb.2009.0134

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. Proceedings of the National Academy of Sciences, 108(supplement_2), 10918-10925. https://doi.org/10.1073/pnas.1100290108

Bradford, E. E., Jentzsch, I., Gomez, J. C., Chen, Y., Zhang, D., & Su, Y. (2018). Cross-cultural differences in adult theory of mind abilities: a comparison of native-English speakers and native-Chinese speakers on the self/other differentiation task. *Quarterly Journal of Experimental Psychology*, *71*(12), 2665-2676. https://doi.org/10.1177/1747021818757170

Brandone, A. C., & Stout, W. (2023). The origins of theory of mind in infant social cognition: Investigating longitudinal pathways from intention understanding and joint attention to preschool theory of mind. *Journal of Cognition and Development*, *24*(3), 375-396. https://doi.org/10.1080/15248372.2022.2146117

Broesch, T., Crittenden, A. N., Beheim, B. A., Blackwell, A. D., Bunce, J. A., Colleran, H., Hagel, K., Kline, M., McElreath, R., Nelson, R. G., Pisor, A. C., Prall, S., Pretelli, I., Purzycki, B., Quinn, E. A., Ross, C., Scelza, B., Starkweather, K., Stieglitz, J., & 346 Mulder, M. B. (2020). Navigating cross-cultural research: Methodological and ethical considerations. Proceedings of the Royal Society B: Biological Sciences, 287(1935), 20201245. https://doi.org/10.1098/rspb.2020.1245

Brooks, M., E., Kristensen, K., Benthem, K., J. ,van, Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H., J., Mächler, M., & Bolker, B., M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378. https://doi.org/10.32614/RJ-2017-066.

Brownell, C. A. (2013). Early development of prosocial behavior: Current perspectives. *Infancy*, *18*(1), 1-9. https://doi.org/10.1111/infa.12004

Bunge, S. A., & Crone, E. A. (2009). Neural correlates of the development of cognitive control. *Neuroimaging in developmental clinical neuroscience*, 22-37.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. Child development, 84(2), 422-428. https://doi.org/10.1111/j.1467-8624.2012.01860.x

Caemmerer, J. M., deLeyer-Tiarks, J. M., Dale, B. A., Winter, E. L., Charamut, N. R., Scudder, A. M., Peters, E. C., Bray, M. A., & Kaufman, A. S. (2024). Does the Bayley-4 measure the same constructs across girls and boys and infants, toddlers, and preschoolers? Psychological Assessment, 36(11), 643-653. https://doi.org/10.1037/pas0001337

Calero, C. I., Salles, A., Semelman, M., & Sigman, M. (2013). Age and gender dependent development of Theory of Mind in 6-to 8-years old children. *Frontiers in human neuroscience*, *7*, 281. https://doi.org/10.3389/fnhum.2013.00281

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., ... & Collins, W. A. (2011). Early social cognition in three cultural contexts. Monographs of the society for research in child development, i-142.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., ... & Singh, S. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science*, *16*(5), 378-384. https://doi.org/10.1111/j.0956-7976.2005.01544.x

Carey, S. (2009). *The origin of concepts.* Oxford University Press.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. Child development, 72(4), 1032-1053. https://doi.org/10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. Infant and Child Development: An International Journal of Research and Practice, 11(2), 73-92. https://doi.org/10.1002/icd.298

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive Function and Theory of Mind: Stability and Prediction From Ages 2 to 3. *Developmental Psychology, 40*(6), 1105–1122. https://doi.org/10.1037/0012-1649.40.6.1105

Carpenter, M., & Call, J. (2013). How joint is the joint attention of apes and human infants. *Agency and joint attention*, *4961*.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 159–181). Boston Review.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, i-174. https://doi.org/10.2307/1166214

Castelli I, Massaro D, Sanfey AG, Marchetti A (2010) Fairness and intentionality in children's decision-making. Int Rev Econ 57: 269-288. https://doi.org/10.1007/s12232-010-0101-x

Castelli, I., Massaro, D., Bicchieri, C., Chavez, A., & Marchetti, A. (2014). Fairness norms and theory of mind in an ultimatum game: judgments, offers, and decisions in school-aged children. PloS one, 9(8), e105024. https://doi.org/10.1371/journal.pone.0105024

Castelli, I., Massaro, D., Sanfey, A. G., & Marchetti, A. (2013). "What is fair for you?" Judgments and decisions about fairness and Theory of Mind. European Journal of Developmental Psychology, 11(1), 49-62. https://doi.org/10.1080/17405629.2013.806264

Chajes, J. R., Grossmann, T., & Vaish, A. (2022). Fairness takes time: development of cooperative decision making in fairness context. *Journal of experimental child psychology*, *216*, 105344. https://doi.org/10.1016/j.jecp.2021.105344

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive development*, *15*(4), 481-498. https://doi.org/10.1016/S0885-2014(01)00037-5

Chasiotis A., Kiessling F., Hofer J., Campos D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. International Journal of Behavioral Development, 30, 249-260. https://doi.org/10.1177/0165025406066759

Childers, J. B., Vaughan, J., & Burquest, D. A. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. Journal of Child Language, 34(2), 199-225. https://doi.org/10.1017/S0305000906007835

Christens, S., Dannenberg, A., & Sachs, F. (2019). Identification of individuals and groups in a public goods experiment. *Journal of Behavioral and Experimental Economics, 82*, 101445. https://doi.org/10.1016/j.socec.2019.06.010

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in cognitive sciences, 15*(5), 218-226. https://doi.org/10.1016/j.tics.2011.03.003

Chudek, M., Heller, S., Birch, S., & Henrich, J. (2012). Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning. Evolution and human behavior, 33(1), 46-56. https://doi.org/10.1016/j.evolhumbehav.2011.05.005

Chiu, P. H., Kayali, M. A., Kishida, K. T., Tomlin, D., Klinger, L. G., Klinger, M. R., & Montague, P. R. (2008). Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron*, *57*(3), 463-473. https://doi.org/10.1016/j.neuron.2007.12.020

Ciairano, S., Bonino, S., & Miceli, R. (2006). Cognitive flexibility and social competence from childhood to early adolescence. Cognitie, Creier, Comportament/Cognition, Brain, Behavior, 10(3).

Ciairano, S., Visu-Petra, L., & Settanni, M. (2007). Executive inhibitory control and cooperative behavior during early school years: A follow-up study. *Journal of abnormal child psychology*, *35*, 335-345. https://doi.org/10.1007/s10802-006-9094-z

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*(1), 591-621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology, 58*(6), 1015. https://doi.org/10.1037/0022-3514.58.6.1015

Clay, Z., Over, H., & Tennie, C. (2018). What drives young children to over-imitate? Investigating the effects of age, context, action type, and transitivity. Journal of experimental child psychology, 166, 520-534. https://doi.org/10.1016/j.jecp.2017.09.008

Clay, Z., & Tennie, C. (2018). Is overimitation a uniquely human phenomenon? Insights from human children as compared to bonobos. *Child development*, *89*(5), 1535-1544.

Clegg, J. M., & Legare, C. H. (2016). A cross-cultural comparison of children's imitative flexibility. Developmental Psychology, 52(9), 1435. https://doi.org/10.1037/dev0000131

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Corriveau, K. H., DiYanni, C. J., Clegg, J. M., Min, G., Chin, J., & Nasrini, J. (2017). Cultural differences in the imitation and transmission of inefficient actions. Journal of Experimental Child Psychology, 161, 1-18. https://doi.org/10.1016/j.jecp.2017.03.002

Cowell, J. M., Lee, K., Malcolm-Smith, S., Selcuk, B., Zhou, X., & Decety, J. (2017). The development of generosity and moral cognition across five cultures. *Developmental science, 20*(4), e12403.

