The Role of Culture in Children's Sex-Typed Preferences for Colours, Toys, and Affordances: A Systems Theory Approach

Jacqueline Topsy Mengersen Davis

Clare College

University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Department of Psychology

School of Biological Sciences

Submitted January 2018

Word Count: 57 140

The Role of Culture in Children's Sex-Typed Preferences for Colours, Toys, and Affordances: A Systems Theory Approach

Jacqueline T. M. Davis

Summary

Children's sex-typed preferences for colours and toys are well-established, and often function as markers of sex-typicality in research on the development of sex-typed behaviour. However, children's sex-typed colour and toy preferences have not been tested cross-culturally, or in remote unindustrialised cultural settings. The present thesis tested children's preferences for sex-typed toys in four cultural settings: Shipibo villages in the Lake Imiria region of the Peruvian Amazon; *kastom* villages in the mountains of Tanna Island in Vanuatu in the South Pacific; children attending school in Lenakel town on Tanna Island; and in a large industrialised city in Australia. It also tested children's colour preferences in three of these cultures. It was hypothesised that colour and toy preferences would show some similarities across cultures, and further, that similarities in toy preferences across cultures would be explained by the different types of play afforded by the toys. Results suggested that colour preferences, specifically, a sex difference in preference for pink, are specific to industrialised cultures. Results further suggested that some sex differences in toy preferences replicate in different cultures, and that the relationship between toy preferences and children's preferences for play affordances is a potentially important area for further research. The present thesis also provided two demonstrations of how new statistical methods, adapted from complex and dynamic systems theory, could be applied to the cross-cultural dataset. A machine learning method suggested that sex, more than culture, affects children's sex-typed toy preferences. A multistate dynamic method further suggested that sex, more than culture, affects the dynamics of children's toy choices.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and Acknowledgements, and as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Faculty of Biological Sciences.

Acknowledgements

Many people contributed to the work presented in this thesis. First, my supervisor, Professor Melissa Hines, gave incredible academic inspiration, many illuminating discussions, practical advice, and endless support and patience. Thank you for mentoring me and for championing this research.

Ellen Robertson, another PhD student supervised by Prof. Hines, travelled with me to Peru and was a core part of the research effort there, including translation, data collection, and interviews.

Professor Mark Nielsen supported this research, and his students Karri Neldner and Rohan Kapitany, travelled with me to Vanuatu, and allowed me to access their study site in Brisbane.

The videos were coded by undergraduate students as part of their research projects: thank you Quain, Jess, Demos, Sophie, Emily, and Isabelle, for doing an excellent job.

This research was supported by funding from Gates Cambridge, the University of Cambridge Department of Psychology, Clare College, the Commonwealth Australia Awards, and the Worts and Smuts Travelling Scholars' funds. Field site access costs were shared with the Queensland University of Technology Eyes on the Wild project (Peru), and the University of Queensland Department of Psychology (Vanuatu and Brisbane).

Thanks to Ronel Cairuna Garcia for research coordination and translation in Peru. Further thanks go to the Directors and families of Caimito, Buenos Aires, Nuevo Loreto, and Nueva Yarina, for hosting and supporting the research.

Thanks to Beverline Mahana for research coordination and translation in Vanuatu. Further thanks go to Joe Narua, Numaline Mahana, and Annie Loughman (cultural information), and Naoke Joseph (JJ), Robert, Selina, Mary, and Rachel (translation). Special thanks to Chief Jacob Kapere (rest in peace), Chief Yapa, and the Principal of Lenakel Harbourview, Stephen. Thanks to members of the Vanuatu Cultural Centre, Tafea Cultural Centre, the Vanuatu Women's Program, and families of Ikunala and Yakel and surrounding villages.

Huge thanks to Professor Kevin Burrage, for his demonstration of Markov models and for helpful discussions.

Members of my research group, the Gender Development Research Centre in Cambridge, provided a supportive and academically stimulating environment, moral support, and many cups of tea. Thanks also to members of my funding group, the Gates Cambridge Scholars, and members of my college, Clare College, for social and practical support.

My family have been a huge help, and I couldn't have done it without them. Thank you to Jay for being a wonderful partner and friend. Finally, my mother, Kerrie Mengersen, who first suggested the trip to Peru after my first fieldwork plans were derailed by Cyclone Pam, and who has provided statistical advice, technical advice, and tonnes of personal support. Thank you for your constant guidance.

Contents

Chapter 1 Literature Review ... 8

Key Terms ... 8

General Background ... 8

Chapter 2 Overview ... 26

Aims and Rationale ... 26

Study Scope ... 27

General Methods ... 30

Thesis Structure ... 42

Chapter 3 Children's Preferences for Blue-Red Hues and High-Low Colour Saturation in Three Cultures ... 48

Background Literature ... 48

Method ... 54

Results ... 59

Discussion ... 64

Chapter 4 Children's Preferences for Sex-Typed Toys in Four Cultures ... 70

Background Literature ... 70

Method ... 74

Results ... 78

Discussion ... 85

Chapter 5 Children's Preferences for Toys with Faces, Propulsion, Dress-Up, and Role Play Toys in Four Cultures ... 90

Background Literature ... 90

Method ... 91

Results ... 92

Discussion ... 97

Chapter 6 A Machine Learning Model of Sex, Culture, and Affordance in Children's Toy Choices ... 101

Background Literature ... 101

Method ... 103

Results ... 106

Discussion ... 110

Chapter 7 A Multistate Dynamic Model of Sex and Culture in Children's Toy Choices ... 114

Background Literature ... 114

Method ... 116

Results 123
Discussion 132
Chapter 8 General Discussion 137
Summary of Chapters 137
Limitations and Caveats 138
Synthesis of Study Results 141
Summary of Contributions 149
References 151
Appendices 171
Appendix 1: Colour Preferences Frequency Tables and Charts for Planned Comparisons 172
Appendix 2: Toy Preferences Summary Statistics 178

Chapter 1

Literature Review

Key Terms

In this thesis, the behaviours and preferences that may theoretically differ, on average, between boys and girls, or between men and women, are referred to as *sex-typed*. The noun *sex* is used, rather than *gender*, because in the cross-cultural study that is the basis for the thesis, children were categorised as boys or girls based on their biological sex, and were not asked about their gender identity. The participle *typed* is used, rather than *related* or *differentiated*, because in the present study, stimuli were chosen and categorised according to sex stereotypes, and were not categorised according to preferences observed in the dataset.

Children's toys that are, on average, expected to be preferred by boys are termed *boy-type toys* (BTT); and toys that are, on average, expected to be preferred by girls are termed *girl-type toys* (GTT). Again, *type* is used to clarify that the expected sex difference is based on stereotypes, and not necessarily on observed preferences.

For consistency, differences between boys and girls, or between men and women, are referred to as *sex differences*. For brevity, *sex-typed behaviour* refers to all the behaviours, including preferences, that have been theorised, stereotyped, or observed to show sex differences.

Regarding cultural contexts, broad terms such as *third world* and *developing* are arguably out of date, and may carry political and value labels. Therefore, in this thesis, parts of the world that do not have access to mass media, mass communication, and industrialised mass-produced goods are referred to as *majority world*, to reflect that most of the world's population lives under these or similar conditions. The remaining cultural contexts have, in general, been described in previous work as *developed*, *Western*, *industrialised*, or *WEIRD* (Henrich, Heine, & Norenzayan, 2010c) but are referred to here as *minority world*, to reflect that the minority of the world's population lives under these conditions. Wherever possible, this thesis will refer to specific cultural contexts.

General Background

The work presented in this thesis represents the intersection of two major branches of psychology: first, theoretical approaches to the development of behavioural *sex differences*, and second, the documentation and explanation of *cultural differences*. These branches will be examined and integrated in the current thesis using *systems theory*, but first, it is useful to give a brief history of each. The current chapter provides a general background to the thesis. Each individual chapter then presents a more focused review of literature relevant to the research question explored in that chapter.

Part 1 of the General Background provides a broad overview of recent approaches to the study of sex differences in human behaviour. It traces the history of theoretical developments in the study of sex-typed behaviour, from early approaches, through social learning and cognitive theories, to hormonal theories, with examples from research on sex differences in children's toy preferences. It argues that with each new theoretical development, sex and gender were seen as more complex and dynamic than they previously had been.

Part 2 of the General Background provides a brief review of recent approaches to the study of cultural differences in human behaviour. It describes some of the ways that cross-cultural data have been integrated into psychology, including ethnographic approaches, evolutionary approaches, and cross-cultural studies, with examples from research on sex-typed behaviour. It reviews previous research on the role of culture in children's sex-typed colour and toy preferences, and it identifies play affordance as a potential functional aspect of sex-typed toys. It concludes by identifying the possibilities for a cross-cultural study of children's sex-typed preferences for colours, toys, and affordances.

Part 1
Sex Differences

Early Approaches

Early scientific approaches to sex-typed behaviour were built on the assumption that masculinity and femininity arose, by definition, from the two biological sexes (Person & Ovesey, 1983). The logical extension of this idea, as stated by Anastasiow (1965), was that any observation of feminine behaviour in boys, or masculine behaviour in girls, was undesirable for healthy development. Consequently, research efforts often focused on atypical sex-typed behaviour, and aimed to correct it.

For example, Rekers and Yates (1976) used atypical sex-typed toy preferences to identify boys who were clinically diagnosed with “childhood gender disturbance” (p. 2) . They tested toy preferences in a free play session, where each child individually was given 5 minutes to play with a set of toys. The toys were arranged on two tables. One table contained girls’ apparel (cosmetics, a wig, shoes, a dress, and jewellery) and boys’ apparel (a football helmet, an army helmet, a sea captain’s hat, an army shirt and belt, and a pretend razor). The other table contained toys that the authors determined to have sex-typed characteristics: girls’ “maternal nurturance” toys (a baby doll in a crib with accessories, and a Barbie doll with two sets of clothes) and boys’ “masculine aggression” toys (dart guns, a rubber knife, plastic handcuffs, and a set of cowboys and Indians) (p. 3). The “gender disturbed” boys were more likely than the control boys were to play with the girls’ dress-up and maternal nurturance toys, and were similarly likely to play with these toys as were the control girls. The authors concluded that a free play session, with multiple toys, could be used as a clinical assessment technique for diagnosing children’s gender disturbance.

More recent approaches are less likely to treat variations in sex-typed behaviour as problematic. But, this early focus on diagnosis was important, because a diagnosis of gender disturbance required an empirical measurement of sex-typed behaviour. In the process of trying to diagnose gender disturbance, Rekers and Yates confirmed an assumption that was foundational to later studies: that children’s sex-typed toy preferences could be used to make inferences about the children’s broader sex-typed development. Children’s toy preferences have since been widely studied as a marker of sex-typed behaviour more broadly[1].

Behaviourism and Social Learning Theories

In contrast to early approaches that saw sex-typed behaviour as an extension of genetic sex, behaviourist approaches saw sex-typed behaviour as something that could be learned. Social learning theorists, such as Bandura (1969) and Mischel (1966), suggested that behavioural modelling was a primary way that children learned sex-typed behaviour.

[1] In later years, however, researchers argued that different aspects of sex-typed behaviour might develop on different pathways, and that toy preferences could not always be used as a marker for other sex-typed behaviours (e.g., Hines, 2005; Liben & Bigler, 2002).

Children copied the sex-typed behaviours that were modelled to them by parents, teachers, and peers, and that they saw in the media.

For example, Barkley and colleagues (Barkley, Ullman, Otto, & Brecht, 1977) found that children changed their sex-typed play behaviour, including toy preferences, depending on what they observed. Barkley and colleagues presented children with a video that showed a model acting out either masculine behaviour (a story with a block, a crayon, and a truck), feminine behaviour (a story with a block, a crayon, a toy kettle, and a toy stove), or a control video. Children who viewed the feminine behaviour imitated more feminine behaviour in a subsequent play session, and a similar trend was observed for masculine behaviour. The authors suggested that children's sex-typed preferences could be influenced by their social environment, because boys and girls are likely to imitate the preferences of their playmates, but play primarily with peers of their own sex.

Other behaviourist research focused on identifying the types of reinforcement that children might receive for sex-typed behaviour. For example, Fagot and Patterson (1969) found that children were reinforced for behaviour that was sex-typed for the child's sex, but not for behaviour that was sex-typed for the other sex. Fagot and Patterson observed the behaviour of children and teachers in a nursery school, over an entire school year. The nursery school teachers positively reinforced feminine behaviours in girls, and masculine behaviours in boys. Peers also positively reinforced the sex-typed behaviour of other children. Later research documented that parents also positively reinforced sex-typed behaviour in their children (Fagot & Hagan, 1991; Fagot, Leinbach, & O'Boyle, 1992). Behaviourists therefore hypothesised that sex-typed behaviour could develop through a combination of behavioural modelling and reinforcement.

Feminism and Cognitive Theories

The rise of feminist thought in the 1960s and 1970s saw a concurrent surge in academic interest about the nature and development of sex-typed behaviour. Feminist thinking brought major conceptual advances, especially an understanding of sex and gender as multidimensional. Theorists identified a difference between biological *sex*, and *gender* as a set of personality traits, social attributes, relationships, interests, and behaviours (reviewed in Huston, 1985; Money, 1973). These multidimensional concepts of sex and gender, and the assignment of distinct definitions to sex and gender, influenced subsequent research and theory.

At the same time, theories of cognitive development were increasingly applied to research on sex differences. Kohlberg (1966) had previously proposed that children actively constructed their gender as part of their larger cognitive construction of their world. Sex differences in children's behaviour were thought to arise from children's understanding that people were male or female, and that they themselves were male or female. Children's knowledge and beliefs about gender were collectively termed *gender schemas* (Bem, 1981; Martin & Halverson, 1981). These schemas were thought to be important to children's demonstrated sex-typed behaviours, as well as to children's attention to new information about gender.

For example, Fagot (1985) related children's toy preferences to their ability to recognise gender in other people. She observed children playing with toys in a free play situation, and then asked children to label pictures of men and women as male or female. She observed that the boys who could not correctly label the pictures of men and women played with dolls about as much as girls did. She concluded that children's sex-typed toy preferences might be partially a product of their ability to understand gender, and predicted that children who had not yet gained an understanding of gender would not show sex differences in toy preferences. Subsequent research also found that children who passed gender knowledge tasks played more with sex-typed toys than children who failed (O'Brien & Huston, 1985). However, not all replication attempts were successful, and other investigations failed to find a connection between gender labelling and toy preferences (Fagot, Leinbach, & Hagan, 1986; Weinraub et al., 1984).

The relationship between cognitive processes and social processes in gender development is a topic of ongoing research. According to the cognitive perspective, children's cognitive understanding of gender causes the development of sex-typed preferences, including preferences for sex-typed toys. However, studies have found sex-typed toy preferences even in children who were too young to have developed the required cognitive understanding. For example, in one study, children younger than two years old could not correctly identify sex-typed objects above chance levels (Blakemore, LaRue, & Olejnik, 1979), but in another study, children younger than two years old preferred to play with sex-typed toys over other toys (O'Brien, Huston, & Risley, 1983). These apparent inconsistencies led theorists to develop perspectives involving interactions between cognitive and social processes.

The interactions between cognitive and social processes were clarified in theoretical approaches that sought to include elements of children's internal psychology together with elements of their external social environment. One such approach was social-cognitive theory (Bussey & Bandura, 1999). According to social-cognitive theory, children's social experiences were fundamental to the construction of their gender concepts. These gender concepts then guided children's motivations and self-regulation in their further interactions with the environment. Social-cognitive theory may be contrasted with cognitive-developmental theory, wherein children's cognitive development was seen as fundamental to children's attention and responses to social experiences (Martin, Ruble, & Szkrybalo, 2002). A further approach, the dual pathways model, posited the existence of two reciprocal causal pathways: an attitudinal pathway, in which children's knowledge of gender stereotypes affected their interests; and a personal pathway, in which children's interests affected their stereotypes (Liben & Bigler, 2002). This is not an exhaustive list, but the key feature of these integrative approaches is that gender development is seen as an active construction that works within a larger social environment, thus allowing for a reciprocal relationship between cognitions and behaviour, and for parallel development of sex-typed preferences and gender concepts.

Increasing Interest in Biological and Hormonal Theories

Recent research on the role of biology in sex and gender development has increasingly seen it as multidimensional and changeable. Although some early approaches regarded biology as fixed and deterministic (e.g., Anastasiow, 1965), others suggested that biology might be flexible and responsive to the environment, based on research in animal models (Arnold, 1985; Maccoby, 1966; Phoenix, Goy, Gerall, & Young, 1959). More recently, a growing body of research has documented these flexible and responsive processes in humans, perhaps based in part on the increasing availability of new technologies such as genetic testing, neuroimaging, and hormone assays (Cahill, 2006; Salk & Hyde, 2012). Like gender, sex is increasingly seen as a multidimensional construct with complex inter-related components, including genetic, developmental, and hormonal components (Fausto-Sterling, 2012a; Joel & Fausto-Sterling, 2016). Consequently, recent research characterises the relationship between human sex-typed behaviour and biology as a complex developmental cascade, rather than as fixed and deterministic (Hines, 2013; LeVay, 2011; Wallen, 2009).