Craig, T. K. J., Bialas, I., Hodson, S., & Cox, A. D. (2004). Intergenerational transmission of somatization behaviour: 2. Observations of joint attention and bids for attention. Psychological Medicine, 34(2), 199-209. https://doi.org/10.1017/S0033291703001120

Dahl, A., & Paulus, M. (2019). From interest to obligation: The gradual development of human altruism. *Child Development Perspectives*, *13*(1), 10-14. https://doi.org/10.1111/cdep.12298

Dannenberg, A., Johansson-Stenman, O., & Wetzel, H. (2022). Status for the good guys: An experiment on charitable giving. *Economic Inquiry, 60*(2), 721-740. https://doi.org/10.1111/ecin.13039

Davidson, C., Shing, Y. L., McKay, C., Rafetseder, E., & Wijeakumar, S. (2023). The first year in formal schooling improves working memory and academic abilities. Developmental Cognitive Neuroscience, 60, 101205. https://doi.org/10.1016/j.dcn.2023.101205

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia, 44(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Deák, G. O., & Wiseheart, M. (2015). Cognitive flexibility in young children: General or task-specific capacity?. Journal of experimental child psychology, 138, 31-53. https://doi.org/10.1016/j.jecp.2015.04.003

Del Rosario, C., Slevin, M., Molloy, E. J., Quigley, J., & Nixon, E. (2021). How to use the Bayley scales of infant and toddler development. Archives of Disease in Childhood-Education and Practice, 106(2), 108-112. https://doi.org/10.1136/archdischild-2020-319063

Devine, R. T., & Apperly, I. A. (2022). Willing and able? Theory of mind, social motivation, and social competence in middle childhood and early adolescence. *Developmental Science*, *25*(1), e13137. https://doi.org/10.1111/desc.13137

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child development*, *85*(5), 1777-1794.

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52*(5), 758–771. https://doi.org/10.1037/dev0000105

Diaz, V., & Farrar, M. J. (2018). The missing explanation of the false-belief advantage in bilingual children: A longitudinal study. *Developmental science*, *21*(4), e12594. https://doi.org/10.1111/desc.12594

Ding, N., Miller, R., & Clayton, N. S. (2023). Inhibition and cognitive flexibility are related to prediction of one's own future preferences in young British and Chinese children. Cognition, 236, 105433. https://doi.org/10.1016/j.cognition.2023.105433

Dobson, A. J. (2002). *An introduction to Generalized Linear Models.* Chapman Hall/ CRC, Boca Raton.

Doebel, S. (2020). Rethinking Executive Function and Its Development. Perspectives on Psychological Science, 15(4), 942-956. https://doi.org/10.1177/1745691620904771

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, *10*(9), 1412–1430. https://doi.org/10.1111/2041-210X.13234

Dreman, S. B., & Greenbaum, C. W. (1973). Altruism or reciprocity: Sharing behavior in Israeli kindergarten children. *Child Development*, 61-68. https://doi.org/10.2307/1127680

Duffey, M. M., Ayuku, D., Ayodo, G., Abuonji, E., Nyalumbe, M., Giella, A. K., ... & McHenry, M. S. (2022). Translation and cultural adaptation of NIH toolbox cognitive tests into Swahili and Dholuo languages for use in children in Western Kenya. Journal of the International Neuropsychological Society, 28(4), 414-423. https://doi.org/10.1017/S1355617721000497

Duranti, A. (2008). Further reflections on reading other minds. Anthropological Quarterly, 81(2), 483-494. https://doi.org/10.1353/anq.0.0002

Eagly, A. H. (2009). The his and hers of prosocial behavior: an examination of the social psychology of gender. *American psychologist*, *64*(8), 644.

Eisenberg, N., Spinrad, T., & Sadovsky, A. (2006). Empathy-related responding in children. In *Handbook of moral development* (pp. 535-568). Psychology Press.

Emler, N. (1990). A social psychology of reputation. *European review of social psychology*, *1*(1), 171-193.

Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology*, *20*, 92-95. https://doi.org/10.1016/j.copsyc.2017.08.024

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One, 7*(10), e48433. https://doi.org/10.1371/journal.pone.0048433

Engelmann, J. M., Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, *16*(6), 952-958. https://doi.org/10.1111/desc.12086

Epley, N., & Dunning, D. (2000). Feeling" holier than thou": are self-serving assessments produced by errors in self-or social prediction?. *Journal of personality and social psychology*, *79*(6), 861. https://psycnet.apa.org/doi/10.1037/0022-3514.79.6.861

Essler, S., Christner, N., Becher, T., & Paulus, M. (2023). The ontogenetic emergence of normativity: How action imitation relates to infants' norm enforcement. *Journal of Experimental Child Psychology*, *227*, 105591. https://doi.org/10.1016/j.jecp.2022.105591

Fabes, R. A., & Eisenberg, N. (1998). Meta-analyses of age and sex differences in children's and adolescents' prosocial behavior. *Handbook of child psychology*, *3*, 1-29. META1.PDF (asu.edu)

Fabes, R. A., Carlo, G., Kupanoff, K., & Laible, D. (1999). Early adolescence and prosocial/moral behavior I: The role of individual processes. *The Journal of Early Adolescence*, *19*(1), 5-16. https://doi.org/10.1177/0272431699019001001

Farrow, K., Grolleau, G., & Ibanez, L. (2017). Social norms and pro-environmental behavior: A review of the evidence. *Ecological Economics, 140*, 1-13. https://doi.org/10.1016/j.ecolecon.2017.04.017

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in cognitive sciences, 8*(4), 185-190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137-140. https://doi.org/10.1038/415137a

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079-1083. https://doi.org/10.1038/nature07155

Fernald, L. C., Weber, A., Galasso, E., & Ratsifandrihamanana, L. (2011). Socioeconomic gradients and child development in a very low income population: Evidence from Madagascar. *Developmental science*, *14*(4), 832-847. https://doi.org/10.1111/j.1467-7687.2010.01032.x

Field, A. (2005). Discovering Statistics using SPSS. Sage Publications, London.

Flavell, J. H., Flavell, E. R., Green, F. L., & Wilcox, S. A. (1981). The development of three spatial perspective-taking rules. *Child Development*, 356-358. https://doi.org/10.2307/1129250

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Freepik. (n.d.). https://www.freepik.com/

Frick, A., Clément, F., & Gruber, T. (2017). Evidence for a sex effect during overimitation: boys copy irrelevant modelled actions more than girls across cultures. *Royal Society Open Science*, *4*(12), 170367. https://doi.org/10.1098/rsos.170367

Fu, G., Heyman, G. D., Qian, M., Guo, T., & Lee, K. (2015). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science, 19*(2), 275-283. https://doi.org/10.1111/desc.12304

Gaffan, E. A., Martins, C., Healy, S., & Murray, L. (2010). Early social experience and individual differences in infants' joint attention. Social Development, 19(2), 369-393. https://doi.org/10.1111/j.1467-9507.2008.00533.x

Garbarini, F., Boero, R., D'Agata, F., Bravo, G., Mosso, C., Cauda, F., ... & Sacco, K. (2014). Neural correlates of gender differences in reputation building. *PLoS One*, *9*(9), e106285. https://doi.org/10.1371/journal.pone.0106285

Gardner, W. L., & Gabriel, S. (2004). Gender Differences in Relational and Collective Interdependence: Implications for Self-Views, Social Behavior, and Subjective Well-Being. In A. H. Eagly, A. E. Beall, & R. J. Sternberg (Eds.), *The psychology of gender* (2nd ed., pp. 169–191). The Guilford Press.

Gee, C. L., & Heyman, G. D. (2007). Children's evaluation of other people's self-descriptions. *Social Development, 16*(4), 800-818. https://doi.org/10.1111/j.1467-9507.2007.00408.x

Gelfand, M. J., Nishii, L. H., & Raver, J. L. (2006). On the nature and importance of cultural tightness-looseness. *Journal of applied psychology*, *91*(6), 1225. https://psycnet.apa.org/doi/10.1037/0021-9010.91.6.1225

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., ... & Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*(6033), 1100-1104. https://doi.org/10.1126/science.1197754

Gelman, A. and Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge, New Jersey.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*, *101*(1), 129-152. https://doi.org/10.1016/j.cognition.2005.05.007

Giner Torréns, M., & Kärtner, J. (2017). The influence of socialization on early helping from a cross-cultural perspective. *Journal of Cross-Cultural Psychology*, *48*(3), 353-368. https://doi.org/10.1177/0022022117690451

Gintis, H. (2004). The genetic side of gene-culture coevolution: internalization of norms and prosocial emotions. Journal of economic behavior & organization, 53(1), 57-67.