The focal point of recent biological theories of gender development has been hormones; particularly, the role of sex hormones in sex-typed behaviour. Many studies of sex development in non-human mammals (e.g., Arnold, 1985; McCarthy, 2010) noted that behaviours that showed a sex difference were influenced by early exposure to androgens, including testosterone. Specifically, during certain periods of prenatal and early postnatal development, testosterone is higher in male than in female mammals, and this hormone difference was found to be partially responsible for later sex differences in behaviour. Researchers theorised that, since toy preferences are a sex-typed behaviour in humans, they might also be affected by early exposure to sex hormones (Hines, 2005, 2006).

The importance of early sex hormone exposure to later sex-typed development is apparent in children with genetic conditions that cause atypical hormone exposure prenatally. The most frequently studied condition in regard to sex-typed behaviour is congenital adrenal hyperplasia (CAH). CAH is an autosomal, recessive condition that causes deficiency in cortisol production, and consequent increased production of androgens. Female foetuses with CAH are therefore exposed to concentrations of androgens, including testosterone, that are markedly elevated compared to those in female foetuses without CAH. Across several studies, researchers noted that girls with CAH made sex-typed toy choices that were somewhere between those made by boys and girls without CAH (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Pasterski et al., 2005; Servin, Nordenström, Larsson, & Bohlin, 2003). The severity of the CAH condition was also related to the degree of sex-typed toy preferences, with the most dramatic masculinisation of toy preferences demonstrated by girls with the most severe forms of CAH (Nordenström, Servin, Bohlin, Larsson, & Wedell, 2002; Servin et al., 2003). Researchers concluded that prenatal exposure to sex hormones might influence children’s later sex-typed toy preferences.

Hormonal effects on sex-typed play have also been measured in typically-developing children, but the findings of these studies were more variable than the results of studies that focused on CAH. For example, Lamminmäki and colleagues measured testosterone in infant urine monthly between 7 days postnatal and age 6 months postnatal, a period of development when testosterone is elevated in boys, more so than in girls. When infants were 14 months of age, they were observed playing with toys in a free play session. Lamminmäki and colleagues hypothesised that higher concentrations of testosterone during early infancy would relate to more boy-type toy preferences, and found that play with a train correlated significantly and positively with testosterone in girls, while play with a baby doll correlated significantly and negatively with testosterone in boys (Lamminmäki et al., 2012). In contrast, van de Beek and colleagues did not find associations between testosterone and later toy preferences, when they measured testosterone in samples of maternal blood during pregnancy and in amniotic fluid at mid-gestation (van de Beek, van Goozen, Buitelaar, & Cohen-Kettenis, 2009). These variable results for studies measuring hormones in typically developing children may suggest that there is no relationship between early hormone exposure and later sex-typed toy preferences in typically-developing children. However, methods for measuring the early hormone environment in typically-developing children are still developing, and may currently be insufficiently sensitive to produce consistent results (Hines, 2013).

Conclusions: The Increasing Complexity of Sex and Gender

Over time, the expanding body or research on sex and gender has led to increased understanding that sex-typed behaviour is complex and dynamic, and is not a simple function of only learning, cognition, or biology. As the above review showed, factors involved in the development of sex-typed behaviour have been a topic of research interest for decades. But the focus of this research has changed, as wider academic trends, and available technology, changed over time. Each shift in research focus seems to have concluded that sex-typed behaviour is more complex, and more dynamic, than previously thought. First, early research saw sex-typed behaviour as a logical extension of biology. Then, behaviourism showed that behaviour, even sex-typed behaviour, might be flexibly learned. Research also shifted to emphasize the multidimensional and flexible nature of gender as a social construct, and researchers became interested in how gender might be cognitively constructed. Lately, new research has suggested that biology, too, is multidimensional and flexible, and there has been increased interest in biological contributions to sex-typed behaviour. With each theoretical development, sex-typed behaviour seems to have acquired more inter-related, changeable, dimensions than before.

These theoretical developments, however, were all made in a specific cultural context: the Western, industrialised, usually English-speaking, world. Results from this cultural context were assumed to generalise across humans, but only a minority of the world's population lives under these conditions. Furthermore, many of these theoretical developments did not explicitly address the role of culture. The following section reviews how culture has been integrated into psychological research, with a focus on sex-typed behaviour.

General Background Part 2

Cultural Differences

Critical Ethnography and Cultural Variation

Early approaches to cultural differences in sex-typed behaviour often related to broader developmental psychology theories. For example, when Freudian psychoanalytic theory was influential, ethnographic studies described cultural differences in child feeding, toilet training, and sexuality (e.g., the Trobriand case, reviewed in Spiro, 1984). Ethnographers used systematic observations to compare child behaviour across different cultural contexts. Research aimed to identify universal patterns in human behaviour, and to relate these patterns to psychological theories.

Ethnographers in this early period also began to use cross-cultural results to challenge the universality of influential theories in developmental psychology, initiating a tradition of critical ethnography that would persist throughout later research in child development. Some ethnographers proposed that psychological theories were universal only until falsified by a single counter-example from cross-cultural research (Kaufmann, 1944). For example, in 1927, Malinowski used evidence from matrilineal societies in the Trobriand Islands to discount the universality of Freud’s concept of the Oedipus complex, and proposed a competing matrilineal complex (reviewed in Malinowski, 2013). Counter-examples, that could potentially falsify developmental theory, would continue to be a key contribution of cross-cultural research.

Critical ethnography became more pronounced with the rising popularity of cognitive accounts of child development. The empirical predictions of cognitive theory were stronger than those of psychoanalytic theory, and therefore were more susceptible to disruption through counter-examples from other cultures. For example, Gilligan (1982) challenged Kohlberg’s stages of moral development (pre-conventional morality, conventional morality, and post-conventional morality; Kohlberg, 1976), contending that these were applicable only to development in boys, and that girls had different patterns of development. Gilligan further contended that boys’ moral development revolved around the construction of artificial hierarchies in ethical priorities, or a “morality of justice”, while girls’ moral development revolved around the consideration of relational networks, or the “morality of care” (p. 5). However, an attempt to document this framework in an Indian Hindu population gave no support for either Kohlberg’s universal stages or Gilligan’s justice-care dichotomy (Miller, 1994). Instead, the cross-cultural study suggested that moral codes reflected larger cultural systems of meaning; whereas American moral development focused on freedom of choice and individual responsibility, Hindu moral development focused on interpersonal obligations and contextual sensitivity.

However, developmental psychology and cross-cultural research have not always been linked. The wider field of developmental psychology sometimes responded to ethnographic critiques, but often ignored them (e.g., Gilligan, 1995; reviewed in LeVine, 2007). Ethnographers were faced with the simultaneous difficulties of gathering cross-cultural results to test psychological theories, and then getting psychologists to acknowledge these results, and often reacted by turning their attention away from testing universal theories in developmental psychology (LeVine, 2007). Instead, many ethnographers emphasised variability between cultures. For example, several studies have described the different ways that children learn new skills in different cultures, without aiming to test psychological theories directly (e.g., Bock, 2002; Boyette, 2016b; Fry, 1992; Hewlett, Fouts, Boyette, & Hewlett, 2011). The focus of these cross-cultural studies is often to gather information on broadly defined parts of life, such as child development, in a way that allows this information to be compared between cultures.

The Six Cultures Study was the first to use systematic observations, across multiple cultures, to provide a comparative dataset on child development. A single detailed field manual (Whiting, Child, Lambert, & others, 1966) was used to collect comparable data in a Mixtecan community in Mexico (Romney & Romney, 1966), an Ilongot community in the Philippines (Nydegger & Nydegger, 1966), a Rajput community in India (Minturn & Hitchcock, 1966), an Okinawan community in Japan (Maretzki & Maretzki, 1966), a Gusii community in Kenya (LeVine & Lloyd, 1966), and a small town in Massachusetts in the US (Fischer & Fischer, 1966). The ethnography was constructed based on naturalistic observations of children, engaged in routine behaviours, in a natural setting. Although sex-typed behaviour was not an original focus of the study, later reanalyses reported sex differences in role play (Edwards, 2000). Since the Six Cultures Study, other cross-cultural ethnographies have generally focused on documenting variations in world cultures, rather than on testing psychological theory (Bloch & Adler, 1994; Boyette, 2016a; Roopnarine, Nathan, & Pellegrini, 2010; Takeuchi, 1994; Whitam & Mathy, 1991).

Evolutionary Theory and Human Behavioural Ecology

Human behavioural ecology applies evolutionary theory to the study of human behaviour. In contrast to ethnographic approaches, that focus on variability, behavioural ecology focuses on universal principles of survival and reproduction (Tudge, Brown, & Freitas, 2010). These universal principles include an important distinction: that behaviour should be regarded in terms of its function, and not in terms of its physiological components (Bateson & Laland, 2013; Tinbergen, 1963).

Behavioural ecology aims to identify universal functions of behaviour, termed *behavioural adaptations*. One of the features of a behavioural adaptation is that it should emerge across all members of a species (reviewed in Schmitt & Pilcher, 2004). In humans, therefore, if a behaviour were observed across cultures, then it might be an adaptation, and might serve some universal evolutionary function (see Caro & Borgerhoff Mulder, 1987). Cross-cultural studies informed by a behavioural ecology approach therefore sought to identify behaviours and preferences that were *culturally universal*. Culturally universal preferences were assumed to have a basis in human evolutionary history (e.g., Buss, 1989).

More recent approaches to human behavioural ecology seek to integrate factors such as geopolitical history, economic forces, cultural transmission, and physical environments into evolutionary theory. The relative importance and interactions between these forces are still under debate. For example, the cultural-ecological framework (Bloch, 1989) posited that the child's immediate setting interacted with historical influences and cultural beliefs, to influence children's play. In contrast, the developmental niche (Super & Harkness, 1986), also acknowledged the child's immediate setting, but emphasised child care customs and caretaker psychology. In general, integrative approaches emphasise historical context and geographic variation, as well as culture (e.g., Lamb, Sternberg, Hwang, & Broberg, 2014), and acknowledge that humans influence the environment at least as much as they are influenced by it (e.g., Laland & Brown, 2006).

Cross-Cultural Experiments

Psychology, including developmental psychology, has recently seen renewed interest in cross-cultural research. Developmental psychologists have typically generalised their conclusions to all humans (e.g., Alexander, 2003; Escudero, Robbins, & Johnson, 2013; Freeman, Sims, Kutsch, & Marcon, 1995; Green, Bigler, & Catherwood, 2004). However, participants in developmental psychology research have usually been recruited from minority world countries, particularly the United States and the United Kingdom. Recent research in experimental psychology has suggested that results from these participants may not generalise across cultures (Henrich, Heine, & Norenzayan, 2010b).

An influential review (Henrich, Heine, & Norenzayan, 2010a) found substantial variation between cultures in many domains of experimental psychology. These cultural variations were noticeable, even in some domains that the authors noted as “fundamental” to human development, such as visual perception, cooperation, and spatial reasoning. Additionally, the review found substantial differences between participants in populations that were typically studied, and participants recruited from the majority world. The authors cautioned that the populations from which psychological research had drawn were “among the least representative populations one could find for generalising about humans” (p. 61).

Research on sex-typed behaviour also suffered from this reliance on limited samples. For example, sex differences in adult spatial reasoning were found in some cross-cultural samples (Vashro & Cashdan, 2015) but not in others (Trumble, Gaulin, Dunbar, Kaplan, & Gurven, 2016). As another example, sex differences in children’s spatial reasoning (Hyde, 1981; Voyer, Voyer, & Bryden, 1995) were not reproduced in low-income children over two different spatial tasks, repeated four times over two years, in a large sample of children in Chicago (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). The authors of the review article cited above (Henrich et al., 2010b) reasoned that if sex differences could not be generalised across income, then conclusions about the development of sex differences in behaviour may not apply across cultures, nor to humans in general.

Consequently, developmental psychology has recently increased its attention to cross-cultural replication studies. Researchers have tested cross-cultural patterns in many developmental domains, including learning (Boyette, 2016b; Caldwell & Whiten, 2002), imitation (e.g., Berl & Hewlett, 2015; Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009), innovation (e.g., Nielsen, Tomaselli, Mushin, & Whiten, 2014), and belief (e.g., Richert, Boyatzis, & King, 2017). However, cross-cultural research on sex and gender development has, so far, focused primarily on the intersectionality of race and gender identities within minority world contexts (e.g., Leaper, 2015; Rogers, Scott, & Way, 2015; Shields, 2008). These cross-cultural replications and extensions have not yet included children’s sex-typed colour and toy preferences in the majority world.

Gender, Toys, and Culture

Definition of Culture. Culture in this thesis is considered as the context in which children develop toy preferences. It includes, but is not limited to, social structures, symbols, and adult gender roles. Previous toy preference research, reviewed below, has considered culture in this broad sense. The focus of the thesis is on removing toys from their cultural context, to understand more about the relationships between cultural explanations and functional explanations for children’s sex-typed toy preferences (see Figure 1.1).

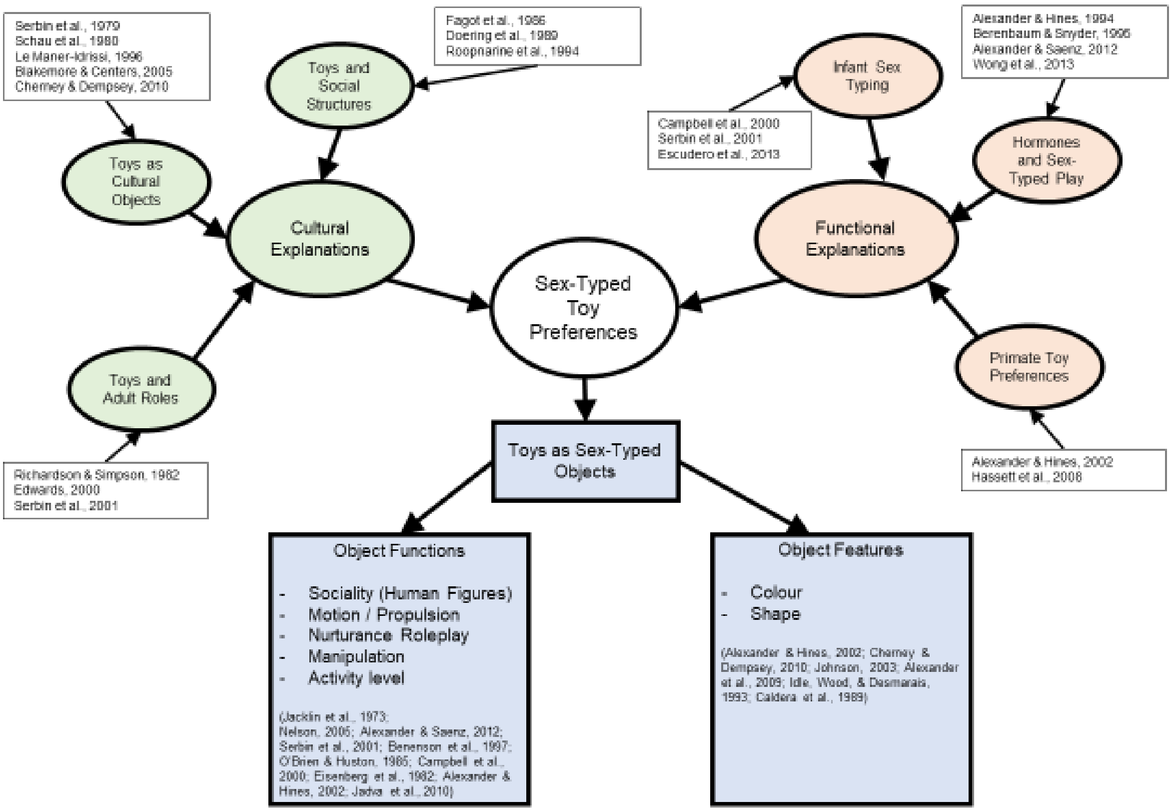


Fig. 1.1. Overview of cultural and functional explanations for sex-typed toy preferences.

Cultural Explanations. Previous work on toy preferences has identified culture as the context in which children's understanding of toys develops, including social structures, symbols, and adult gender roles. Consequently, authors often refer to the gendered cultural meanings of toys. For example, an analysis of children's letters to Santa Claus concluded that the physical characteristics of toys were secondary to their symbolic meaning as representations of gendered social structures (Richardson & Simpson, 1982). Children's selection of gender-typed toys may therefore be a product of their conformity to larger social structures (Fagot & Patterson, 1969). Consistent with this view, early studies of toy preferences sometimes explained that the toys' gender categories were culturally defined (e.g., Schau, Kahn, Diepold, & Cherry, 1980; Serbin et al., 1979). Where authors discussed features of the toys, they sometimes specified that these features derived their gender-typed values from social structures around children (e.g., Eisenberg, Murray, & Hite, 1982), or from their relevance to adult gender roles (e.g., Edwards, 2000; Richardson & Simpson, 1982; Serbin, Poulin-Dubois, Colburne, Sen, & Eichstedt, 2001). More recently, some toy preference studies have aimed to categorise toys according to cultural gender expectations in the minority world (e.g. Blakemore & Centers, 2005; Cherney & Dempsey, 2010).

Functional Explanations. While some perspectives see toys as cultural objects, other perspectives see toys as functional objects, with features that may be differentially attractive to boys and girls for reasons other than, or in addition to, culture. Several lines of research support the functional view. First, infants show some sex differences in toy preferences at an earlier age than would be expected if toy preferences only resulted from cultural conditioning (Campbell, Shirley, Heywood, & Crook, 2000; Serbin et al., 2001). Sex differences in infants' toy preferences may imply a causal mechanism that is inborn, and thus not culturally defined, according to some authors (Campbell et al., 2000; Jadva, Hines, & Golombok, 2010). This view is challenged, however, by other authors, who did not find sex-related toy preferences in 4- and 5-month-old babies (Escudero et al., 2013).

Further evidence for the functional view is found in research examining the role of early androgen exposure in sex differences in children's toy preferences. As an illustration, Alexander and colleagues interpreted their study of hormone-behaviour association in infants to imply that "prenatal androgen levels may be important for the early organization of preferences for object features that characterize male-preferred and female-preferred toys" (Alexander, Wilcox, & Farmer, 2009). However, proponents of cultural explanations (e.g., Fagot & Patterson, 1969; Serbin et al., 2001; Cherney & Dempsey, 2010), might argue that, since toys are cultural objects, androgens could be acting on sex-typed development more generally, with toys simply representing a culturally appropriate manifestation of sex-typed behaviour.

Alternatively, proponents of the functional view might point to toy preference research in primates, who do not have a cultural context for sex-typed toys, but who appear to show similar sex-typed toy preferences to those seen in children (e.g., Alexander et al., 2009; Hassett et al., 2008). Sex-typed toy preferences, paralleling those of humans, have been found in two species of non-human primates (Alexander & Hines, 2002; Hassett, Siebert, & Wallen, 2008). Since non-human primates do not have culturally gendered roles for toys, these studies have been cited as evidence that there is some inborn, non-cultural, component to children's toy preferences (Alexander & Hines, 2002; Alexander, Wilcox, & Farmer, 2009; Hassett et al., 2008; Jadva et al., 2010; Wong, Pasterski, Hindmarsh, Geffner, & Hines, 2013).

These lines of research do not exclude a role for culture in children's sex-typed toy preferences. However, they are often interpreted to mean that there is some inborn, non-cultural component to toy preferences (e.g. Alexander, 2003; Alexander & Hines, 2002;

Alexander & Saenz, 2012; Cherney & Dempsey, 2010; Jadva et al., 2010). The specifics of this inborn component are not well-developed, but several commenters point to functional features of toys such as their colour, or the type of play afforded by the toy. The next section reviews some of these functional features.

Toys as Sex-Typed Objects. If sex differences in children's toy preferences are not based solely on culture, then on what might they be based? Some authors have identified colour, specifically, the use of pink as a signal of female gender (Alexander, 2003; Del Giudice, 2017; LoBue & DeLoache, 2011; Weisgram, Fulcher, & Dinella, 2014; Wong & Hines, 2015). Others have pointed to the kinds of activities afforded by the toy (Campbell et al., 2000; Moller & Serbin, 1996; Serbin et al., 2001), or to specific toy features such as wheels (Campbell et al., 2000), faces (Campbell et al., 2000; Escudero et al., 2013; Jacklin, Maccoby, & Dick, 1973; Nelson, 2005), social stimuli (Alexander & Saenz, 2012; Jacklin et al., 1973), object motion or propulsion (Alexander, 2003; Benenson, Liroff, Pascal, & Cioppa, 1997; Benenson, Tennyson, & Wrangham, 2011; Cherney & Dempsey, 2010; O'Brien & Huston, 1985; Serbin et al., 2001), activity level (Alexander & Hines, 2002; Alexander & Saenz, 2012; Berenbaum & Hines, 1992), or nurturance (Alexander & Hines, 2002; Cherney & Dempsey, 2010; Serbin et al., 2001). Recent literature has begun to test children's preferences for some of these functions and features directly, particularly colour (Weisgram et al., 2014), propulsion (Dinella, Weisgram, & Fulcher, 2016), and faces or wheels (Escudero et al., 2013).

One approach to beginning to separate possible influences of cultural context and functional features of toys, might be to remove the toys from their cultural context. Do sex differences in children's toy preferences replicate in different cultures, particularly in those that are not minority world cultures or where the toys are unfamiliar? To this end, the current study assessed children's preferences for sex-typed toys, as well as for colours and toy affordances, in a minority world culture, and in three majority world cultures. The study aims, rationale, and general methods are presented in the following chapter. Next, I discuss how statistical methods, based on systems theory, will be used to integrate the results of the current study.

Systems Theory and the Current Study

Systems Theory. In recent decades, scientists have increasingly viewed sex and gender development as a complex and dynamic, but mathematically describable, phenomenon (e.g., Fausto-Sterling, Garcia Coll, & Lamarre, 2012; Hines, 2013; Martin & Ruble, 2010). This change has happened in the context of a wider adoption of systems perspectives in child development (van Geert, 2011), and a broad shift towards systems-based developmental science (Overton, 2007; Overton & Lerner, 2014); see Figure 1.2. The thesis refers to this perspective broadly as *systems theory*. Systems theory has primarily been used as a conceptual framework, but recent approaches have started to apply systems theory to the statistical analysis of empirical research in the behavioural sciences (e.g., Tenenbaum, Griffiths, & Kemp, 2006).

One of the key contributions of systems theory to developmental psychology is that it introduces new statistical methodology (van Geert, 2011). New statistical methods, based on systems theory, have allowed developmental psychologists to address important theoretical issues such as non-linear learning (Gopnik & Tenenbaum, 2007), individual developmental trajectories (Oostenbroek et al., 2016), and the spatial movement of play dyads (DiDonato et al., 2012). Thus, statistical methods based on complex systems have contributed new theoretical insights to developmental psychology.

Further, in the context of sex-typed behaviour, systems theory may offer a solution to the multidimensionality and variability of sex and gender in development. Investigators have acknowledged that social, cognitive, and biological perspectives each cover part of a wider, multidimensional system. For example, researchers with a social and cognitive focus acknowledge the role of biology in providing a foundation for experience (e.g., Martin & Ruble, 2010), and researchers with a biological focus acknowledge the importance of social environments and cognition (e.g., Hines, 2013; Pasterski et al., 2005). Increasingly, researchers use systems theory to explain complexity in gender research results (Fausto-Sterling et al., 2012; Hines, 2013; Martin & Ruble, 2010). Thus, there is an increasing interest in systems explanations for gender development, including toy preferences. To date, however, research on sex-typed behaviour contains very few examples of statistical methods based on complex systems (DiDonato, England, Martin, & Amazeen, 2013).

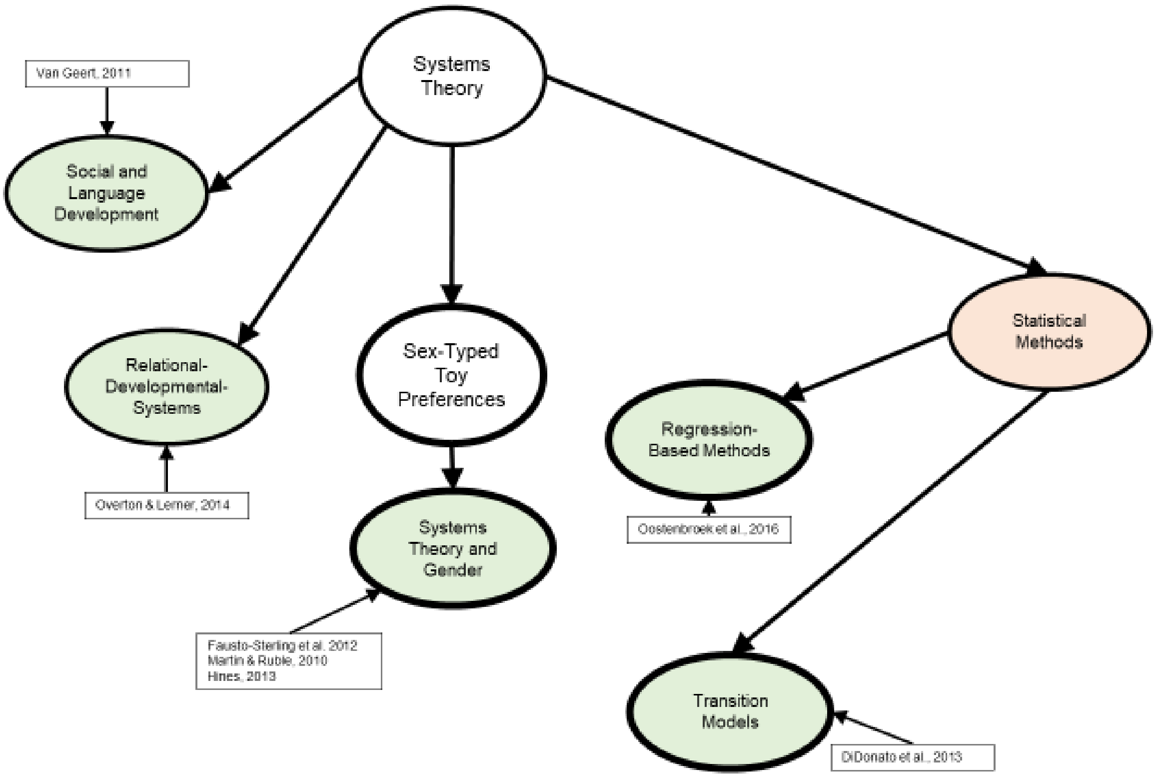


Fig. 1.2. Overview of sex-typed toy preferences in the context of prior work on systems theory.

The data collected in this thesis are well-suited to statistical methods based on systems theory. A system is *complex* if it is made up of multiple inter-related parts, and *dynamic* if it changes over time, and can be described mathematically using simple rules (van Geert, 2011). In comparison, the toy preference dataset collected in this thesis includes multiple inter-related variables (child gender, toy sex-type, cultural context, and toy affordance); and the play session is video-recorded continuously over time (see Chapter 2). The thesis dataset is therefore well-suited to statistical methods based on complex and dynamic systems.

The increasing interest in systems theory for gender development research, and the contributions of statistical methods based on systems theory to other areas of developmental psychology, together indicate that it would be useful to demonstrate how statistical methods could be used for gender research. Additionally, the toy preference dataset collected as part of this thesis is well-suited to statistical methods based on systems theory. Therefore, this thesis includes two demonstrations of how statistical methods based on systems theory can be applied to the toy preference data. The following chapter explains the thesis aims, rationale, and general methods.

Chapter 2
Overview

The research presented in this thesis was gathered over a series of field studies. The primary objective of these field studies was to test whether children's preferences for sex-typed toys would replicate cross-culturally. As secondary objectives, analyses of children's colour preferences and affordance preferences were also added. A methods contribution was also made in the context of systems theory.

In this overview chapter, I detail the rationale for these objectives, and give a description of the scope of the thesis. I also introduce the field sites, describe the general methods and their strengths and weaknesses, and outline the structure of the thesis document.

Aims and Rationale

The primary objective of this thesis was to test whether gender differences in children's preferences for sex-typed toys would replicate in cultures where the toys were unfamiliar. Sex differences in children's toy preferences are well-established and reliable. But are these sex differences due to the gendered cultural meanings of the toys, or to features of the toys that might appeal differently to boys and girls everywhere, such as colour or play affordance? Taking sex-typed toys out of their cultural context might help to answer this question.

To this end, the current study assessed children's preferences for sex-typed toys, as well as for colours and affordances, in one minority world, industrialised, English-speaking culture and in three majority world, non-industrialised, non-English-speaking cultures:

- Brisbane, Australia – a large, English-speaking, industrialised city;
- Shipibo villages, Peru – small, indigenous villages on the Ucayali River in the Amazon Basin;
- Lenakel town, Vanuatu – the largest town on Tanna Island, in the country of Vanuatu in the South Pacific; and,

- Navhal kastom villages, Vanuatu – small, indigenous villages in the mountains of Tanna Island.

Children in each study site completed the same two tasks. One task was a toy preference test, in which children were asked to play alone with a set of toys for five minutes. The other task was a colour preference test, in which children were asked to indicate their preferred colour by pointing to a laminated square. Results from the toy preference test were also analysed to explore children's preferences for play affordances.

Given the two important topics considered in this thesis – gender and culture – there are a great number of possible alternative studies that could have been conducted. In addition, the two secondary foci of the thesis – colour and affordance – might have been studied differently if they were the primary focus. However, toy preference was selected as the main focus of the study because toy preference is one of the largest and most reliably demonstrated behavioural sex differences in present research (Hines, 2005), and because of the availability of methods for measuring toy preference that could be adapted for high-variability field conditions. The next sections discuss some issues with adapting a toy preference study for cross-cultural research, and how these issues were addressed on the basis of background interviews and an initial pilot study.

Study Scope

Clearly, attempting a cross-cultural replication of a well-studied psychological effect raises complex and difficult issues. The process of planning this research considered such issues as ecological validity, use of mass-produced toys, and variability in field conditions. These issues were resolved based on information from background interviews and a pilot study, as described in the following section.

Basis for the Study Design

The study design was refined based on discussions with colleagues and with the thesis supervisor. Additional design decisions were based on background interviews with contact people who knew the cultural settings well, and on an exploratory pilot study.

Background interviews. In the initial scoping stages of the research, adult contacts in the field sites were asked to give some information as background to the study. These interviews informed the study design. Key questions were: what do men and women do during the day? What do children do during the day? Are there any special activities that are for men only or for women only? Are there any special objects that are for boys only or for girls only? Do boys and girls do different things? What do children do when they play? Do parents and children play together? Interviews were informal and free-form, but some were recorded for later reference. The interviews were intended to only provide background information to inform the results of the play sessions, and to determine whether the study procedures would be culturally appropriate. A full ethnographic study, or qualitative interviews on the development of sex and gender, were not part of the scope of this thesis.

Pilot study. The study methods were piloted on a sample of children (N=10) in the Gender Development Research Centre at the University of Cambridge. The purpose of the pilot study was to identify any logistical issues with the procedure. As such, some changes to the design were made based on the pilot study.

Design Considerations

Designing a cross-cultural behavioural psychology study requires making some difficult design decisions. Key decisions are presented and discussed below, including: balancing ecological and internal validity; the decision to use mass-produced toys and not local toys; and variability in field conditions. All of these decisions were made in consultation with local experts well in advance of data collection, to try to ensure the optimum outcome for the research, and to try to limit any negative impact on the study population.

Ecological vs. internal validity. In constructing the work for this thesis, it became clear that there would be a trade-off between ecological validity and internal validity. The primary aim of this thesis was a cross-cultural replication study, so design decisions prioritised a procedure that would be comparable across each of the field sites.

The decision to prioritise replication resulted in some limitations. In particular, because of the differences in language across the cultures studied, the research had to use a non-verbal assessment of children's toy preferences. I thus chose to watch children's behaviour in a play area, where they were given the opportunity to interact with several toys, but where verbal prompts or replies were minimized. The outcome measure is the amount of time spent with each toy or type of toy (e.g., boy-type toys, girl-type toys). Although this approach has been used in much prior research on sex differences in toy preferences, solo play is unusual for children in the majority world cultures included in the current study (Shipibo villages, Peru; Lenakel town, Vanuatu; and Navhal kastom villages, Vanuatu). However, children in minority world cultures also spend a large amount of time playing with others, rather than on their own. Also, it is not clear how this limitation should affect any sex difference in children's observed toy preference.

Because the initial background interviews suggested that the toy preference test and the novelty of the toys might feel strange to some children, I anticipated that some children might not engage with the toys. Partly for this reason, I concluded that I would need large samples in this study, larger than power analyses of prior, minority world studies suggested. However, most majority world children did engage with the toys during the play session and their expressions and behaviour as seen on the video recordings of the play sessions indicated that they enjoyed the experience.

Use of mass-produced toys. Another key decision in this research was whether to use the same toys in each study site, or whether to choose different gendered objects in each location. The aim was to see whether children's sex-typed toy preferences related to the cultural role of the toys as sex-typed objects, or to features of the toys themselves. This question could only be answered by using the same toys – with the same features – in each location. These toys had to be the ones used most commonly in previous research in minority world cultures, so that the study findings could be meaningfully compared to previous research. I therefore decided to use commercially available, mass-produced toys that were sex-typed in the local minority world culture (Cambridge, UK).

Early stages of the study design considered an additional step: identifying locally relevant gendered objects (preferably toys) in each cultural group, and then transporting these objects to all other groups, along with the toys from the minority world. Children would then have completed a second toy preference test, using the objects relevant to specific majority world cultures. However, following a literature review, it was clear that this procedure would itself require extensive research to justify the selection of culturally relevant gendered objects. Additionally, based on background interviews, a second toy preference test could have introduced issues from children becoming fatigued and may have decreased community engagement due to the time commitment for participants. As such, this path of research was considered out of scope for this thesis.