Godamunne, P., Liyanage, C., Wimaladharmasooriya, N., Pathmeswaran, A., Wickremasinghe, A. R., Patterson, C., & Sathiakumar, N. (2014). Comparison of performance of Sri Lankan and US children on cognitive and motor scales of the Bayley scales of infant development. BMC research notes, 7, 1-5. https://doi.org/10.1186/1756-0500-7-300

Grace, D. M., David, B. J., & Ryan, M. K. (2008). Investigating preschoolers' categorical thinking about gender through imitation, attention, and the use of self-categories. *Child development*, *79*(6), 1928-1941. https://doi.org/10.1111/j.1467-8624.2008.01234.x

Graham, K. E., Buryn-Weitzel, J. C., Lahiff, N. J., Wilke, C., & Slocombe, K. E. (2021). Detecting joint attention events in mother-infant dyads: Sharing looks cannot be reliably identified by naïve third-party observers. *PLoS One*, *16*(7), e0255241. https://doi.org/10.1371/journal.pone.0255241

Gummerum, M., Hanoch, Y., Keller, M., Parsons, K., & Hummel, A. (2010). Preschoolers' allocations in the dictator game: The role of moral emotions. *Journal of Economic Psychology*, *31*(1), 25-34. https://doi.org/10.1016/j.joep.2009.09.002

Gummerum, M., Keller, M., Takezawa, M., & Mata, J. (2008). To give or not to give: Children's and adolescents' sharing and moral negotiations in economic decision situations. *Child Development*, *79*(3), 562-576. https://doi.org/10.1111/j.1467-8624.2008.01143.x

Gurven, M. (2004). To give and to give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences, 27*(4), 543-560.

Haft, S. L., & Hoeft, F. (2017). Poverty's impact on children's executive functions: Global considerations. *New directions for child and adolescent development*, *2017*(158), 69-79. https://doi.org/10.1002/cad.20220

Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, *476*(7360), 328-331. https://doi.org/10.1038/nature10278

Harbaugh, W. T., & Krause, K. (2000). Children's contributions in public good experiments: The development of altruistic and free-riding behaviors. *Economic inquiry*, *38*(1), 95-109. https://doi.org/10.1111/j.1465-7295.2000.tb00006.x

Hardecker, S., & Tomasello, M. (2017). From imitation to implementation: How two-and three-year-old children learn to enforce social norms. *British Journal of Developmental Psychology*, *35*(2), 237-248. https://doi.org/10.1111/bjdp.12159

Hareli, S., Kafetsios, K., & Hess, U. (2015). A cross-cultural study on emotion expression and the learning of social norms. Frontiers in psychology, 6, 1501. https://doi.org/10.3389/fpsyg.2015.01501

Hareli, S., Moran-Amir, O., David, S., & Hess, U. (2013). Emotions as signals of normative conduct. Cognition & emotion, 27(8), 1395-1404. https://doi.org/10.1080/02699931.2013.791615

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... & Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behavioral and brain sciences, 28(6), 795-815. https://doi.org/10.1017/S0140525X05000142

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, *33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767-1770. https://doi.org/10.1126/science.1127333

Hepach, R., Vaish, A., & Tomasello, M. (2017). Children's intrinsic motivation to provide help themselves after accidentally harming others. *Child development*, *88*(4), 1251-1264. https://doi.org/10.1111/cdev.12646

Hepach, R., & Warneken, F. (2018). Editorial overview: Early development of prosocial behavior: Revealing the foundation of human prosociality. Current Opinion in Psychology, 20, 4-8. https://doi.org/10.1016/j.copsyc.2018.02.001

Hepach, R., Haberl, K., Lambert, S., & Tomasello, M. (2017). Toddlers help anonymously. *Infancy*, *22*(1), 130-145. https://doi.org/10.1111/infa.12143

Hepach, R., Engelmann, J. M., Herrmann, E., Gerdemann, S. C., & Tomasello, M. (2023). Evidence for a developmental shift in the motivation underlying helping in early childhood. *Developmental Science*, *26*(1), e13253. https://doi.org/10.1111/desc.13253

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362-1367. https://doi.org/10.1126/science.1153808

Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2013). Direct and indirect reputation formation in nonhuman great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus*) and human children (*Homo sapiens*). *Journal of Comparative Psychology, 127*(1), 63–75. https://doi.org/10.1037/a0028929

Heyes, C. (2024). Rethinking norm psychology. Perspectives on Psychological Science, 19(1), 12-38. https://doi.org/10.1177/17456916221112075

Heyman, G. D., Fu, G., & Lee, K. (2007). Evaluating claims people make about themselves: The development of skepticism. *Child development, 78*(2), 367-375. https://doi.org/10.1111/j.1467-8624.2007.01003.x

Heyman, G., Barner, D., Heumann, J., & Schenck, L. (2014). Children's sensitivity to ulterior motives when evaluating prosocial behavior. *Cognitive science, 38*(4), 683-700. https://doi.org/10.1111/cogs.12089

Hill, V., & Pillow, B. H. (2006). Children's understanding of reputations. *The Journal of genetic psychology, 167*(2), 137-157. https://doi.org/10.3200/GNTP.167.2.137-157

Hine, B. A. (2017). Identifying the male prosocial niche: The gender-typing of prosocial behaviour across childhood and adolescence. *European Journal of Developmental Psychology*, *14*(2), 206-220. https://doi.org/10.1080/17405629.2016.1186009

Hine, B., & Leman, P. J. (2013). The developing relationship between gender and pro-social behaviour. In *Gender and development* (pp. 78-108). Psychology Press.

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). 'Over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*, 90-108. https://doi.org/10.1016/j.dr.2018.12.002

Holden, E., Buryn-Weitzel, J. C., Atim, S., Biroch, H., Donnellan, E., Graham, K. E., ... & Slocombe, K. E. (2022). Maternal attitudes and behaviours differentially shape infant early life experience: A cross cultural study. *Plos one*, *17*(12), e0278378. https://doi.org/10.1371/journal.pone.0278378

Hong, Y. Y., Morris, M. W., Chiu, C. Y., and Benet-Martinez, V. (2000). Multicultural minds: a dynamic constructivist approach to culture and cognition. Am. Psychol. 55, 709-720. https://doi.org/10.1037/0003-066X.55.7.709

Horner, V., Whiten, A. Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). Anim Cogn 8, 164-181 (2005). https://doi.org/10.1007/s10071-004-0239-6

House, B. R. (2017). Diverse ontogenies of reciprocal and prosocial behavior: cooperative development in Fiji and the United States. *Developmental science*, *20*(6), e12466. https://doi.org/10.1111/desc.12466

House, B. R. (2018). How do social norms influence prosocial development*?. Current Opinion in Psychology, 20*, 87-91. https://doi.org/10.1016/j.copsyc.2017.08.011

House, B. R., & Tomasello, M. (2018). Modeling social norms increasingly influences costly sharing in middle childhood. *Journal of experimental child psychology, 171*, 84-98. https://doi.org/10.1016/j.jecp.2017.12.014

House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebioglu, S., Crittenden, A. N., Erut, A., Lew-Levy, S., Sebastian-Enesco, C., Smith, A. M., Yilmaz, S., & Silk, J. B. (2020a). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour, 4*(1), 36-44. https://doi.org/10.1038/s41562-019-0734-z

House, B. R., Kanngiesser, P., Barrett, H. C., Yilmaz, S., Smith, A. M., Sebastian-Enesco, C., Erut, A., & Silk, J. B. (2020b). Social norms and cultural diversity in the development of third-

party punishment. *Proceedings of the Royal Society B, 287*(1925), 20192794. https://doi.org/10.1098/rspb.2019.2794

House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., Boyette, A. H., Hewlett, B.S., McElreath, R., & Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences, 110*(36), 14586-14591. https://doi.org/10.1073/pnas.1221217110

House, B., Henrich, J., Sarnecka, B., & Silk, J. B. (2013). The development of contingent reciprocity in children. *Evolution and Human Behavior*, *34*(2), 86-93. https://doi.org/10.1016/j.evolhumbehav.2012.10.001

House, B., Silk, J. B., & McAuliffe, K. (2023). No strong evidence for universal gender differences in the development of cooperative behaviour across societies. *Philosophical Transactions of the Royal Society B*, *378*(1868), 20210439. https://doi.org/10.1098/rstb.2021.0439

Howard, L. H., Henderson, A. M., Carrazza, C., & Woodward, A. L. (2015). Infants' and young children's imitation of linguistic in-group and out-group informants. *Child development*, *86*(1), 259-275. https://doi.org/10.1111/cdev.12299

Howard, L. H., Henderson, A. M., Carrazza, C., & Woodward, A. L. (2015). Infants' and young children's imitation of linguistic in-group and out-group informants. Child development, 86(1), 259-275. https://doi.org/10.1111/cdev.12299

Hruschka, D. J. (2020). "What we look with" is as important as "What we look at". Evolution and Human Behavior, 41(5), 458-459. https://doi.org/10.1016/j.evolhumbehav.2020.07.011 https://doi.org/10.1016/j.evolhumbehav.2020.07.011

Hua, J., Li, Y., Ye, K., Ma, Y., Lin, S., Gu, G., & Du, W. (2019). The reliability and validity of Bayley-III cognitive scale in China's male and female children. Early human development, 129, 71-78. https://doi.org/10.1016/j.earlhumdev.2019.01.017

Huber, L., Popovová, N., Riener, S., Salobir, K., & Cimarelli, G. (2018). Would dogs copy irrelevant actions from their human caregiver?. *Learning & Behavior*, *46*(4), 387-397. https://doi.org/10.3758/s13420-018-0336-z

Huber, L., Salobir, K., Mundry, R., & Cimarelli, G. (2020). Selective overimitation in dogs. *Learning & Behavior*, *48*, 113-123. https://doi.org/10.3758/s13420-019-00400-w

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental psychology*, *43*(6), 1447.

Hughes, C., Ensor, R., & Marks, A. (2011). Individual differences in false belief understanding are stable from 3 to 6 years of age and predict children's mental state talk with school friends. *Journal of Experimental Child Psychology*, *108*(1), 96-112. https://doi.org/10.1016/j.jecp.2010.07.012

Huppert, E., Cowell, J. M., Cheng, Y., Contreras-Ibáñez, C., Gomez-Sicard, N., Gonzalez-Gadea, M. L., Huepe, D. H., Ibanez, A., Lee, K., Mahasneh, R., Malcolm-Smith, S., Salas, N., Selcuk, B., Tungodden, B., Wong, A., Zhou, X., & Decety, J. (2019). The development of children's preferences for equality and equity across 13 individualistic and collectivist cultures. *Developmental science, 22*(2), e12729.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. Developmental Psychology, 52(8), 1192-1205. https://doi.org/10.1037/dev0000140

Izuma, K., Saito, D. N., & Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of cognitive neuroscience, 22*(4), 621-631. https://doi.org/10.1162/jocn.2009.21228

Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E., & Galinsky, A. D. (2018). The critical role of second-order normative beliefs in predicting energy conservation. *Nature Human Behaviour, 2*(10), 757-764. https://doi.org/10.1038/s41562-018-0434-0

Jaeger, C. M., & Schultz, P. W. (2017). Coupling social norms and commitments: Testing the underdetected nature of social influence. *Journal of Environmental Psychology, 51*, 199-208. https://doi.org/10.1016/j.jenvp.2017.03.015

Jaeggi, A. V., Burkart, J. M., & Van Schaik, C. P. (2010). On the psychology of cooperation in humans and other primates: combining the natural history and experimental evidence of prosociality. Philosophical Transactions of the Royal Society B: *Biological Sciences, 365*(1553), 2723-2735.

Javo, C., Rønning, J. A., & Heyerdahl, S. (2004). Child-rearing in an indigenous Sami population in Norway: A cross-cultural comparison of parental attitudes and expectations. Scandinavian Journal of Psychology, 45(1), 67-78. https://doi.org/10.1111/j.1467-9450.2004.00380.x

Kâğıtçıbaşı, C. (1996). The Autonomous-Relational Self: A New Synthesis. European Psychologist, 1(3), 180–186.

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and social psychology bulletin, 26*(8), 1002-1012. https://doi.org/10.1177/01461672002610009

Kanngiesser, P., Schäfer, M., Herrmann, E., Zeidler, H., Haun, D., & Tomasello, M. (2022). Children across societies enforce conventional norms but in culturally variable ways. *Proceedings of the National Academy of Sciences*, *119*(1), e2112521118. https://doi.org/10.1073/pnas.2112521118

Kärtner, J., Keller, H., & Chaudhary, N. (2010). Cognitive and social influences on early prosocial behavior in two sociocultural contexts. Developmental psychology, 46(4), 905. https://doi.org/10.1037/a0019718

Keller, H., Lamm, B., Abels, M., Yovsi, R., Borke, J., Jensen, H., ... & Chaudhary, N. (2006). Cultural models, socialization goals, and parenting ethnotheories: A multicultural analysis. *Journal of cross-cultural psychology*, *37*(2), 155-172.

Keller, M., Gummerum, M., Canz, T., Gigerenzer, G., & Takezawa, M. (2013). The is and ought of sharing: The equality heuristic across the lifespan. In *Simple heuristics in a social world* (pp. 171-195). Oxford University Press.

Kenward, B., Karlsson, M., & Persson, J. (2011). Over-imitation is better explained by norm learning than by distorted causal learning. Proceedings: Biological Sciences, 278(1709), 1239-1246. https://doi.org/10.1098/rspb.2010.1399

Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. Journal of Experimental Child Psychology, 116(2), 392-406. https://doi.org/10.1016/j.jecp.2013.07.002

Kim, Z., Óturai, G., Király, I., & Knopf, M. (2015). The role of objects and effects in action imitation: Comparing the imitation of object-related actions vs. gestures in 18-month-old infants. Infant Behavior and Development, 41, 43-51. https://doi.org/10.1016/j.infbeh.2015.07.002

Kim, Z., Óturai, G., Király, I., Hirte, M., & Knopf, M. (2018). A variation of the social context in the warm-up period influences 18-month-olds' imitation. *Cognitive Development*, *45*, 105-112. https://doi.org/10.1016/j.cogdev.2018.01.003

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental science*, *14*(1), 106-111. https://doi.org/10.1111/j.1467-7687.2010.00965.x

Király, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of experimental child psychology*, *116*(2), 471-486. https://doi.org/10.1016/j.jecp.2012.12.003

Klein-Radukic, S., & Zmyj, N. (2023). The predictive value of the cognitive scale of the Bayley Scales of Infant and Toddler Development-III. Cognitive Development, 65, 101291. https://doi.org/10.1016/j.cogdev.2022.101291

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child development*, *67*(2), 490-507. https://doi.org/10.1111/j.1467-8624.1996.tb01747.x

Köster, M., & Hepach, R. (2024). Preverbal infants' understanding of social norms. Scientific Reports, 14(1), 2983. https://doi.org/10.1038/s41598-024-53110-3

Köster, M., & Hepach, R. (2024). Preverbal infants' understanding of social norms. Scientific Reports, 14(1), 2983. https://doi.org/10.1038/s41598-024-53110-3

Kristen, S., Sodian, B., Thoermer, C., & Perst, H. (2011). Infants' joint attention skills predict toddlers' emerging mental state language. Developmental psychology, 47(5), 1207. https://doi.org/10.1037/a0024808