Variability in field conditions. Large variability in field conditions meant that tasks could not be standardised to the extent that they would be in a lab study. In consultation with local contacts, it was decided that a shorter procedure was preferable to limit the time requirement for the young participants. The tasks were therefore shorter than typical laboratory tests; for example, the colour preference task included no repetition of questions, and the toy preference task was only 5 minutes long. These changes may have increased the statistical variability of the data and thereby decreased statistical power. To account for the increased statistical variability, the sample size was increased to a larger number than was required according to the power analysis (see Chapters 3 and 4 for power analyses) or seen in many lab-based studies.

Each of these design decisions had consequences for the interpretation of the study results. These consequences are discussed in detail in the final chapter of the thesis, under “Limitations and Caveats”.

Statistical Methods Based on Systems Theory

This thesis includes two chapters focused on statistical methods based on systems theory (Chapters 6 and 7). New statistical methods, based on systems theory, have offered new insights in other areas of developmental psychology (Gopnik & Tenenbaum, 2007; van Geert, 2011). However, there are few demonstrations of statistical methods, based on complex systems, applied to data on gender development, specifically toy preferences (DiDonato, England, Martin, & Amazeen, 2013). This is despite an increasing interest in systems thinking by gender theorists (e.g., Fausto-Sterling, Garcia Coll, & Lamarre, 2012; Hines, 2013; Martin & Ruble, 2010). Thus, a demonstration of these methods may be useful for gender researchers with an interest in systems theory, and these new statistical methods may provide additional insights into the results of the thesis research. One of the objectives of this thesis, then, was to demonstrate how statistical methods, based on systems theory, might be applied to the toy preference data. This methodological objective is discussed further below, and in Chapters 6 and 7.

General Methods

Selection of Study Sites

The aim of the thesis was to test whether sex differences in children’s toy preferences replicated in cultural contexts where the toys were unfamiliar. Therefore, the most important criterion for selecting cultures to study was that these cultural settings contain as little context as possible for the toys. Ideally, the cultures would have no access to mass media (that might contain advertisements about the toys), mass communications (that might communicate the opinions of minority world members about the toys), or mass-produced leisure goods (the toys themselves).

Nevertheless, children in all cultures might recognise some of the toys as related to their real-life analogues, particularly for toys that represent adult roles (e.g., the baby doll and the sword), or for miniature adult figures (e.g., the Barbie doll and the Hulk action figure). Apart from these toys, however, it would be ideal for children to not have observed adults interacting with the real-life objects that the toys represent. Therefore, the ideal setting for this study would be one that had no vehicles and no mass-produced clothes, in addition to no mass media, mass communication, or access to manufactured toys.

Practical considerations limited the selection of cultural contexts for this study. Communities had to be willing to receive researchers, and accessible within a reasonable time frame and budget. The final communities selected for study were: a Shipibo community separated from major towns by the Ucayali River in Peru; a kastom community separated from major towns by mountains on Tanna Island, Vanuatu; and a larger coastal town on Tanna Island. A large city, Brisbane, in Australia provided a comparison sample from the minority world.

Selected Field Sites

This section provides background information on each field site. Figure 2.1 gives an overview of the geographic distribution of the field sites.

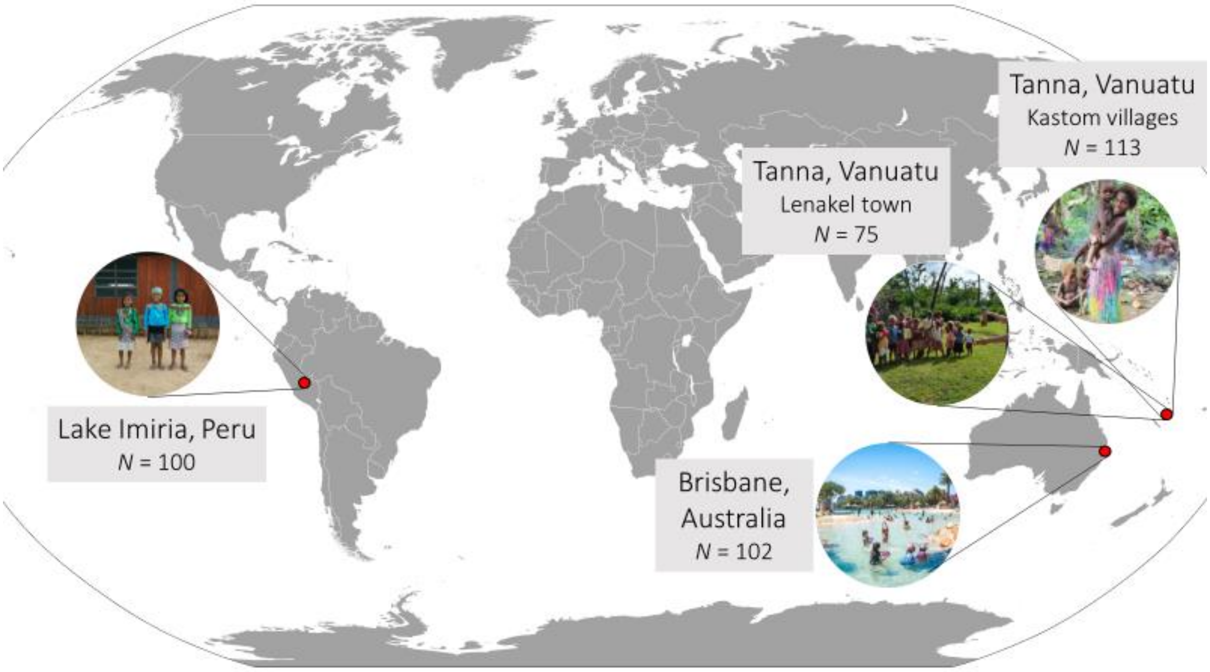


Fig. 2.1. Geographic distribution of the study sites.

Brisbane City, Australia

Brisbane is a large (population approx. 2.35 million), English speaking, developed city with compulsory free education for children aged 5 years and older. Brisbane is typical of a minority world culture usually studied in academic research on sex and gender development. Children have access to mass-produced toys, mass media, and mass communication. Through these channels, children in Brisbane might receive information about the cultural gender context of toys like those in the current study.

Children in Brisbane were recruited as a convenience sample in a public, free-entry museum. Consent was given verbally and electronically by parents, and children were given a wristband for participating.

Shipibo Villages in the Ucayali River, Peru

Caimito, Buenos Aires, Nuevo Loreto, and Nueva Yarina are four villages of indigenous Shipibo people, situated in the Lake Imiria region of the Ucayali River in the Amazon rainforest, Peru. Villages in this region have sporadic contact with outside influences because the dense jungle prohibits access via any means but the river, and the river is too flooded to pass in the wet season and contains impassable sand banks in the dry season. The villages have no central electricity or communications infrastructure, but they do typically have a generator that supplies electricity to the town when there is fuel available, and a single landline phone per village to allow residents to receive calls from the major cities. Caimito village also has mobile reception, although as of our visit the mobile tower had only been set up in the last few months and almost no residents owned a mobile phone. There is no access to Internet.

Children attend state-regulated formal education. School is taught in Shipibo language and in Spanish. Villagers speak Shipibo and some adults also speak Spanish. Villagers live on small-scale agricultural and fishing subsistence, but each village contains a small number of people who are paid a government wage for teaching, tax collecting, or other professions. Peruvian government and aid agencies occasionally visit Caimito village, and each village contains a small shop selling some mass-produced food and soap items, but no toy or leisure items. Adults may travel to nearby cities, where mass-produced goods are available including clothing and toys. Travel to the city is exclusively by river and is expensive and uncommon. Adults and children wear clothes from the minority world, typically sourced second hand. Some of these clothes, especially clothes for children, depict media characters or children's mass-produced toys.

Shipibo culture is patrilineal, and roles are gender differentiated (for example, only men may fish in the river). Village chiefs and administrators are not exclusively men in theory, but in practice, during this research, only men were observed in these roles.

Children in the Shipibo villages were recruited through a formal process. The research project was presented to the village in a public forum, where anyone could ask questions about the project or about the researchers. In smaller villages (Buenos Aires, Nueva Yarina, and Nuevo Loreto), all children who fell within the required age range were recruited. In the larger village (Caimito), children were selected for the study via a process suggested and overseen by the village leaders. Mothers nominated their children for participation in a separate village meeting. Children were given a small gift for participating (some books and pencils for school).

Tanna, Vanuatu

The South Pacific island of Tanna, in the nation of Vanuatu, provides a unique testing ground for cross-cultural research. Most adults on Tanna live primarily through subsistence farming (especially yam, taro, and plantain), with some cash crops (coffee, sandalwood, vanilla) sold to overseas buyers when cash is needed (Lindstrom, 2008). Male and female roles are well-defined and separate, and several important cultural rituals are male-only, such as the drinking of kava and the circumcision ceremony (Lindstrom, 2008). The availability of gender-typed manufactured goods, such as toys, is limited by the difficulty involved in physically transporting these items to the island. Foreign goods other than food and clothing are only found in the area immediately surrounding the capital city of Lenakel and are not familiar to residents of the remote inland mountain villages. The trade language is Bislama, but most children are more familiar with local languages that vary regionally. The most common local language in our testing region was Navhal.

Participants on Tanna were recruited from two separate locations: kastom villages in the mountains, and a school in Lenakel, the largest town on the island.

Navhal-speaking kastom villages. Ikunala and Yakel villages are kastom villages in the remote mountain regions of Tanna Island, Vanuatu. Both villages are part of the Navhal language group and live according to kastom tradition. Kastom, as practised in these villages, limits contact with modern inventions; the ideal lifestyle requires no modern facilities, and villagers wear skirts and penis sheaths made of grass, live in grass houses, and have no access to electricity, foreign trade, manufactured goods, mass media, or mass communication. Villagers are of Melanesian indigenous descent, and adults and children speak the indigenous Navhal language, although some adults also speak Bislama. Children do not receive a formal education and do not typically travel to large towns or cities, although adults may visit large towns to take produce to market. Travel is typically by foot, but vehicles occasionally visit the village when the road is dry.

Kastom villages subsist on small-scale agriculture, with inter-village trade of valuable goods such as woven mats and baskets. Money is rarely used, and having money is not seen as culturally desirable in village members. Ikunala village is not typically accessible to outside visitors, but Yakel village allows access from paying visitors including tourists and film crews. The kastom culture is patrilocal, patrilineal, and monogamous. Women may own land and livestock, but positions of power in the village (chief, medicine man, spiritual leader) are always held by men.

Research permission was given by the Vanuatu and Tanna Cultural Centres. In each village, two translators were recruited who each spoke Navhal, Bislama, and English, and who knew children of the village personally. According to translators' advice, we provided each village with appropriate gifts to thank them for their participation: Ikunala village with trade gifts (coconuts, tinned food, rice, and kava, approximately 70GBP worth), and Yakel village with money (10 000 vatu, equivalent in value to the gift for Ikunala village). Children were recruited through village chiefs and heads of families. All children in each village who appeared to be in the target age group, and who were otherwise available and eligible to participate, were included in our study sample.

Lenakel Town. The second set of participants on Tanna was recruited from a school in Lenakel. Lenakel is the largest town on Tanna Island in the Republic of Vanuatu. Children in Lenakel were recruited in the public primary school, Lenakel Harbourview. The school is attended by children who live in the town and in villages nearby. Children from different villages may speak different native languages and have varying access to electricity and communications infrastructure at home. Typical households in Lenakel may have some electricity and mobile phone access, but access to mass media, mass communication, or the Internet is rare. Individuals living in Lenakel can access shops that sell a variety of mass-produced food items, household goods, stationary, and clothes. Mass-produced leisure items such as toys, however, tend to not be prioritised for transport to the island over essentials like food, household items, and clothing. For this reason, shops in Lenakel do not typically sell mass-produced toys or games. The most toy-like items observed in shops over the duration of data collection were footballs, volleyballs, and coloured pencils. Travel to islands with larger towns and cities is expensive, and not accessible to most children. Some adults, however, travel to Port Vila or to cities in New Zealand and Australia for seasonal work. These adults may return with gifts purchased in the minority world, although from personal observation such gifts tend to be functional (e.g., kitchen items) rather than purely recreational (e.g., toys).

School instruction takes place in Bislama - a creole of French, English, and indigenous languages, used as a trade language throughout Vanuatu. Teaching materials are distributed by the Vanuatu government to schools, in the form of some textbooks and class curricula. However, teachers' access to standardised teaching materials is limited, and classrooms do not have electricity or connections to mass media, mass communication, or mass-produced goods.

Gender role conventions in Lenakel are a mix of traditional kastom culture (reviewed below) and colonial cultures. As a result, the social and legal status of men and women vary, but in general men are more likely to have power in social roles (Douglas, 2002; Vanuatu Women's Centre, 2014). All of the teachers at the school were women, and the Headmaster was a man.

Children at Lenakel Harbourview were recruited through permission from the Headmaster and then from teachers of each class. Children were given a pencil and eraser, purchased in local shops, as a gift for participating. A small donation was given to the school, in thanks for allowing the research team to use one of their cyclone recovery tents as a setting for the study. Information about child age was taken from class lists, and the accuracy of this information depended on the teacher's records. In some cases, exact birth dates were available, but in others, no age information had been recorded by teachers. For this reason, age information is not complete in the Lenakel sample.

Lenakel was added as a location after consultation with local contacts on Tanna, who reported that the kastom villages represented a specific subculture and not the island as a whole. Due to its late addition, data collection in Lenakel took place over a shorter timeframe than the other locations. There are fewer participants in Lenakel than in the other locations, and participants in Lenakel completed only the toy preference task, and not the colour preference task.

Recruitment and Participant Information

In each site, participants were children between the ages of 4 and 11 years. Recruitment in Vanuatu and Peru was via a three-stage system. First, I identified villages through a local contact (Peru: Mr Ronel Garcia, Shipibo tribe; Vanuatu: Mr Jacob Kapere, Tafea Cultural Centre) and asked them for preliminary permission to visit. Next, I asked village chiefs and elders for formal permission to visit, recruit, and conduct the study in their villages. Finally, when the study team arrived in the village, we held a village information session where information about the study was read to interested adults and children, and they had the opportunity to ask questions and to register their interest in participating. In Australia, participants were approached individually inside a state museum, where a university research station was already set up and regularly recruited local children.

Determining age in the Navhal kastom villages is difficult, because these groups do not typically keep a record of birth dates or ages, and the indigenous language has no numbers greater than five. For this reason, age estimates in the kastom villages are approximate, and based on best estimates made by an English speaker from a neighbouring village, who was acquainted with the village children. Age estimates in Lenakel are also approximate: we asked children their ages and checked school records, but children were not confident in reporting their own ages, and school records were incomplete.

All children who asked to participate were allowed to, but not all of these children were in the correct age range. Additionally, some children did not play with the toys for very long, particularly in the Vanuatu samples. Therefore, some children were removed from the sample before data analysis. The exact number of children who were excluded is given separately for the colour preference test in Chapter 3, and for the toy preference test in Chapter 4.

Children participated first in a toy preference task, and then in a colour preference task, as described below. Priority was given to the toy preference task if there was limited time available, so fewer children participated in the colour preference task than participated in the toy preference task. No children in Lenakel participated in the colour preference task. Details of the final sample are given in Table 2.1.

Location	**Number of children who completed the toy preference task**	**Number of children who completed the colour preference task**	**Average age (years)**
Brisbane, Australia	102 (52.50% male)	50 (60.00% male)	5 years 11 months
Shipibo villages, Peru	100 (45.74% male)	100 (45.00% male)	6 years 11 months
Lenakel, Vanuatu	75 (50.94% male)	0	6 years 0 months
Navhal villages, Vanuatu	113 (61.25% male)	51 (60.78% male)	6 years 2 months
Total N	**390**	**201**	

Table 2.1. Participant information.

Study Team

The study team in each location consisted of Ms Davis, one or two university collaborators, and one or more local people who were employed as research assistants and translators. In Vanuatu, the study team included two students from the Early Cognitive Development Centre at the University of Queensland (Karri Neldner and Rohan Kapitany), and four field workers supplied by the Tanna Cultural Centre (Robert, Selina, Mary, and Rachel) who translated and administered the research procedure under supervision from the thesis author. In Peru, the university collaborator was another member of the Gender Development Research Centre at the University of Cambridge (Ellen Robertson), who also assisted with translation from English to Spanish and in administering the toy and colour preference tests. Ronel Garcia provided further translation between Spanish and Shipibo, and he assisted in administering the toy and colour preference tests, under supervision from the thesis author.

Local contacts, and researchers experienced in each of the field sites, were asked to comment on the protocol. For the Vanuatu sites, ongoing consultation took place with Chief Jacob Kapere (Tafea Cultural Centre), Mr Joe Narua, Ms Beverline Mahana, and Teacher Annie Loughman for information about local children, and with Professor Mark Nielsen, who has experience running developmental psychology experiments in similar locations. For the Peru sites, ongoing consultation took place with Mr Ronel Garcia (Shipibo tribe), who was a contact with the indigenous villages, and with Ms Vanessa Hunter, who has experience as an external researcher in field locations around Peru.

Materials

Figure 2.2 gives an overview of the twelve toys used in the study and how they were divided into sex-typed toys and affordance pairs. Toys were selected based on a previous meta-analysis showing the most commonly used sex-typed toys in toy preference research (Davis & Hines, forthcoming). Toys with no or neutral gender type were also included, as a third option for children to play with, to avoid artificially inflating the size of the sex difference. The main analysis of toy preference focused on sex differences in children's preferences for toys grouped by sex type (boy-type toys, girl-type toys, and neutral toys; Chapter 4).

Data from the same toys was also used to examine children's preferences for toy affordances. For example, propulsive movement affordance could be provided by a toy racing car (a boy-type toy) or a pony carriage (a girl-type toy). Since some types of play may appeal to boys more than girls, or vice versa, the sex-typed toys were matched as far as possible on play affordance. A secondary analysis focused on sex differences in children's preferences for toys with different affordances (propulsion, figure play, dress-up, and roleplay affordances; Chapter 5). Only sex-typed toys were included in the analysis of toy affordances.

Practical considerations further shaped the selection of toys for the study. Toys had to be portable by hand across variable terrain, ruling out particularly heavy or delicate toys. To avoid confounding factors of toy appearance, toys were matched as closely as possible on size and distinguishing features, and toys were all purchased new and kept in as good condition over the course of the study. Two sets of identical toys were used to ensure that the toy set was never changed due to toys being lost or damaged. To avoid influence from recent media releases, toys were chosen, as far as possible, that did not reference mass marketed films or TV shows.

Despite these criteria, it was not possible to match toys perfectly, nor to entirely remove branding or reference to other media. The toys in the study were intended to represent toys that were commonly available in the immediate environment of actual children in the minority world. For this reason, toys were purchased in a large brick and mortar store in the United Kingdom. Toy selection was therefore limited to what was available in the store, and some options (such as a gender-neutral propulsion toy) were not available.

The final set of twelve toys was selected according to the above parameters. The set of boy-type toys (BTT set) comprised: a racing car (propulsive BTT), a superhero Hulk action figure with accessories (figure BTT), a pirate costume (dress-up BTT), and a plastic sword (role play BTT). The set of girl-type toys (GTT set) comprised: a pony and carriage (propulsive GTT), a female Barbie doll with accessories (figure GTT), a princess costume (dress-up GTT), and a baby doll (role play GTT). The set of neutral toys comprised: a dinosaur costume, a set of plastic African animals, pencils and paper, and a stuffed dog.

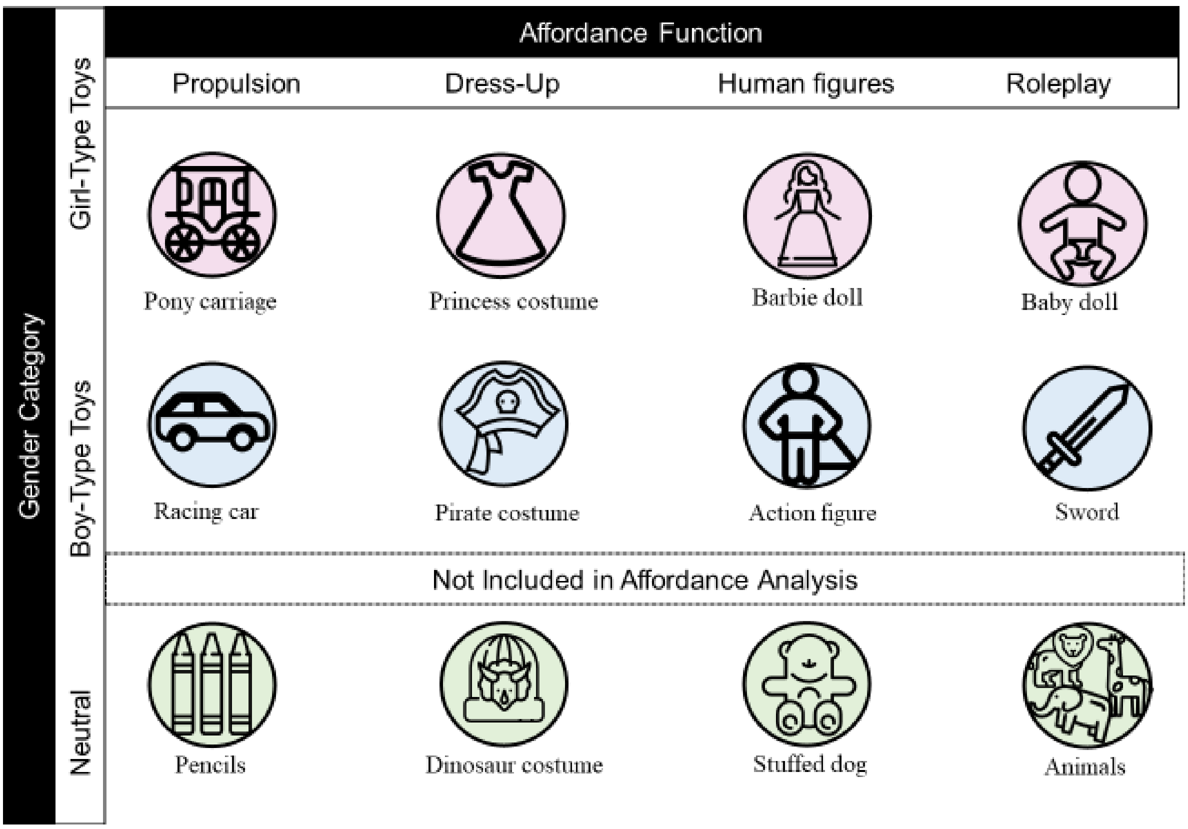


Figure 2.2. Representation of the 12 toys and how they were grouped by gender (Chapter 4), and by affordance (Chapter 5).

General Procedure

Selection of study method. Given the inherent difficulties of translating complex concepts (like *gender*) across cultural contexts, it can be challenging to show that translated measurements reflect the same underlying construct in different cultures (Cha, Kim, & Erlen, 2007; Sperber, 2004). Relying on measurements of observed behaviour (such as play time with toys) overcomes this issue, allowing for comparable measurements across cultures as there is no translation requirement. Two commonly used behavioural methods for studying toy preferences are free play and forced choice methods. In free play studies, children are observed playing with a set of toys, usually selected by the experimenter, over a set time frame (e.g., Berenbaum & Hines, 1992; Pasterski et al., 2005; Serbin, Connor, Burchardt, & Citron, 1979). In forced choice studies, children are asked to select their preferred toy from a small set, usually in a series of trials (e.g., Alexander & Hines, 1994; DeLucia, 1963). Free play methods have some advantages over forced choice methods, including (1) free play can be conducted without the experimenter present, reducing potential demand characteristics, and (2) in free play the child can play with multiple toys, or no toys, allowing for a greater range of possible outcomes than forced choice methods. For these reasons, the current study used a free play paradigm for behavioural measurement. The precise method is described in more detail in the Methods section of Chapter 4.

Overview of tasks. Children were asked to complete two tasks., a toy preference test and a colour preference test. The toy preference test was a free play session with twelve toys. Data from the toy preference test were analysed to answer research questions about children's preferences for sex-typed toys (Chapter 4), and toy affordances (Chapter 5). Data from the toy preference test was also used to demonstrate the application of two systems models (Chapters 6 and 7).

The colour preference test was a forced choice questionnaire using printed pictures. Data from the colour preference test were analysed to answer a research question about children's preferences for pink and blue (Chapter 3). Table 2.2 summarises the tasks, dependent variables, and thesis chapters.

Activity	Dependent Variable	Thesis Chapter
Colour preference test	Colour preference	3
Toy preference test	Sex-typed toy preference	4
	Toy affordance preference	5
	Method: Machine learning	6
	Method: Dynamic model	7

Table 2.2. Tasks, dependent variables, and thesis chapters.

Procedure. In each location, a large private space was constructed, with the materials for the play session inside. Parents were permitted to remain and watch if they wished, from outside the play space and unseen by children playing. The play space always had an open section so that the child could easily leave the area at any time if he or she wished to do so.

Children were introduced one-by-one into the play area and given instructions. Instructions were given in the local language, sometimes in several languages (e.g., Spanish and Shipibo; Bislama and Navhal) to ensure comprehension. Instructions were given by the translator/research assistant, who then checked that children understood the procedure and answered any questions. Also present was the author of this thesis, to check that the standard procedure was being followed and to monitor the cameras. The child was then left alone to play with the toys for 5 minutes, and the play session was recorded on video. More details about the toy preference test are given in Chapter 4.

After the toy preference test, the child was asked to complete a colour preference test. The translator/research assistant explained to the child that they would see a series of pages with two pictures on each page, and that they would have to point to the picture that they preferred. More details about the colour preference test are given in Chapter 3.

The same children completed both the toy preference task and the colour preference task. In the Navhal villages, and in Brisbane, priority was given to the toy preference task if children had limited time, so while all children completed the toy preference task, not all children completed the colour preference task. In Lenakel, children completed only the toy preference task, and not the colour preference task. In the Shipibo villages, all children completed both tasks.

Data from the toy preference test were analysed in several ways to answer different research questions, as described below.

Thesis Structure

Each chapter in the thesis presents a separate research question:

- Do sex differences in children's colour preferences replicate in different cultures? (Chapter 3);
- Do sex differences in children's toy preferences replicate in different cultures? (Chapter 4);
- Can toy affordances explain the sex differences in children's toy preferences? (Chapter 5);
- Can new statistical methods, based on systems theory, be applied to the toy play data? (Chapters 6 and 7).

Due to the depth of prior research around each of these questions, a separate literature review and discussion section is provided for each chapter. Results of all chapters are then integrated in a final Discussion chapter. Figure 2.3 illustrates how the overall study design and data collection corresponds with each thesis chapter.

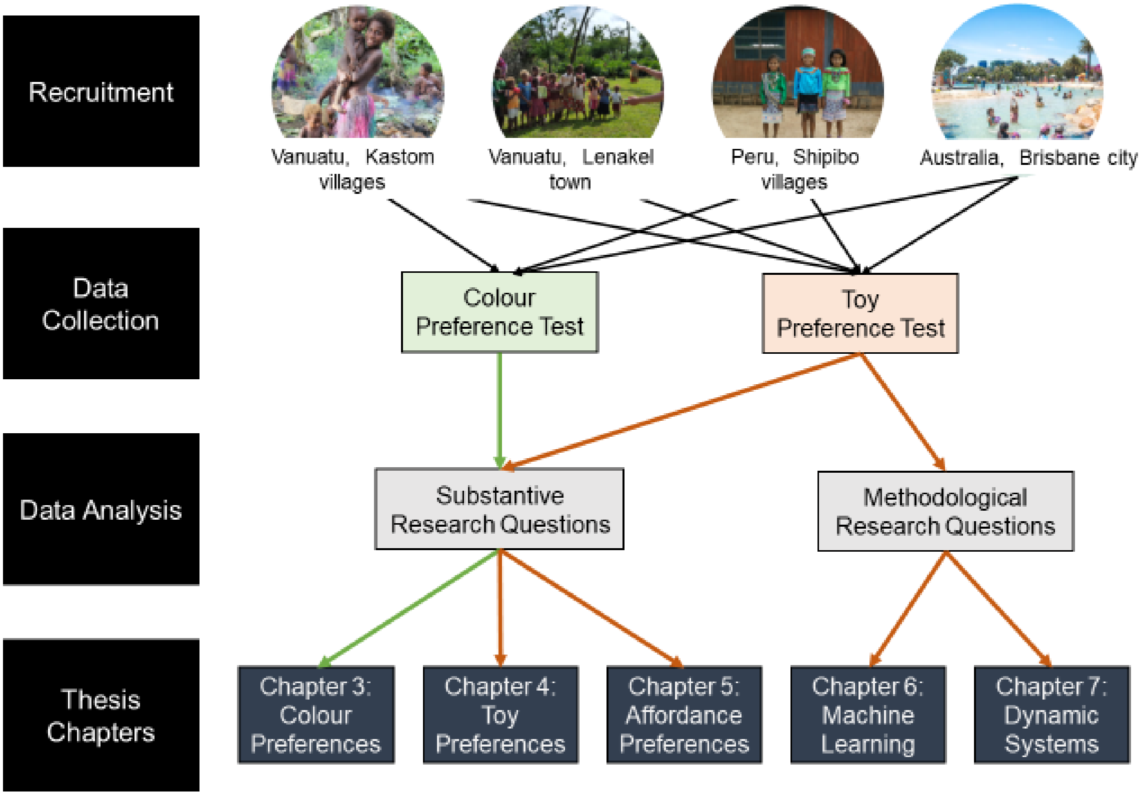


Fig 2.3. Overview of recruitment, data collection, research questions, and corresponding thesis chapters.

Summary of Thesis Contributions

The aim of this thesis is to examine the role of culture in children's sex-typed preferences for colours, toys, and affordances, and to demonstrate statistical methods based on systems theory. This general aim addresses two research gaps. First, whether children's preferences for sex-typed colours, toys, and affordances replicate across different cultures. Second, how systems theory can be statistically applied to empirical research on children's sex-typed toy preferences. These two gaps are addressed by two thesis objectives: a substantive objective, that describes the results of the cross-cultural study; and a methodological objective, that describes the statistical application of systems theory to the cross-cultural dataset. Each chapter of the thesis is presented as a stand-alone piece of research.

Substantive Chapters: 3, 4, and 5

The research presented in this thesis contributes to existing literature primarily by providing a replication study of children's toy preferences in cultures where the toys are unfamiliar. Additionally, functional explanations for sex-typed toy preferences are explored in the analyses of colour preferences and affordance preferences. A summary of substantive contributions is given in Figure 2.4. Focused reviews of relevant literature are also given in each of Chapters 3, 4, and 5.

The substantive objective of the thesis is to test whether children's sex-typed preferences for colours, toys, and affordances replicate across cultures. The cross-cultural study focuses on cultural contexts where the toys, that are sex-typed in minority world cultures, might be novel or unfamiliar. The cross-cultural study includes boys and girls from three majority world cultures: Shipibo villages in the Lake Imiria region of the Peruvian Amazon; Navhal-speaking *kastom* villages in the mountains of Tanna Island, Vanuatu; and a school in Vanuatu. It also includes boys and girls from a minority world culture, like the cultures where these preferences have been extensively studied: Brisbane, Australia. Children's preferences for colours are tested using a forced choice method, and children's preferences for sex-typed toys and for toy affordances are tested using a free play method. Cross-cultural results are presented separately for colour preferences, sex-typed toy preferences, and affordance preferences.

Chapter 3 presents the results of a cross-cultural study of children's colour preferences. It includes an empirical comparison of colour preferences in three cross-cultural samples, and a statistical model that compares the effects of sex, culture, hue, and saturation in children's colour preferences. The aim of Chapter 3 is to test whether sex differences in children's colour preferences replicate across cultures.

Chapter 4 presents the results of a cross-cultural study of children's toy preferences. It includes an empirical comparison of boys' and girls' preferences for sex-typed toys in four cross-cultural samples. The aim of Chapter 3 is to test whether sex differences in children's toy preferences replicate across cultures.

Chapter 5 presents a re-analysis of the toy preferences data from Chapter 3, with a focus on children's preferences for different toy affordances. It includes a discussion of the possible role of toy affordance as a functional explanation for gender differences in toy preferences, and an empirical comparison of boys' and girls' preferences for toy affordances in four cross-cultural samples. The aim of Chapter 4 is to explore sex differences in children's preferences for play affordances across cultures.

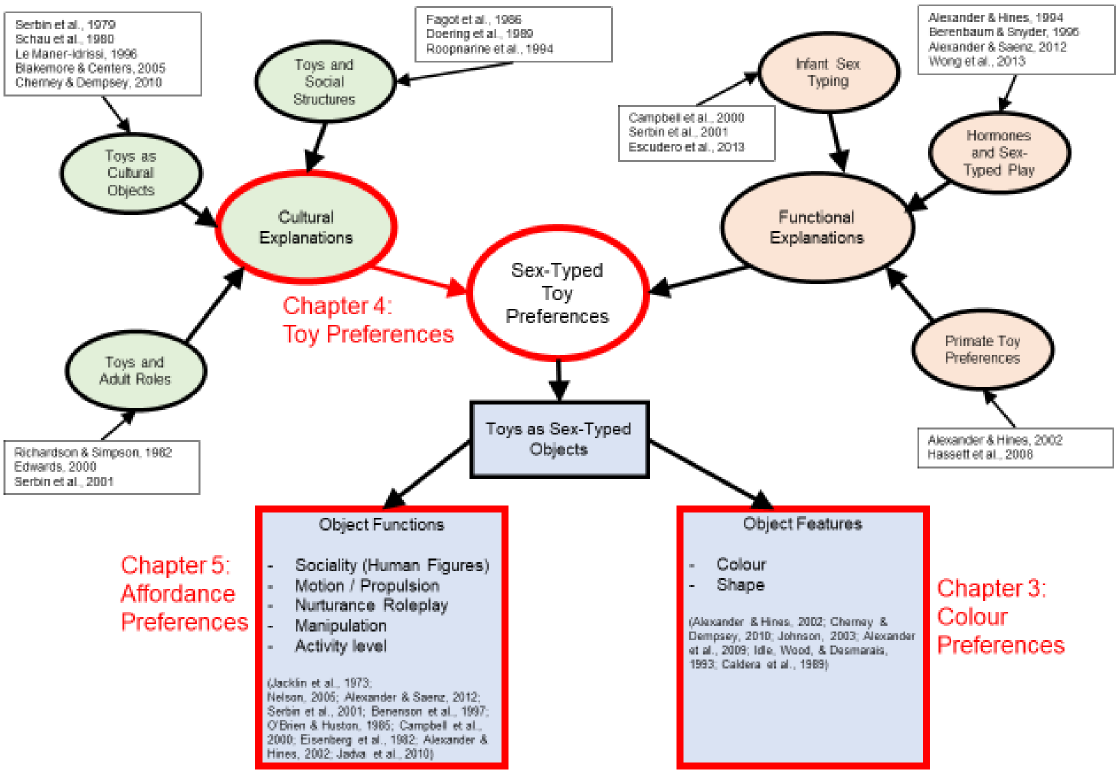


Fig. 2.4. Overview of thesis contributions to research on sex-typed toy preferences. Thesis contributions are shown in red. Each chapter contains a thorough review of the relevant subset of literature.