Krogh, M. T., & Væver, M. S. (2019). Does gender affect Bayley-III scores and test-taking behavior?. Infant Behavior and Development, 57, 101352. https://doi.org/10.1016/j.infbeh.2019.101352

Kuntoro, I. A., Saraswati, L., Peterson, C., & Slaughter, V. (2013). Micro-cultural influences on theory of mind development: A comparative study of middle-class and pemulung children in Jakarta, Indonesia. *International Journal of Behavioral Development*, *37*(3), 266-273. https://doi.org/10.1177/0165025413478258

Lacetera, N., & Macis, M. (2008). Motivating Altruism: A Field Study. *IZA Discussion Paper No. 3770*, Available at SSRN: https://ssrn.com/abstract=1290039

Lan, X. (2023). Perceived parenting styles, cognitive flexibility, and prosocial behavior in Chinese Youth with an immigrant background: A three-group comparison. Current Psychology, 42(24), 20718-20736. https://doi.org/10.1007/s12144-022-03140-9

Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. Child development, 87(3), 633-642. https://doi.org/10.1111/cdev.12542

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. Trends in cognitive sciences, 19(11), 688-699. https://doi.org/10.1016/j.tics.2015.08.005

Legare, C. H., Dale, M. T., Kim, S. Y., & Deák, G. O. (2018). Cultural variation in cognitive flexibility reveals diversity in the development of executive functions. Scientific reports, 8(1), 16326. https://doi.org/10.1038/s41598-018-34756-2

Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PloS one, 7*(10), e48292. https://doi.org/10.1371/journal.pone.0048292

Leman, P. J., Keller, M., Takezawa, M., & Gummerum, M. (2009). Children's and adolescents' decisions about sharing money with others. *Social Development*, *18*(3), 711-727. https://doi.org/10.1111/j.1467-9507.2008.00486.x

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive psychology*, *50*(1), 45-85. https://doi.org/10.1016/j.cogpsych.2004.06.002

Li, L., Britvan, B., & Tomasello, M. (2021). Young children conform more to norms than to preferences. *Plos one*, *16*(5), e0251228. https://doi.org/10.1371/journal.pone.0251228

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, *4*(3-4), 231-244. https://doi.org/10.1556/jcep.4.2006.3-4.3

Lillard, A. (1998). Ethnopsychologies: cultural variations in theories of mind. Psychological bulletin, 123(1), 3. https://doi.org/10.1037/0033-2909.123.1.3

List, J. A., & Levitt, S. D. (2005). What do laboratory experiments tell us about the real world. *NBER working paper*, 14-20. Accessed via: http://pricetheory.uchicago.edu/levitt/Papers/LevittList2005.pdf

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: a meta-analysis of false-belief understanding across cultures and languages. Developmental psychology, 44(2), 523. https://doi.org/10.1037/0012-1649.44.2.523

Lucas, M. M., Wagner, L., & Chow, C. (2008). Fair game: The intuitive economics of resource exchange in four-year olds. *Journal of Social, Evolutionary, and Cultural Psychology*, *2*(3), 74. https://doi.org/10.1037/h0099353

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. Child development, 76(3), 697-712. https://doi.org/10.1111/j.1467-8624.2005.00872.x

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, *104*(50), 19751-19756. https://doi.org/10.1073/pnas.0704452104

Macdonald, J. A., Beauchamp, M. H., Crigan, J. A., & Anderson, P. J. (2014). Age-related differences in inhibitory control in the early school years. *Child Neuropsychology*, *20*(5), 509-526. https://doi.org/10.1080/09297049.2013.822060

Mageo J. M. (1998). Theorizing self in Samoa: Emotions, genders, and sexualities. Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.15560

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive development*, *33*, 40-55. https://doi.org/10.1016/j.cogdev.2014.07.001

Marlowe, F. W. (2007). Hunting and gathering: the human sexual division of foraging labor. *Cross-cultural research, 41*(2), 170-195.

Martin, R., & Randal, J. (2008). How is donation behaviour affected by the donations of others?. *Journal of Economic Behavior & Organization, 67*(1), 228-238.

Marsh, L. E., Ropar, D., & Hamilton, A. F. D. C. (2019). Are you watching me? The role of audience and object novelty in overimitation. *Journal of Experimental Child Psychology*, *180*, 123-130. https://doi.org/10.1016/j.jecp.2018.12.010

Marshall, S., Lavender, E. & Slocombe, K.E. (In prep). Cross-Cultural Differences in Infant Play Behaviour Between Two Samples: the UK and Uganda. *Developmental Psychology.*

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mayer, A., & Riese, J. (2013). False Belief Understanding in Samoa: Evidence for Continuous Development and Cross-Cultural Variability. Theory of Mind in the Pacific. Reasoning Across Cultures, 79-108. https://philarchive.org/archive/WASTOM#page=90

Mayer, A., & Träuble, B. E. (2013). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. International Journal of Behavioral Development, 37(1), 21-28. https://doi.org/10.1177/0165025412454030

McAuliffe, K., Raihani, N. J., & Dunham, Y. (2017). Children are sensitive to norms of giving. *Cognition, 167,* 151-159. https://doi.org/10.1016/j.cognition.2017.01.006

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the national academy of sciences*, *98*(20), 11832-11835. https://doi.org/10.1073/pnas.211415698

McCoy, D. C., Zuilkowski, S. S., & Fink, G. (2015). Poverty, physical stature, and cognitive skills: Mechanisms underlying children's school enrollment in Zambia. *Developmental Psychology, 51*(5), 600–614. https://doi.org/10.1037/a0038924

McCullagh, P., & Nelder, J. A. (1989). Binary data. In *Generalized linear models* (pp. 98-148). Springer US.

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3- and 5-year-old children. Cognitive Development, 22(3), 353-364. doi: 10.1016/j.cogdev.2007.01.001

McGuigan, N., & Whiten, A. (2009). Emulation and "overemulation" in the social learning of causally opaque versus causally transparent tool use by 23-and 30-month-olds. *Journal of experimental child psychology*, *104*(4), 367-381. https://doi.org/10.1016/j.jecp.2009.07.001

McHenry, M. S., Oyungu, E., Yang, Z., Hines, A. C., Ombitsa, A. R., Vreeman, R. C., ... & Monahan, P. O. (2021). Cultural adaptation of the Bayley scales of infant and toddler development, for use in Kenyan children aged 18-36 months: a psychometric study. Research in developmental disabilities, 110, 103837. https://doi.org/10.1016/j.ridd.2020.103837

McHenry, M. S., Roose, A., Abuonji, E., Nyalumbe, M., Ayuku, D., Ayodo, G., ... & Kaat, A. J. (2023). A psychometric evaluation of the NIH Toolbox fluid cognition tests adapted for Swahili and Dholuo languages in Kenyan children and adolescents. *Journal of the International Neuropsychological Society*, *29*(10), 933-942. doi:10.1017/S1355617723000632

Melis, A. P., Hare, B., & Tomasello, M. (2006). Chimpanzees recruit the best collaborators. *Science*, *311*(5765), 1297-1300. https://doi.org/10.1126/science.1123007

Melnyk, V., Herpen, E. V., Fischer, A. R., & van Trijp, H. C. (2011). To think or not to think: The effect of cognitive deliberation on the influence of injunctive versus descriptive social norms. Psychology & marketing, 28(7), 709-729. https://doi.org/10.1002/mar.20408

Memisevic, H., & Biscevic, I. (2018). Exploring the link between inhibitory control and cognitive flexibility in preschool children. Cognition, Brain, Behavior, 22(1), 1-11. https://doi.org/10.24193/cbb.2018.22.01

Messer, E. J., Lumsden, A., Burgess, V., & McGuigan, N. (2021). Young children selectively adopt sharing norms according to norm content and donor age. *Cognitive Development*, *59*, 101088. https://doi.org/10.1016/j.cogdev.2021.101088

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature*, *415*(6870), 424-426. https://doi.org/10.1038/415424a

Misch, A., & Dunham, Y. (2021). (Peer) Group influence on children's prosocial and antisocial behavior. Journal of Experimental Child Psychology, 201, 104994. https://doi.org/10.1016/j.jecp.2020.104994

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Moll, H., Meltzoff, A. N., Merzsch, K., & Tomasello, M. (2013). Taking versus confronting visual perspectives in preschool children. *Developmental Psychology, 49*(4), 646–654. https://doi.org/10.1037/a0028633

Moore, C. (2009). Fairness in children's resource allocation depends on the recipient. *Psychological Science*, *20*(8), 944-948.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science*, *16*(5), 269-274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Najnin, S., & Banerjee, B. (2018). Pragmatically framed cross-situational noun learning using computational reinforcement models. Frontiers in Psychology, 9, 5. https://doi.org/10.3389/fpsyg.2018.00005

Nelson, P. B., Adamson, L. B., & Bakeman, R. (2008). Toddlers' joint engagement experience facilitates preschoolers' acquisition of theory of mind. Developmental science, 11(6), 847-852. https://doi.org/10.1111/j.1467-7687.2008.00733.x

Nielsen, M. (2006). Copying actions and copying outcomes: social learning through the second year. *Developmental psychology*, *42*(3), 555.

Nielsen, M., & Blank, C. (2011). Imitation in young children: when who gets copied is more important than what gets copied. *Developmental psychology*, *47*(4), 1050.

Nielsen, M., & Susianto, E. W. (2010). Failure to find over-imitation in captive orangutans (Pongo pygmaeus): Implications for our understanding of cross-generation information transfer.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. Journal of experimental child psychology, 162, 31-38. https://doi.org/10.1016/j.jecp.2017.04.017

Nielsen, M., Mushin, I., Tomaselli, K., & Whiten, A. (2014). Where culture takes hold:"Overimitation" and its flexible deployment in Western, Aboriginal, and Bushmen children. Child development, 85(6), 2169-2184. https://doi.org/10.1111/cdev.12265

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental science*, *11*(5), 722-731. https://doi.org/10.1111/j.1467-7687.2008.00722.x

Nielsen, M., & Tomaselli, K. (2010). Over-imitation in Kalahari Bushman children and the origins of human cultural cognition. Psychological Science (Sage Publications Inc.), 21(5), 729-736. doi:10.1177/0956797610368808

Nieuwenhuis, R., Te Grotenhuis, H. F., & Pelzer, B. J. (2012). *Influence. ME: tools for detecting influential data in mixed effects models.* https://repository.ubn.ru.nl/bitstream/handle/2066/103101/103101.pdf

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and social psychology bulletin, 34*(7), 913-923. https://doi.org/10.1177/0146167208316691

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*(6685), 573-577. https://doi.org/10.1038/31225

Obradović, J., & Willoughby, M. T. (2019). Studying executive function skills in young children in low-and middle-income countries: Progress and directions. *Child Development Perspectives*, *13*(4), 227-234. https://doi.org/10.1111/cdep.12349

Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. Child development, 79(1), 80-99. https://doi.org/10.1111/j.1467-8624.2007.01112.x

Olafsen, K. S., Rønning, J. A., Kaaresen, P. I., Ulvund, S. E., Handegård, B. H., & Dahl, L. B. (2006). Joint attention in term and preterm infants at 12 months corrected age: The significance of gender and intervention based on a randomized controlled trial. *Infant Behavior and Development*, *29*(4), 554-563. https://doi.org/10.1016/j.infbeh.2006.07.004

Over, H. (2020). The social function of imitation in development. *Annual Review of Developmental Psychology*, *2*(1), 93-109. https://doi.org/10.1146/annurev-devpsych-033020-024051

Over, H., & Carpenter, M. (2012). Putting the social into social learning: explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*(2), 182.https://psycnet.apa.org/doi/10.1037/a0024555

Over, H., & Carpenter, M. (2013). The social side of imitation. *Child development perspectives*, *7*(1), 6-11.https://doi.org/10.1111/cdep.12006

Packard, C. D., & Schultz, P. W. (2023). Emotions as the Enforcers of Norms. Emotion Review, 15(4), 279-283. https://doi.org/10.1177/17540739231194315

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: a field experiment on collective norms and harassment behavior in a school social network. Journal of personality and social psychology, 103(6), 899. https://doi.org/10.1037/a0030015

Paulus, M. (2015). Children's inequity aversion depends on culture: A cross-cultural comparison. *Journal of experimental child psychology*, *132*, 240-246. https://doi.org/10.1016/j.jecp.2014.12.007

Paulus, M. (2014). The emergence of prosocial behavior: Why do infants and toddlers help, comfort, and share?. *Child Development Perspectives*, *8*(2), 77-81. https://doi.org/10.1111/cdep.12066

Paulus, M., & Moore, C. (2014). The development of recipient-dependent sharing behavior and sharing expectations in preschool children. *Developmental psychology*, *50*(3), 914. https://psycnet.apa.org/doi/10.1037/a0034169

Pelto, P. J. (1968). The differences between" tight" and" loose" societies. *Trans-action*, *5*, 37-40.

Pendergast, L. L., Schaefer, B. A., Murray-Kolb, L. E., Svensen, E., Shrestha, R., Rasheed, M. A., Scharf, R. J., Kosek, M., Vasquez, A. O., Maphula, A., Costa, H., Rasmussen, Z. A., Yousafzai, A., Tofail, F., Seidman, J. C., & The MAL-ED Network Investigators. (2018). Assessing development across cultures: Invariance of the Bayley-III Scales Across Seven International MAL-ED sites. School Psychology Quarterly, 33(4), 604-614. https://doi.org/10.1037/spq0000264

Perkovich, E., & Yoshida, H. (2024, May). Infant Sex Effect on Naturally Occurring Attention Behaviors During Interactive Object Play. In *2024 IEEE International Conference on Development and Learning (ICDL)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICDL61372.2024.10644681

Perloff, R. M. (1982). Gender constancy and same-sex imitation: A developmental study. *The Journal of Psychology*, *111*(1), 81-86. https://doi.org/10.1080/00223980.1982.9923516

Pexels. (n.d.). https://www.pexels.com/

Povis, K. T., & Crowley, K. (2015). Family learning in object-based museums: The role of joint attention. Visitor Studies, 18(2), 168-182. https://doi.org/10.1080/10645578.2015.1079095

Powell, L. J., & Carey, S. (2017). Executive function depletion in children and its impact on theory of mind. *Cognition*, *164*, 150-162. https://doi.org/10.1016/j.cognition.2017.03.022

Prencipe, A., & Zelazo, P. D. (2005). Development of affective decision making for self and other: Evidence for the integration of first-and third-person perspectives. *Psychological Science*, *16*(7), 501-505. https://doi.org/10.1111/j.0956-7976.2005.01564.x

Quinn, G. P., & Keough, M. J. (2002). *Experimental Designs and Data Analysis for Biologists.* Cambridge University Press, Cambridge.

R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/

Raihani, N. J., & McAuliffe, K. (2014). Dictator game giving: The importance of descriptive versus injunctive norms. *PloS one, 9*(12), e113826. https://doi.org/10.1371/journal.pone.0113826

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. Nature Reviews Psychology, 1(4), 223-235. https://doi.org/10.1038/s44159-022-00037-z

Rakoczy, H., & Schmidt, M. F. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*(1), 17-21. https://doi.org/10.1111/cdep.12010

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental psychology*, *44*(3), 875. https://psycnet.apa.org/doi/10.1037/0012-1649.44.3.875

Rao, N., & Stewart, S. M. (1999). Cultural Influences on Sharer and Recipient Behavior: Sharing in Chinese and Indian Preschool Children. *Journal of Cross-Cultural Psychology*, *30*(2), 219-241. https://doi.org/10.1177/0022022199030002005

Raver, C. C., & Leadbeater, B. J. (2014). Factors influencing joint attention between socioeconomically disadvantaged adolescent mothers and their infants. In *Joint Attention* (pp. 251-271). Psychology Press.