Methods Chapters: 6 and 7

Statistical methods, based on systems theory, could offer additional insights to research on sex-typed toy preferences. Chapters 6 and 7 provide two demonstrations of new statistical methods, based on systems theory, and apply these methods to the toy preference data. Chapter 6 demonstrates a machine learning method that is designed to deal with multiple complex interacting predictor variables. Chapter 7 demonstrates a multistate transition model that describes the transitions in children's toy choices dynamically over a play session. These contributions are presented in Figure 2.5 and are described in detail in Chapters 6 and 7.

The methodological objective of the thesis is to demonstrate how novel statistical methods, adapted from systems theory, can be applied to the toy preference data. The thesis also aims to use these novel methods to gain additional insights into the interactions between sex and culture, and their contributions to children's sex-typed toy preferences.

Chapter 6 presents a machine learning method for analysing the cross-cultural data. It includes a review of the advantages of complex systems in describing natural behaviour, an explanation of the boosted regression tree (BRT) and classification and regression tree (CART), and a demonstration of the BRT and CART applied to the cross-cultural sample. The aim of Chapter 6 is to demonstrate how machine learning models might offer insights into the factors that determine cultural variation in sex-typed behaviour.

Chapter 7 presents a theory-driven transitional method for analysing the cross-cultural data. It includes a review of dynamic systems theory, an explanation of the Markov model, and a demonstration of the Markov model applied to the cross-cultural dataset. The aim of Chapter 7 is to demonstrate how dynamic systems models might offer insights into the mechanisms underlying sex-typed behaviour.

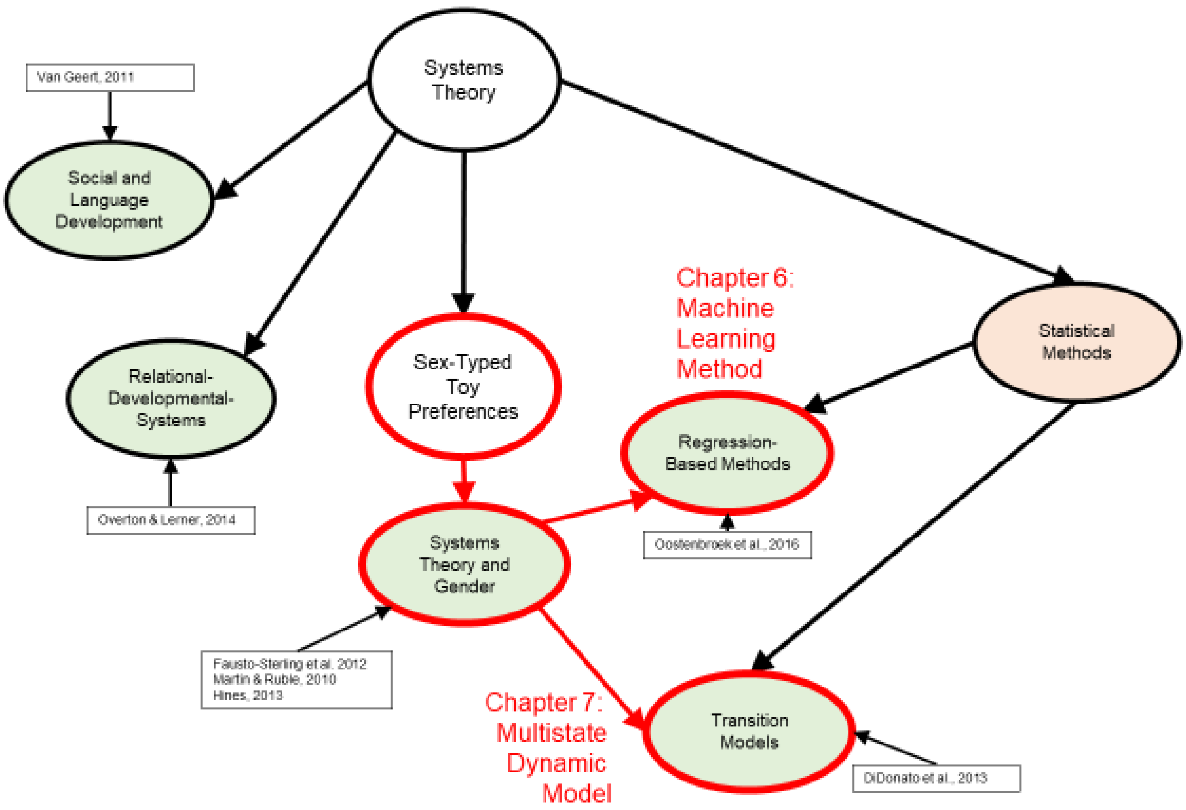


Fig. 2.5. Overview of thesis chapters in the context of current systems theory literature. Thesis contributions are shown in red. An explanation of the literature in the figure is given in each chapter.

Discussion Chapter: Chapter 8

The final Discussion chapter (Chapter 8) presents an overview of the conclusions from this set of studies, and presents some conclusions about the role of culture in children's sex-typed preferences for colours, toys, and affordances. It discusses whether sex differences replicated across cultures, according to the tests of colour preferences in Chapter 3, toy preferences in Chapter 4, and affordance preferences in Chapter 5. The Discussion further argues that systems theory may provide a useful theoretical approach for cross-cultural results, using the results of the systems theory analyses in Chapters 6 and 7. The Discussion also includes a review of the limitations and contributions of this thesis, and it concludes by identifying a need for current theories of sex and gender development to consider the role of culture.

Chapter 3

Children's Preferences for Blue-Red Hues and High-Low Colour Saturation in Three Cultures

The colour pink is widely used as a marker for female gender, in mass communication (Merler, Cao, & Smith, 2015; Vaisman, 2016), mass media (Koller, 2008), and mass-produced children's clothes and toys (Auster & Mansbach, 2012; Sweet, 2013). The use of pink to signal female gender has been criticised, however, as limiting children's opportunities to engage with toys, activities, and careers that are signalled as other-gender by their typical colour (Liben & Bigler, 2002). The resolution of this debate depends, in part, on the cause for the cultural pairing of pink and female. Some argue that an inborn, universal, preference for pink in the female sex has been disseminated into wider culture (Alexander, 2003; Del Giudice, 2017). However, others argue that cultural gender stereotypes cause girls to adjust their preferences, through repeated exposure to pink clothes and toys (Palmer & Schloss, 2010), or through active construction of gender stereotypes and signals of social gender (Martin & Ruble, 2004).

Background Literature

Theories of Universal Colour Preferences

Some theories propose that colour preferences, and sex differences in colour preferences, are universal across all human cultures. Biological process, in particular, have been proposed as universal explanations for sex differences in human colour preferences (Alexander, 2003). However, accounts differ as to the exact dimensions of colour preferences that might be universally sex-typed. Some authors focus on sex differences in preferences for specific colours – for example, a female preference for reddish-pink hues (Del Giudice, 2017). Others focus on sex differences in preferences for other aspects of colour, such as brightness (Semin & Palma, 2014). The following section reviews the most widely discussed biological theory of colour preferences, cone-contrast theory (Hurlbert & Ling, 2007). Unless otherwise noted, all research reviewed here was done in a minority world context.

Cone-Contrast Theory. Colour preferences are of interest to researchers seeking to establish whether humans have universal preferences for sensory stimuli. Cone-contrast theory (Hurlbert & Ling, 2007) posits that colour preferences are a biological adaptation based on evolved characteristics of the human visual system: specifically, the opponent-process model of neural encoding of colour (Hurvich & Jameson, 1957). Retinal cone photoreceptors are sensitive to different, but overlapping, wavelengths of colours, either short (bluish colours), medium (greenish colours), or long (reddish colours). The neural signals received by different photoreceptors can be contrasted to give a measurement of colour processing, approximated by comparing long-wavelength and medium-wavelength cone signals (the L-M opponent process) and then by comparing short-wavelength signals with the combined long- and medium-wavelength cone signals (the S-(L+M) opponent process) (De Valois & De Valois, 1993). Since the underlying retinal and neural structures are similar in all modern humans, the cone-contrast theory predicts that preferences for specific colours are also universal, since these probably evolved in line with the behavioural uses of colour vision (Hurlbert & Ling, 2007).

The cone-contrast theory has also been extended to predict universal sex differences in colour preferences (Hurlbert & Owen, 2015). For example, using a computerised task, Hurlbert and Ling (2007) asked young adults to select their preferred colour from a series of pairs of rectangles, and then analysed the preference responses with colour curve analyses. Colour preferences were best explained by two components that corresponded to short-axis (S-axis) and long-medium-axis (L-M axis) neuronal cone-opponent contrast channels. Their analysis found sex differences in participants' preferences for the L-M axis, which runs approximately from reddish to blue-green hues. Female participants tended to prefer reddish-purple hues, independently of brightness and saturation, while male participants tended to prefer blue-green hues, although to a lesser extent. The study also claimed a cross-cultural finding, since 37 out of the 171 participants were Chinese (the remainder were British). The utility of this cross-cultural data is disputable, however, since the testing took place at a British university and was conducted in English, and therefore the participants, both British and Chinese, may all have been exposed to British cultural information regarding gender and colour preferences.

Hurlbert and Ling (2007) concluded that the sex difference in colour preference was universal, and furthermore, that it was a direct biological adaptation to female-specific environmental cues that were important in human evolutionary history, such as gathering red fruits, or selection of healthy mates. Their findings, however, have not replicated in samples of non-human primates that are evolutionarily close to humans. A study of great apes found no sex differences in gorillas, chimpanzees, or the combined group of primates, in preferences for touching or looking at blue, green, or red stimuli (Wells, McDonald, & Ringland, 2008). A general preference was found for blue-green coloured stimuli over reddish stimuli in both male and female non-human primates.

The mate selection hypothesis is also difficult to reconcile with findings that sex, not sexual preference, predicts colour preferences in humans. A study of heterosexual and non-heterosexual North American college students found that gay males showed similar colour preferences to straight males, and not to straight females (Ellis & Ficek, 2001). Males overall showed a preference for blue, whereas females did not show a preference for pink. Furthermore, other research suggests that although women like pink more than men do, they do not prefer pink to blue (Wong and Hines, 2015). This absence of a preference for pink over blue in women also argues against the mate selection hypothesis as the sole explanation for sex differences in colour preferences.

The cone-contrast theory hypothesises sex-differentiated preferences for specific colours (reddish colours for females, blueish colours for males) that are culturally universal and arise from evolutionarily adaptive processes in all modern humans. It predicts that males and females should have similar colour preferences in different cultures (Hurlbert & Ling, 2007), and that hue preference should be maintained across saturation combinations (Hurlbert & Owen, 2015). Some authors further suggest that the most important contrast in previous studies for explaining variation in male and female preference is the red-blue contrast, not the blue-yellow or red-green dimensions suggested by the cone-contrast model (Hurlbert & Owen, 2015).

Theories of Non-Universal Colour Preferences

Other theories propose that colour preferences are not universal across human cultures. Individual experiences, in particular, have been proposed as explanations for non-universal colour preferences. The following section reviews the most influential theories of non-universal colour preferences: ecological valence theory (Palmer & Schloss, 2010), which proposes that individuals learn colour preferences from positive and negative experiences with coloured objects; and cognitive theories of gender (Martin & Ruble, 2004), which state that colour preference develops as a conscious signal of gender group membership.

Ecological Valence Theory. Ecological valence theory (Palmer & Schloss, 2010) holds that human colour preferences arise from an individual's experience with coloured objects. According to this framework, human preferences for specific colours are context-dependent. Humans develop preferences for colours that are associated with pleasant experiences. For example, Palmer and Schloss (2010) created object-related valence scores for 32 chromatic colours by asking adult participants to name objects associated with those colours, and then asking a different group of participants to rate the emotional valence of those objects, separately from colour. These object-related valence scores explained colour preferences in a separate sample, and also performed better than scores based on the cone-contrast model (Hurlbert & Ling, 2007) or based on direct ratings of the colours' emotional valence (Ou, Luo, Woodcock, & Wright, 2004). The authors concluded that colour preferences were learned from individual experience with coloured objects.

Ecological valence theory does not provide specific predictions for sex differences in colour preferences, but does predict that individuals' colour preferences should change according to their experience. One way to test this prediction is to compare the colour preferences of individuals with the same genetic sex, but different experiences of social gender. In a study of the impact of gender identity disorder (GID) on colour preferences, children aged between 3 and 12 years were asked to point to their three favourite colours from a printed display of 11 colours that varied in luminance, hue, and saturation (Chiu et al., 2006). Children who were genetically male, but identified as female, chose pink and purple colours significantly more than children who were genetically female but identified as male. Children in a control sample (without GID) showed stereotypical sex differences in colour preference, with females choosing pink and purple colours more than males. The researchers concluded that children choose favourite colours based on their subjective construction of social gender roles, and not based on genetic sex. Ecological valence might suggest that children who were genetically male, but identified as female, may prefer pink and purple because they may have had pleasant experiences with toys and clothes that come in these colours.

Ecological valence theory is supported by studies finding that children's sex-typed colour preferences develop after their preferences for coloured objects. In a preferential looking task, infants aged 12 months and 24 months showed no sex differences in colour preferences, but did show sex differences in preferences for toy dolls and vehicles (Jadva et al., 2010). The authors suggested that children develop colour preferences based on the colours that their preferred toys tend to come in. For example, many of the toys that female children play with are pink, and so they may learn to associate the colour pink with enjoyable activity, and thereby develop a preference for it.

Cognitive Theories of Gender. Cognitive theories of gender vary in their exact formulation, but often refer to gender schemas and stereotypes as drivers of sex-typed behaviour. Some cognitive theories state that children assemble large associative networks of information about gender, called gender schemas (Bem, 1981), and that the attribution of gender-typed information to other people may be called stereotypes (Carter & Levy, 1988; Martin & Dinella, 2012). Past research has shown that children are likely to prefer objects and activities that they identify as relevant to their gender, based on these schemas and stereotypes. For example, a study of children's attitudes towards occupations, activities, and traits (Liben & Bigler, 2002) indicated that children were likely to indicate a personal preference for things that they perceived to be appropriate for their gender. Furthermore, the relationship between children's personal preferences and gender attitudes was domain-general, in that children who had highly sex-typed preferences and attitudes in one domain were likely to have highly sex-typed preferences and attitudes in other domains. The authors concluded that gender schemas and stereotypes formed an important underlying driver of children's preferences.

Studies of children's toy preferences suggest that colour is a salient part of children's gender schemas and stereotypes. Children's toys are often coloured differently, according to whether the toy is intended for boys or for girls (Koller, 2008; Pomerleau, Bolduc, Malcuit, & Cossette, 1990). Weisgram and colleagues (Weisgram et al., 2014) reported that altering the colours of sex-typed toys could influence children's preferences for the toys. Colour also influenced children's judgements of whether the toy was suitable for boys or for girls. The authors concluded that colour was a salient marker for gender stereotypes, and furthermore, that children's toy preferences were partly driven by these stereotypes.

Adult Colour Preferences in Non-Industrialised Contexts

Cross-cultural colour preferences are most effectively studied in cultures with little or no industrialisation, and perhaps in remote locations, because of these cultures' relative isolation from other cultures' stereotypes. Stereotypes, especially widely accepted stereotypes such as pink as a marker of female gender, are easily accessed and spread through mass media, mass communication, and mass-produced goods (Aubrey & Harrison, 2004; Auster & Mansbach, 2012; Goldstein, Buckingham, & Brougère, 2004; Sweet, 2013). Previous studies have found similarities in colour preferences between industrialised cultures (Hurlbert & Ling, 2007; Ou et al., 2004; Yokosawa, Schloss, Asano, & Palmer, 2016), but it is impossible to separate the effect of culture from biology in these studies, since the industrialised cultures always have access to stereotypes disseminated by mass media, mass communication, and mass-produced goods. To separate the effect of cultural stereotypes from biology, then, it is necessary to find a cultural context that does not include regular access to mass media, mass communication, or mass-produced goods (Taylor, Clifford, & Franklin, 2013).