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of public Economics, 88*(7-8), 1625-1644. https://doi.org/10.1016/S0047-2727(03)00021-5

Reinstein, D., & Riener, G. (2012). Reputation and influence in charitable giving: an experiment. *Theory and decision, 72*(2), 221-243. https://doi.org/10.1007/s11238-011-9245-8

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. Psychological science, 18(5), 407-413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Matthew, S., Newton, E. K., Naar, N., Newson, L., Ross, C., Smaldino, P. E., Waring, T. M., & Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences, 39*, 1-68. https://doi.org/10.1017/S0140525X1400106X

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. Neuroimage, 22(4), 1694-1703. https://doi.org/10.1016/j.neuroimage.2004.04.015

Rochat, P., Dias, M. D., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., & Berg, B. (2009). Fairness in distributive justice by 3-and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology*, *40*(3), 416-442. https://doi.org/10.1177/0022022109332844

Ronald, A., Viding, E., Happé, F., & Plomin, R. (2006). Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: A twin study. *Social neuroscience*, *1*(3-4), 412-425. https://doi.org/10.1080/17470910601068088

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological science*, *24*(1), 27-33. https://doi.org/10.1177/0956797612447819

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2007). Gender development. *Handbook of child psychology*, *3*. https://doi.org/10.1002/9780470147658.chpsy0314

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and US preschoolers. Psychological science, 17(1), 74-81. https://doi.org/10.1111/j.1467-9280.2005.01667.x

Sadeghi, S., Shalani, B., & Nejati, V. (2022). Sex and age-related differences in inhibitory control in typically developing children. *Early Child Development and Care*, *192*(2), 292-301. https://doi.org/10.1080/03004430.2020.1755668

Sally, D., & Hill, E. (2006) The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. Journal of Economic Psychology 27: 73-97. https://doi.org/10.1016/j.joep.2005.06.015

Samek, A.S. & Sheremeta, R.M. (2014) Recognizing contributors: an experiment on public goods. *Experimental Economics, 17*(4), 673–690. https://doi.org/10.1007/s10683-013-9389-1

Sansom, L. J., Minh, T. P. N., Hill, I. E., Ha, Q. N. T., Trong, T. D., Vidaillac, C., ... & Kestelyn, E. (2020). Towards a fair and transparent research participant compensation and reimbursement framework in Vietnam. *International health*, *12*(6), 533-540. https://doi.org/10.1093/inthealth/ihaa066

Saxon, T. F., & Reilly, J. T. (1999). Joint Attention and Toddler Characteristics: Race, Sex and Socioeconomic Status∗ . Early Child Development and Care, 149(1), 59-69. https://doi.org/10.1080/0300443991490105

Scharpf, F., Paulus, M., & Wörle, M. (2017). The impact of social relationships on Ugandan children's sharing decisions. *European Journal of Developmental Psychology*, *14*(4), 436-448. https://doi.org/10.1080/17405629.2016.1231062

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral ecology*, *20*(2), 416-420. https://doi.org/10.1093/beheco/arn145

Schirmbeck, K., Rao, N., & Maehler, C. (2020). Similarities and differences across countries in the development of executive functions in children: A systematic review. Infant and Child Development, 29(1), e2164. https://doi.org/10.1002/icd.2164

Schleihauf, H., & Hoehl, S. (2021). Evidence for a dual-process account of over-imitation: Children imitate anti-and prosocial models equally, but prefer prosocial models once they become aware of multiple solutions to a task. PLoS One, 16(9), e0256614. https://doi.org/10.1371/journal.pone.0256614

Schmidt, M. F., Butler, L. P., Heinz, J., & Tomasello, M. (2016). Young children see a single action and infer a social norm: Promiscuous normativity in 3-year-olds. *Psychological Science*, *27*(10), 1360-1370. https://doi.org/10.1177/0956797616661182

Schmidt, M. F. H., Vaish, A., & Rakoczy, H. (2024). Don't Neglect the Middle Ground, Inspector Gadget! There Is Ample Space Between Big Special and Small Ordinary Norm Psychology. *Perspectives on Psychological Science*, *19*(1), 69-71. https://doi.org/10.1177/17456916231187408

Schwartz, S. H. (2014). Rethinking the concept and measurement of societal culture in light of empirical findings. *Journal of cross-cultural Psychology*, *45*(1), 5-13. https://doi.org/10.1177/0022022113490830

Selcuk, B., Gonultas, S., & Ekerim-Akbulut, M. (2023). Development and use of theory of mind in social and cultural context. *Child Development Perspectives*, *17*(1), 39-45. https://doi.org/10.1111/cdep.12473

Shimpi, P. M., Akhtar, N., & Moore, C. (2013). Toddlers' imitative learning in interactive and observational contexts: The role of age and familiarity of the model. *Journal of experimental child psychology*, *116*(2), 309-323. https://doi.org/10.1016/j.jecp.2013.06.008

Shutts, K., Kinzler, K. D., Katz, R. C., Tredoux, C., & Spelke, E. S. (2011). Race preferences in children: Insights from South Africa. Developmental science, 14(6), 1283-1291. https://doi.org/10.1111/j.1467-7687.2011.01072.x

Silk, J. B., & House, B. R. (2012). 20 The Phylogeny and Ontogeny of Prosocial Behavior. *The Oxford handbook of comparative evolutionary psychology*, 381.

Silverman, I. W. (2021). Gender differences in inhibitory control as assessed on simple delay tasks in early childhood: A meta-analysis. *International Journal of Behavioral Development*, *45*(6), 533-544. https://doi.org/10.1177/01650254211020385

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260-274. https://doi.org/10.1016/j.cognition.2019.03.019

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PloS one, 8(*3), e59510. https://doi.org/10.1371/journal.pone.0059510

Smith, J. R., & Louis, W. R. (2008). Do as we say and as we do: The interplay of descriptive and injunctive group norms in the attitude–behaviour relationship. *British Journal of Social Psychology*, *47*(4), 647-666. https://doi.org/10.1348/014466607X269748

Smith, R., & LaFreniere, P. J. (2009). Development of children's ability to infer intentions from nonverbal cues. *Journal of Social, Evolutionary, and Cultural Psychology*, *3*(4), 315. https://doi.org/10.1037/h0099313

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. Developmental psychology, 51(9), 1190. https://doi.org/10.1037/dev0000039

Soetevent, A. R. (2005). Anonymity in giving in a natural context—a field experiment in 30 churches. *Journal of public Economics, 89*(11-12), 2301-2323. https://doi.org/10.1016/j.jpubeco.2004.11.002

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes what do they signify?. *Psychology of women quarterly*, *24*(1), 44-62.https://doi.org/10.1111/j.1471-6402.2000.tb01021.x

Spiegler, O., & Leyendecker, B. (2017). Balanced cultural identities promote cognitive flexibility among immigrant children. Frontiers in Psychology, 8, 1579. https://doi.org/10.3389/fpsyg.2017.01579

Staller, A., & Petta, P. (2001). Introducing emotions into the computational study of social norms: A first evaluation. Journal of artificial societies and social simulation, 4(1), U27-U60.

Stengelin, R., Hepach, R., & Haun, D. (2019). Being observed increases overimitation in three diverse cultures. *Developmental Psychology*, *55*(12), 2630.