Studies of adults' colour preferences in remote, non-industrialised cultures report varying results. For example, Taylor and colleagues (2013) compared colour preferences in 42 adults from Britain and 38 adults from the semi-nomadic Himba group in northern Namibia. Participants were asked to rate their preference, on a computer screen, for 12 colours individually on an 11-point scale. Himba adults had different colour preferences to British adults, and their colour preferences were not statistically predicted by the cone-contrast model or the object valence model. According to the authors' analyses, Himba preferences were best explained by colour saturation, with Himba adults preferring more saturated colours. The authors suggest that the cone-contrast model fails to explain cross-cultural variation in colour preferences because it accounts only for hue, and not for other colour dimensions such as saturation or lightness. Himba adults did not show any sex difference in colour preference, and the study also found no preference for reddish hues in British women.

However, another cross-cultural study suggested that sex differences in adult colour preferences might replicate across cultures. Sorokowski and colleagues (Sorokowski, Sorokowska, & Witzel, 2014) presented a printed colour wheel of 12 colours to 108 adults in the remote Yali group in Papua New Guinea, and asked them to select a single favourite and a single least favourite colour. Yali adults showed a preference for red and yellow hues, and a significant sex difference in selection of a favourite colour, with females more likely to select reddish colours and males more likely to select blue-green colours. The cone-contrast theory did not statistically explain participants' preferences, however, and the authors suggested that contextual variables may be responsible for similar sex-typed colour preferences in both the Yali adults and in their comparison sample of Polish adults.

Adult and Child Colour Preferences

In general, adults show different colour preferences to children (Zentner, 2001) and infants (Taylor, Schloss, Palmer, & Franklin, 2013). Additionally, the magnitudes of sex differences in colour preferences appear to change with age. Infants show few sex differences in colour preferences, but children show larger sex differences than adults do. For example, Wong and Hines (2015) assessed toddlers' preferences for pink and blue, between the ages of 18 months and 4 years. Children were asked to point to their preferred colour in a series of paired comparison colour cards, with a total of nine pairs including three blueish colours and three pinkish colours. At age 2, both boys and girls preferred pink, but at age 3, boys had changed their average preference to blue. In contrast, adult women did not show a preference for pink (Wong & Hines, 2015). Similarly, LoBue & DeLoache (2011) offered a large cross-sectional sample of children aged 7 months to 5 years a choice between pairs of objects, where one object in the pair was always coloured pink. Male and female children chose pink objects with similar frequency before age 2, but female children older than 2.5 years showed a significant preference for pink objects, while male children showed an increasing avoidance of pink with age. Overall, previous research suggests that sex differences in colour preferences are likely to be larger in children over the age of 3 years, than in adults or infants. However, no study has previously tested colour preferences in children from cultures without access to mass communication, mass media, and mass-produced goods. This gap is the focus of the current study.

The Current Study

The current study presents a cross-cultural investigation of children's colour preferences, specifically preferences for blue and red hues, including pink, across three cultures: two non-industrialised, remote cultures, and one industrialised culture, for comparison.

Specifically, this study provides:

1. Planned comparisons of male and female children's preferences for pink and blue across three cultures;
2. Planned comparisons of male and female children's preferences for high saturation (red and blue) and low saturation (pink and light blue) versions of the stimuli, across three cultures;
3. Colour curves, to assess the best predictor of colour preferences across sex and culture groups.

Method

Participants

Culture and Age Requirements. The purpose of the current study was to replicate sex-typed colour preferences, especially girls' preference for pink, in two majority world cultures. Therefore, the cultures selected had to include ones in which children had no exposure to mass communication, mass media, or mass-produced clothing or toys that might give contextual information about pink as a gender marker in minority world cultures. Furthermore, adults and infants show different colour preferences. Since the purpose of the current study was to examine the universality of girls' preference for pink, and not to examine wide-ranging colour preferences, the target age for the current study was children of approximately primary school age.

As part of a larger cross-cultural study of children's toy and other preferences, data on colour preferences were gathered from three populations: children living in remote villages in the Shipibo communities of lowland Peru; children living in remote villages in the *kastom* Navhal-speaking communities of Tanna island, Vanuatu; and for comparison, children in Brisbane, Australia.

The colour preference measure was administered after a toy preference task, and only if there was time available. Therefore, the number of children that completed the colour preference task is different in each site[2].

A power analysis showed that with an effect size similar to that seen in previous studies of sex differences in children's colour preferences, a minimum of 8 boys and 8 girls from each culture would be required to provide 80% confidence of detecting a statistically significant effect with alpha set at 0.05. To account for increased statistical noise due to variable field conditions, more participants were recruited than were required according to the power analysis.

Brisbane Sample. Colour preference data were collected for 50 children (30 male and 20 female) in a public, free entry museum in Brisbane, Australia. Brisbane is a large city with access to radio, television, mobile, and internet networks that is typical of a large Western industrialised city. Children were predominantly white and had English as a first language. The task was conducted in English.

Shipibo Sample. Colour preference data were collected for 100 children (45 male and 55 female) in four villages in remote areas of the Peruvian Amazon basin, along the Ucayali

[2] There is a fourth population in the larger study of toy preferences, but time constraints prevented the collection of data on colour preferences in that population.

River. These villages were not accessible by road, only by boat, and had no radio, television, mobile, or internet coverage. All children were part of the Shipibo indigenous group and the task was delivered in both Spanish and Shipibo languages.

Navhal Sample. Colour preference data were collected for 51 children (31 male and 20 female) in two villages in remote areas of Tanna Island, Vanuatu. These villages were sometimes accessible by road, but most villagers rarely travelled except on foot. The villages had no radio, television, mobile, or internet coverage. Villagers live according to *kastom*, which specifically discourages engagement with mass-produced goods, mass media, mass communication, or industrialised cultures. Children were part of the Navhal indigenous language groups and the task was delivered in either Bislama (the trade language of Vanuatu) or Navhal language, depending on the translator's assessment of which language the child understood best.

Materials and Procedure

Colour Preference Measure. Although colour preference studies often cover many colours, the method used in this thesis focused on red-blue colour contrasts, and particularly on girls' preference for pink. There was limited time available to administer colour preference tests after the toy preference test, and before participants became fatigued. The colour preference measure was therefore narrowed in focus, to contrasts that were most theoretically important to finding sex differences. Red-blue contrasts are the most important for finding sex differences according to the cone contrast theory (Hurlbert & Owen, 2015), and also include pink, which is the colour that is most sex-typed in children's environments in minority world cultures (Pomerleau et al., 1990), and therefore most important to ecological valence theory. Pink has also been singled out in cognitive theories of gender (Weisgram et al., 2014), and prior studies of children's sex-typed colour preferences have focused on red and blue hues (Jadva et al., 2010; Wong & Hines, 2015).

Stimuli were presented as printed, laminated pages. Children were shown a page depicting two coloured squares. The child was asked to point to the square that they preferred. This process was repeated three times, for a total of three comparisons. These questions were adapted from a larger measure of colour and shape preference (Jadva et al., 2010), and using colour cards in a similar method to a previous test of colour preference in children (Wong & Hines, 2015).

Colour Selection. Three pairs of colours were compared. The key comparison was a pink square, of the "baby pink" colour used in girls' toys, with a blue square. Pink is a less saturated colour than blue, and children might therefore be reacting to the intensity of the colour rather than the colour itself. Therefore, we also asked children to choose between a red square and a blue square (high saturation versions of both colours), and to choose between a pink square and a less saturated blue square, of a "light blue" colour (low saturations versions of both colours). Children thus completed three choices overall: pink vs. blue (gender-typical comparison), red vs. blue (high-saturation comparison), and pink vs. light blue (low-saturation comparison). In this design, a reddish hue was always compared to a blueish hue, and saturation was matched for two comparisons.

Statistical Methods

Planned Comparisons: Pink and Blue, Red and Blue, Pink and Light Blue. Children's overall preferences for each of the paired stimuli were analysed using binomial tests. The binomial test compares the proportion of successful trials, out of a total number of trials, to that expected by chance. In this case, children's selection of a reddish hue was assigned a score of 1, and selection of a blueish hue was assigned a score of 0. The number of children expected to choose a reddish hue purely by chance, out of two possible options, is 50%. A significant result indicates that children chose one of the two options more frequently than would be expected if they were responding randomly.

Planned Comparisons: Sex Differences. The proportion of girls who chose reddish hues in each trial was compared to the proportion of boys who chose reddish hues in each trial using odds ratios. The odds ratio compares the proportion of successful trials in one group to the proportion of successful trials in another group, and it is robust to differences in group sample size. An odds ratio of 1 indicates equal odds, that is, that both groups are similarly likely to succeed. In this case, an odds ratio of 1 indicates that both boys and girls are equally likely to select reddish hues. Odds ratios are tested for statistical significance using Fisher's exact *p* value. A significant result indicates that one group of children (boys or girls) was more likely to choose reddish hues than the other group.

Colour Curves. Colour curves plot individuals' preferences for different colours along a continuous scale, and then compare these curves between male and female participants to evaluate sex differences. To add to the interpretation of the planned comparisons, colour curves were created following methods in previous literature (e.g., Hurlbert & Ling, 2007). The advantage of colour curves is that the effects of hue, saturation, gender, and culture can be evaluated in a single model that accounts for the correlated structure of the data. Each child responded to multiple items, and children's preferences within each culture may be correlated. Colour curves can account for this dependency. Colour preference scores were calculated as the mean preference for a single colour over all trials in which that colour was presented (range = 0-1). A structural equation modelling approach was adopted, in which linear mixed models were compared to one another to establish which combination of predictor variables best explained variation in colour preference.

The structural equation models were linear mixed models based on a normal (Gaussian) distribution; models based on binomial and Poisson distributions were also tested, as these might have been more suited to count data, but they had lower statistical fit to the data than the Gaussian models did, and were discarded. The models included a random effect of participant, to account for the correlated scores obtained when each participant provided preference scores for more than one colour. The outcome variable was the colour preference score, and the predictor variables included individual variables for the participant (gender and culture) and colour variables (hue and saturation).

Candidate models were: a null model (1), in which the only predictor was the random effect (in other words, children's preferences were only predicted by the individual child, and by no characteristics of the colour and no group effects); a partially specified model (2), which included main effects of participant and colour variables; and a fully specified model (3), which included main effects and interactions between participant and colour variables. The best performing model was selected according to a combination of model selection criteria (AIC, BIC, and deviance), and then interpreted.

$$y_{ij} = 1|j \quad (1)$$

$$y_{ij} = hue_i + saturation_i + gender_j + culture_j + 1|j \quad (2)$$

$$y_{ij} = hue_i + saturation_i + gender_j + culture_j + interactions + 1|j \quad (3)$$

Structural equations for null (1), partially specified (2), and fully specified (3) models of colour preference, where:

y_{ij} = preference score for colour *i* in individual *j*

$1|j$ = random effect for correlated colour preferences from individual *j*

hue_i = hue of colour *i* (red or blue)

$saturation_i$ = saturation of colour *i* (high or low)

$gender_j$ = sex of individual *j* (male or female). Gender refers to biological sex, but is called *gender* here to avoid confusion with *saturation* of colour in the acronyms for interactive terms

$culture_j$ = culture of individual *j* (Brisbane, Shipibo, or Navhal)

$interactions$ = a set of terms for the interactions between gender and colour variables. Specifically: hue*saturation, gender*hue, gender*saturation, gender*culture, and gender*hue*saturation

Results

Planned Comparisons: Pink and Blue

Preferences Across Both Sexes. Across male and female children combined, there was no overall preference for pink or blue in the Brisbane (binomial p = .672), Shipibo (binomial p = .555), or Navhal (binomial p = .542) samples. Tables in Appendix 1 present the frequencies of children's selections of pink and blue, with results of binomial tests and odds ratios.

Sex Differences. In the Brisbane sample, children showed sex differences in preference for pink or blue, such that girls preferred pink more than boys did, and boys preferred blue more than girls did, odds ratio (OR) = 0.26, 95% CI = 0.06 – 1.03, p = .039. There were no sex differences in preference for pink or blue in either the Shipibo (OR = 1.00, 95% CI = 0.42 – 2.39, *p* > .999) or Navhal (OR = 1.14, 95% CI = 0.30 – 4.48, *p* > .999) samples.

Planned Comparisons: High Saturation: Red and Blue

Preferences Across Both Sexes. Across male and female children combined, there was no overall preference for red or blue in either the Brisbane (binomial $p = .332$), or Shipibo (binomial $p > .999$) samples. In the Navhal sample, there was a significant preference, such that children overall preferred red to blue (binomial $p = .001$). Tables in Appendix 1 present the frequencies of children's selections of red and blue, with results of binomial tests and odds ratios.

Sex Differences. There were no significant sex differences in preference for red or blue in the Brisbane (OR = 0.48, 95% CI = 0.12 – 1.79, $p = .246$), Shipibo (OR = 0.70, 95% CI = 0.29 – 1.67, $p = .423$) or Navhal (OR = 2.93, 95% CI = 0.58 – 16.80, $p = .157$) samples.

Planned Comparisons: Low Saturation: Pink and Light Blue

Preferences Across Both Sexes. Across male and female children combined, there was a significant overall preference for light blue over pink in every culture: Brisbane (binomial $p < .001$), Shipibo (binomial $p < .001$), and Navhal (binomial $p = .013$). Tables in Appendix 1 present the frequencies of children's selections of pink and light blue, with results of binomial tests and odds ratios.

Sex Differences. There were no significant sex differences in preference for pink or light blue in the Brisbane (OR = 1.16, 95% CI = 0.19 – 8.52, $p > .999$), Shipibo (OR = 0.65, 95% CI = 0.24 – 1.70, $p = .380$) or Navhal (OR = 0.86, 95% CI = 0.22 – 3.44, $p > .999$) samples.

Colour Preference Curves

Linear mixed models tested the independent contribution of individual person variables (sex and culture) and colour variables (hue and saturation) across the whole sample of 201 participants. Three models were tested: a null model, a partially specified model including main effects, and a fully specified model including main effects and sex interactions. Model comparison parameters are shown in Table 3.1.

Model	Random effects	Main effects	Interactions	Model fit statistics	Model comparison
Null (1)	Child	None	None	Df = 3 AIC = 906.34 BIC = 920.32 Dev = 900.34	
Partially specified (2)	Child	Hue Saturation Gender Culture	None	Df = 8 AIC = 901.96 BIC = 939.23 Dev = 885.96	$\chi^2(5) = 14.38$ $p = .013$
Fully specified (3)	Child	Hue Saturation Gender Culture	S*H G*H G*S G*C G*S*H	Df = 14 AIC = 855.19 BIC = 920.42 Dev = 827.19	$\chi^2(6) = 58.77$ $p < .001$

Table 3.1. Model fit estimates for null (1), partially specified (2), and fully specified (3) structural equation models for colour curves. Significant model comparison values indicate significantly better fit to the data than the previous model, using ANOVA for model comparison. Random effects are random intercepts for each child, allowing correlated scores for different colour preferences from each child. Gender refers to genetic sex, but it is called gender here (G) to avoid confusion with saturation of colour (S) in the acronyms for interactive terms. Interactive terms: S*H = Saturation*Hue, G*H = Gender*Hue, G*S = Gender*Saturation, G*C = Gender*Culture, G*S*H = Gender*Saturation*Hue. Df = degrees of freedom of the model. Dev = deviance.

Model selection. The best performing model, according to AIC and deviance parameters, was the fully specified model. The BIC parameters indicated that the most preferred model was the null model, which is expected because BIC penalizes model fit more strongly for having more parameters. BIC also prefers the fully specified model to the partially specified model, and given the strong penalties for extra parameters, supports the fully specified model. Overall, the fully specified model fits the data significantly better than a partially specified or null model. The fully specified model (3) was selected for interpretation.

Model description. The fully specified model was a linear mixed model with a random effect of child, main effects of child gender, cultural context, colour hue, and colour saturation, interaction effects of hue with saturation and gender with hue, saturation, and cultural context, and a three-way interaction of gender, saturation, and hue. The model covariates are summarised in Table 3.2 and are interpreted below the table.

Covariate	**Estimate (b)**	**Standard error**	**t value**	***p* value**
Gender (0=Female)	0.02	0.08	0.22	.826
Culture (0 = Brisbane)				
Shipibo	0.04	0.12	0.30	.761
Navhal	0.07	0.15	0.46	.648
Hue (0 = Blue)	0.22	0.14	1.61	.108
Saturation (0=High)	0.26	0.14	1.92	.055
Hue*Saturation	-0.38	0.19	-1.97	.049
Gender*Culture				
Shipibo	0.01	0.07	0.16	.872
Navhal	0.05	0.09	0.61	.540
Gender*Hue	-0.06	0.08	-0.77	.441
Gender*Saturation	<0.01	0.08	0.02	.988
Gender*Hue *Saturation	-0.05	0.12	-0.40	.693

Table 3.2. Results of the fully specified model of colour preference. *p* values produced by normal approximation. Interaction effects are indicated with *.