Sternberg, R. J., Grigorenko, E., & Bundy, D. A. (2001). The predictive value of IQ. Merrill-Palmer Quarterly, 47(1), 1-41. https://doi.org/10.1353/mpq.2001.0005

Subiaul, F., Vonk, J., Okamoto-Barth, S., & Barth, J. (2008). Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal cognition*, *11*, 611-623. https://doi.org/10.1007/s10071-008-0151-6

Suhonen, E., Sajaniemi, N. K., Alijoki, A., & Nislin, M. A. (2018). Children's biological givens, stress responses, language and cognitive abilities and family background after entering kindergarten in toddlerhood. Early Child Development and Care, 188(3), 345-358. https://doi.org/10.1080/03004430.2016.1218157

Sun, L., Sabanathan, S., Thanh, P. N., Kim, A., Doa, T. T. M., Thwaites, C. L., ... & Wills, B. (2019). Bayley III in Vietnamese children: lessons for cross-cultural comparisons. Wellcome open research, 4, 98. https://doi.org/10.12688/wellcomeopenres.15282.1

Takagishi H, Kameshima S, Schug J, Koizumi M, Yamagishi T (2010) Theory of mind enhances preference for fairness. Journal of Experimental Child Psychology 105: 130-137. https://doi.org/10.1016/j.jecp.2009.09.005

Takagishi, H., Fujii, T., Koizumi, M., Schug, J., Nakamura, F., & Kameshima, S. (2015). The development of the effect of peer monitoring on generosity differs among elementary school-age boys and girls. *Frontiers in psychology*, *6*, 895. https://doi.org/10.3389/fpsyg.2015.00895

Taniguchi, Y., & Sanefuji, W. (2017). The boundaries of overimitation in preschool children: Effects of target and tool use on imitation of irrelevant actions. *Journal of Experimental Child Psychology*, *159*, 83-95. https://doi.org/10.1016/j.jecp.2017.01.014

Tennie, C., Call, J. & Tomasello, M. (2009) Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–15. https://doi.org/10.1098/rstb.2009.0052.

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development*, *12*(2), 199-212 .https://doi.org/10.1016/S0885-2014(97)90013-7

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2016). Cultural learning redux. Child development, 87(3), 643-653. https://doi.org/10.1111/cdev.12499

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Carpenter, M. (2005). Intention reading and imitative learning. *Perspectives on imitation: From neuroscience to social science*, *2*, 133-148.

Tomasello, M., & Warneken, F. (2008). Share and share alike. *Nature*, *454*(7208), 1057-1058. https://doi.org/10.1038/4541057a

Townsend, C., Aktipis, A., Balliet, D., & Cronk, L. (2020). Generosity among the Ik of Uganda. *Evolutionary Human Sciences*, *2*, e23. https://doi.org/10.1017/ehs.2020.22

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological review*, *96*(3), 506. https://psycnet.apa.org/doi/10.1037/0033-295X.96.3.506Turiel, E. (1983). *Development of social knowledge*. NY: Cambridge University Press.

Užgiris IC. 1981. Two functions of imitation during infancy. Int. J. Behav. Dev. 4:1–12

Urassa, M., Lawson, D. W., Wamoyi, J., Gurmu, E., Gibson, M. A., Madhivanan, P., & Placek, C. (2021). Cross-cultural research must prioritize equitable collaboration. Nature Human Behaviour, 5(6), 668-671. https://doi.org/10.1038/s41562-021-01076-x

Utech, D. A., & Hoving, K. L. (1969). Parents and Peers as Competing Influences in the Decisions of Children of Differing Ages. The Journal of Social Psychology, 78(2), 267-274. https://doi.org/10.1080/00224545.1969.9922366

Uz, I. (2015). The index of cultural tightness and looseness among 68 countries. *Journal of Cross-Cultural Psychology*, *46*(3), 319-335. https://doi.org/10.1177/0022022114563611

Vaish, A., Carpenter, M., & Tomasello, M. (2016). The early emergence of guilt-motivated prosocial behavior. Child development, 87(6), 1772-1782. https://doi.org/10.1111/cdev.12628

van Hoorn, J., van Dijk, E., Meuwese, R., Rieffe, C., & Crone, E. A. (2016). Peer influence on prosocial behavior in adolescence. *Journal of research on adolescence*, *26*(1), 90-100. https://doi.org/10.1111/jora.12173

Vierhaus, M., Lohaus, A., Kolling, T., Teubert, M., Keller, H., Fassbender, I., ... & Schwarzer, G. (2011). The development of 3-to 9-month-old infants in two cultural contexts: Bayley longitudinal results for Cameroonian and German infants. *European Journal of Developmental Psychology*, *8*(3), 349-366. https://doi.org/10.1080/17405629.2010.505392

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, *77*(3), 640-663. https://doi.org/10.1111/j.1467-8624.2006.00895.x

Warneken, F., & Tomasello, M. (2013). Parental presence and encouragement do not influence helping in young children. *Infancy*, *18*(3), 345-368. https://doi.org/10.1111/j.1532-7078.2012.00120.x

Watson-Jones, R. E., Legare, C. H., Whitehouse, H., & Clegg, J. M. (2014). Task-specific effects of ostracism on imitative fidelity in early childhood. *Evolution and Human Behavior*, *35*(3), 204-210. https://doi.org/10.1016/j.evolhumbehav.2014.01.004

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, *288*(5467), 850-852. https://doi.org/10.1126/science.288.5467.850

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of Theory-of-Mind Understandings in Chinese Children. Psychological Science, 17(12), 1075-1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child development*, *82*(3), 780-792. https://doi.org/10.1111/j.1467-8624.2011.01583.x

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, *75*(2), 523-541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wharton Credibility Lab. (n.d.). *AsPredicted.* https://aspredicted.org/

Whiten, A. (2021). The burgeoning reach of animal culture. *Science*, *372*(6537), eabe6514.. https://doi.org/10.1126/science.abe6514

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2417-2428. https://doi.org/10.1098/rstb.2009.0069

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., ... & Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682-685. https://doi.org/10.1038/21415

Whiten, A., Allan, G., Devlin, S., Kseib, N., Raw, N., & McGuigan, N. (2016). Social learning in the real-world:'Over-imitation'occurs in both children and adults unaware of participation in an experiment and independently of social interaction. *PloS one*, *11*(7), e0159920. https://doi.org/10.1371/journal.pone.0159920

Whiting, B., & Edwards, C. P. (1973). A cross-cultural analysis of sex differences in the behavior of children aged three through 11. *The Journal of Social Psychology*, *91*(2), 171-188. https://doi.org/10.1080/00224545.1973.9923040

World Bank Photo Collection. (n.d.). *Uganda*. Flickr. flickr.com/photos/worldbank/albums/72157630544765096/with/7556640330/

Wu, Z., Pan, J., Su, Y., & Gros-Louis, J. (2013). How joint attention relates to cooperation in 1-and 2-year-olds. International Journal of Behavioral Development, 37(6), 542-548. https://doi.org/10.1177/0165025413505264

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., Emmen, R. A., & Prevoo, M. J. (2014). Cognitive flexibility children across the transition to school: A longitudinal study. Cognitive Development, 31, 35-47. https://doi.org/10.1016/j.cogdev.2014.02.004

Yu, C. L., & Wellman, H. M. (2024). A meta-analysis of sequences in theory-of-mind understandings: Theory of mind scale findings across different cultural contexts. Developmental Review, 74, 101162. https://doi.org/10.1016/j.dr.2024.101162

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PloS one, 8(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Zarbatany, L., Hartmann, D. P., Gelfand, D. M., & Vinciguerra, P. (1985). Gender differences in altruistic reputation: Are they artifactual?. *Developmental psychology*, *21*(1), 97.

Zelazo, P. D. (2004). The development of conscious control in childhood. Trends in cognitive sciences, 8(1), 12-17. https://doi.org/10.1016/j.tics.2003.11.001

Zelazo, P. The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nat Protoc* **1**, 297–301 (2006). https://doi.org/10.1038/nprot.2006.46

Zelazo, P. D., & Lourenco, S. F. (2003). Imitation and the dialectic of representation. *Developmental Review*, *23*(1), 55-78. https://doi.org/10.1016/S0273-2297(03)00005-4

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... & Carlson, S. M. (2003). The development of executive function in early childhood. Monographs of the society for research in child development, i-151. https://doi.org/10.1111/j.0037-976X.2003.00261.x

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child development perspectives*, *6*(4), 354-360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Qu, L., Kesek, A. C., Calkins, S. D., & Bell, M. A. (2010). Child development at the intersection of emotion and cognition. *American Psychological Association*, 97-111.