Person variables. No individual person variables explained significant variation in colour preferences across the whole sample. Gender and culture were not statistically predictive of children's colour preferences. Gender did not significantly interact with hue or saturation, indicating no universal gender effect on colour preference.

Colour variables. Hue and saturation did not significantly predict colour preferences, but the direction of the main effects indicated that children preferred reddish hues and lower saturation. The interaction between hue and saturation was statistically significant, so colour preference curves were created to examine this interaction.

Colour preference curves. Children's colour preferences were plotted as colour curves, with the x-axis representing a gradient from low saturation reddish hues to low saturation blueish hues, and high saturation hues in the middle. Variation in these colour preference curves was statistically explained by a general preference for low saturation blues and high saturation reds, as shown in Figure 3.1. There were no sex differences in the overall shape of the colour preference curves, but children in the Brisbane sample showed a slightly different curve than children in the Shipibo and Navhal samples, due to a sex difference in preference for pink that was present in the Brisbane sample but not in the other cultures.

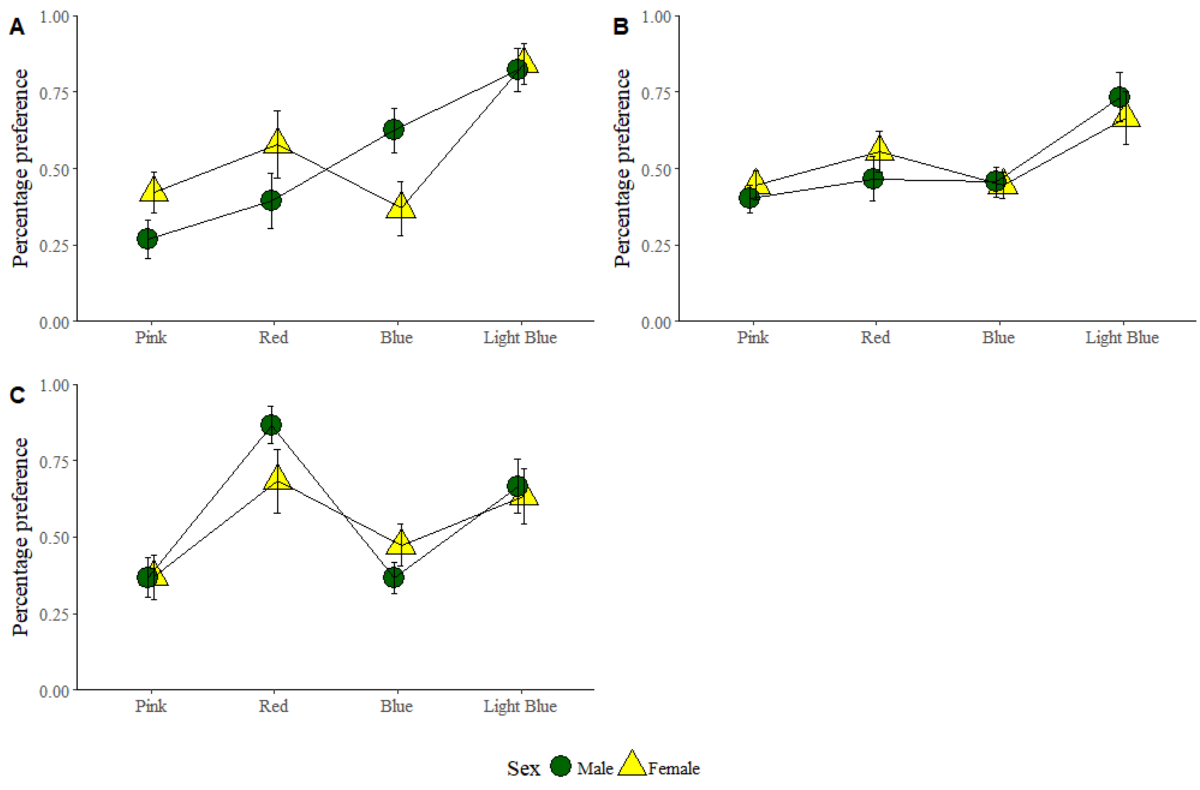


Figure 3.1. Colour curves for red-blue hues and low-high saturation in **(A)** Brisbane **(B)** Shipibo **(C)** Navhal samples. Black lines with circle point markers represent boys and grey lines with triangle point markers represent girls. Error bars show the standard error of the mean preference score.

Discussion

Sex differences in children's colour preferences did not replicate across cultures. Although children in Brisbane, Australia showed sex differences in colour preferences, children in two non-industrialised majority world samples, the Shipibo people in the Lake Imiria region of Peru, and the Navhal language group of *kastom* villages in Tanna Island, Vanuatu, did not.

No Universal Female Preference for Pink

No sex difference in preference for pink was found in either the Shipibo or Navhal samples. Girls showed a preference for pink only in the industrialised, English-speaking, minority world sample. This finding suggests that, rather than being universal, girls' preference for pink is specific to industrialised minority world cultures.

Although the present study is the first study of sex differences in children's colour preferences in non-industrialised cultures, there are a few prior studies of adult colour preferences in non-industrialised cultures. One prior study of adult colour preferences in nomadic Himba adults in Namibia, also found no sex differences in colour preferences (Taylor, Clifford, et al., 2013), similar to the findings of the present study. In contrast, another study found a sex difference in Yali adults' favourite colours in remote Papua New Guinea (Sorokowski et al., 2014). This apparently contradictory result may be a consequence of the latter study's method for measuring colour preferences in adults. In that study, adults were asked to choose only a single favourite colour from a set of 12, but most studies of colour preferences ask participants either to rate individual colours (e.g., Taylor, Clifford, et al., 2013) or to choose from a series of paired comparisons (e.g., Hurlbert & Ling, 2007; Jadva et al., 2010; Wong & Hines, 2015; the present study). Asking for only a single favourite colour may have limited the reliability of the colour preference score, compared to asking about each colour separately.

The results of the present study suggest that sex differences in children's colour preferences are learned from cultural context, and this result is consistent with previous research on colour preferences in infants. Infants before the age of 2.5 years show no sex differences in preference for pink or reddish hues (Franklin, Gibbons, Chittenden, Alvarez, & Taylor, 2012; Jadva et al., 2010; LoBue & DeLoache, 2011; Wong & Hines, 2015; Zemach, Chang, & Teller, 2007), suggesting that these sex differences develop later in childhood. Furthermore, in the present study, across all three cultures studied, girls chose light blue hues more often than pink. Overall, the findings suggest that girls' preference for pink is likely to be culturally specific.

The results of the present study further suggest that sex differences in colour preferences may be specific to minority world cultures. Previous research on a range of psychological traits, including visual perception and spatial reasoning, also suggests that much research on these traits may be specific to participants from minority world cultures (Henrich et al., 2010b). Like many areas of psychological research, colour preference research has often assumed that its participants are representative of all humans, and therefore, that the sex differences found in these minority world populations generalise universally to all humans (e.g., Alexander, 2003; Ellis & Ficek, 2001; Hurlbert & Ling, 2007). The findings presented here suggest instead that sex differences in minority world participants' colour preferences may not generalise to other human cultures.

Preferences for Light Blue

The present study found a general interaction between hue and saturation, such that children generally preferred light blue to other colours. This colour preference is more similar to minority world results for adults than for infants. A previous review found an overall preference for blue hues in adults (Hurlbert & Owen, 2015) and previous studies have found a general preference for reddish hues in infants (Adams, 1987; Franklin et al., 2012; Jadva et al., 2010), although some have found that infants prefer both blue and red over other colours (Teller, Civan, & Bronson-Castain, 2004; Zemach et al., 2007) or that infants have no hue preferences (Taylor, Schloss, et al., 2013).

The interaction between hue and saturation in the present study indicates that both hue and saturation are important to children's colour preferences. In the present study, children's preference for light blue was not only due to a preference for blue over red hues, since children did not prefer saturated blue to saturated red; indeed, Navhal children preferred saturated red to saturated blue. These combined results suggest that there may be some universal rules governing children's colour preferences across different cultures, but that these rules are more complex than simple universal preferences for particular hues. The interaction between hue and saturation is clearly important to children's colour preference, and future work should include colours that vary both in hue and in saturation.

Relevance to Theories of Universal Colour Preference

The cone-contrast theory of colour preferences states that human colour preferences are evolved behavioural adaptations to specific biological features of the human visual-neural colour processing system (Hurlbert & Ling, 2007). Proponents of the cone-contrast theory predict that humans across different cultures should show similar preferences for hues, and that females across cultures should show a greater preference than males for reddish hues (Hurlbert & Owen, 2015). This prediction was not supported by evidence from Shipibo or Navhal children in the current study, suggesting low evidentiary support for the cone-contrast theory in its current form.

However, the colour curves analysis in the current study provides some support for the possibility of universal patterns in human colour preferences. Colour variables (hue and saturation) were more important to predicting colour preference scores than were individual person variables (gender and culture). The importance of colour variables over person variables suggests that there may be some universal rules governing preference for hues and saturation, and that these may operate similarly across sex and culture.

Relevance to Theories of Non-Universal Colour Preference

Ecological valence theory states that human colour preferences vary according to individual life experiences with coloured objects (Palmer & Schloss, 2010). As such, ecological valence theory suggests that humans should show different preferences for hues if their environment contains different coloured objects, and these objects are associated with pleasant or unpleasant experiences. The ecological valence theory is supported by one of the results of the current study: girls' preference for pink in the Brisbane sample, but not in the Shipibo or Navhal samples. Girls in the Shipibo and Navhal villages did not have access to mass-produced toys or clothes in pink. Although the Shipibo villages had some second-hand mass-produced children's clothing from donations, observations of children's dress patterns during data collection suggested that girls' and boys' clothing was not differentiated on the basis of colour. Children in the Navhal villages did not typically have any access to mass-produced clothing, and although clothing was gender differentiated (grass skirts for girls and penis sheaths for circumcised boys), the grass clothing was not typically dyed or coloured. Children in Brisbane, however, were raised in a cultural context where pink clothing and toys are marketed to, and possessed by, girls (Auster & Mansbach, 2012; Pomerleau et al., 1990; Sweet, 2013). Girls in Brisbane, then, would likely have had positive experiences with pink coloured objects, while girls in Shipibo and Navhal villages would not. According to ecological valence theory, these cultural differences in girls' experience with pink objects could explain cultural differences in girls' ecological valence for the colour pink, and subsequent preference.

Cognitive theories of gender, unlike ecological valence theory, do not require that children have experience with pink objects directly. Instead, cognitive theories of gender might predict lower preferences for pink in cultural contexts where pink is not stereotypically associated with female gender. In the present study, children in the Shipibo and Navhal samples were not likely to have had access to mass media and mass communication. They therefore were not likely have had access to information about minority world stereotypes, including the use of pink as a marker of female gender (Koller, 2008; Vaisman, 2016). Children in the Shipibo and Navhal samples, therefore, would not be expected to have gender schemas or stereotypes for pink, and so would not be expected to show a sex difference in preference for it. Consequently, the cross-cultural results presented here provide some support for the position that children's colour preferences are affected by broad underlying cognitions, such as gender schemas and gender stereotypes.

Limitations and Future Directions

The present study has some limitations. First, the colour preference task was limited to blue and red hues. Blue and red hues have been identified as most important to determining sex differences by proponents of universal colour preferences in general (Alexander, 2003) and cone contrast theory in particular (Hurlbert & Owen, 2015), and pink has been singled out by researchers focusing on sex differences in ecological valence (Taylor, Schloss, et al., 2013) and on cognitive theories of gender (Weisgram et al., 2014). The present study was designed specifically to elicit sex differences if they existed, and therefore focused on red and blue hues. Future investigations could test the cone-contrast theory more extensively, however, by including comparisons of colours that are more explicitly placed along the short-axis (S-axis) and long-medium-axis (L-M axis) neuronal cone-opponent contrast channels.

Additionally, the test of colour preferences used in the present study relied on a single presentation of each colour pair, using laminated cards, in varying outdoor field conditions. Previous studies have typically used computer screens, in controlled indoor environments (e.g., Jadva et al., 2010), although some have used laminated cards (e.g., Wong & Hines, 2015). Variability in the field conditions may have added statistical noise to the data, and the low number of presentations of the items may have reduced reliability, and both of these considerations may have contributed to the failure of this study to find significant gender differences in some cultures. However, the sample size was well in excess of the statistical power required to find a gender difference, so it is unlikely that statistical noise or reliability alone account for the lack of gender differences in some cultures. Additionally, the Brisbane sample displayed significant sex differences in colour preferences in the expected direction, suggesting that the null results in other samples were not due to unreliability. Further replication, with more repetition and using a range of presentation methods, may test some of these possibilities.

Future research might also usefully incorporate systematic measurements of coloured objects in children's environments, and measurements of culturally relevant gender schemas and stereotypes. The present study used procedures designed to be replicable across different cultural contexts, and to reproduce procedures used to assess behavioural sex differences in minority world cultures. It did not include protocols for gathering culturally specific information. Observations of the role of colour in Shipibo and Navhal communities in the current study are taken from field journals, which included some observations of daily life, but these were not made systematically. Future investigations could test the role of ecological valence more directly, by including a measurement of coloured objects in children's environments, and their emotions in relation to the objects. Future investigations could also test the role of gender schemas and stereotypes more directly, by including measurements of these.

Conclusions

The sex difference in preference for pink is not culturally universal, and it may be best explained as a combination of experience with pink objects and exposure to cultural stereotypes about pink as a marker for female gender. Both saturation and hue appear to be important to children's colour preferences, and future work might usefully investigate both. Findings of previous studies that used cross-cultural samples taken from other industrialised nations (e.g., Hurlbert & Ling, 2007; Yokosawa et al., 2016) might be best explained by the widespread dissemination of minority world gender stereotypes across mass media, mass communication, and mass-produced children's clothes and toys. It is possible that pink, as a marker for female gender, is specific to the context of minority world cultures.

Chapter 4

Children's Preferences for Sex-Typed Toys in Four Cultures

Sex-typed preferences, and their origin and development, remain a controversial and important topic. The sex-typed preferences of children are particularly of interest because these are observed early in development before pivotal mechanisms such as pubertal hormone changes and adult social roles become influential. In children, a particularly large sex difference is found for toy preferences: girls prefer to play with different toys than boys do. Toy preferences are also easily measured, using behavioural observations. For these reasons, toy preferences are a useful case for examining theories of sex and gender development more broadly.

However, theories of sex and gender development typically do not make explicit predictions about how children's toy preferences might differ cross-culturally, particularly in cultures that differ markedly from cultures where most of the existing research has been conducted. But these sex differences may not appear in cultures where these toys are relatively novel. The current study presents the first empirical comparison of children's preferences for sex-typed toys in remote cultures.

Background Literature

The current study examines children's toy preferences in remote cultures, and aims to discover whether sex differences in toy preferences may be culturally universal. *Culturally universal* and *culturally variable* here refer to similar categories of influence to those that are sometimes termed *nature* and *nurture* in broader gender psychology. The terms *nature* and *nurture* remain highly recognised and important in sex and gender studies, but have been typically used to refer respectively to biological and social influences (Eagly & Wood, 2013). However, biology is not the only category of explanation that can be culturally universal, and biological explanations also allow for cultural variation. Furthermore, as explored below, most theorists acknowledge a role for both nature and nurture in development. Therefore, the current study avoids the terms *nature* and *nurture*.

Many influential theories exist to explain sex differences in children's toy preferences, and these theories can be broadly divided into biological, social, and cognitive. Biological, social, and cognitive approaches do not typically consider culture as an important part of theory. These theories are very influential in toy preference research, however, and so are reviewed in the following section.

Biological Approaches

One approach that has the potential to explain culturally universal patterns in sex-typed toy preferences is derived from animal research on processes involved in sexual differentiation. According to this approach, genetic information coded on the sex chromosomes is responsible for sex differences in early hormone exposure, and these then contribute to differential preferences for sex-typed toys in children (Hines, 2010; Hines, Constantinescu, & Spencer, 2015). Typical variation in hormone levels has been linked to broadly defined sex-typed behaviours (Hines et al., 2016), but in the case of toy preferences specifically, extreme early hormonal environments show the most consistent predictive value (e.g., Nordenström et al., 2002; Pasterski et al., 2005; Servin et al., 2003). In these studies, girls who were exposed to high levels of androgen prenatally, due to CAH, show more male-typical patterns of toy and play preferences than their unaffected female relatives or matched controls (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Hines et al., 2002; Pasterski et al., 2011).

Sex differences in hormone exposure during early development, arising from genetic information on the sex chromosomes, have been proposed by some commentators to apply to all modern humans (e.g., Alexander, 2003). In support of this view, hormonal influences on behavioural development are well-established in non-human mammals (Hines, 2006), and sex differences have been observed in toy preferences in non-human animals (Alexander & Hines, 2002; Hassett et al., 2008; Hines & Alexander, 2008; Williams & Pleil, 2008). In the view of some theorists, these findings imply that preferences for specific sex-typed toys might be a human universal (e.g., Alexander, 2003; Hurlbert & Ling, 2007), although many other theorists suggest that social and cultural influences interact with biology to modify sex-typed preferences (e.g., Hines, 2013; Wallen, 1996).

Social Approaches

In contrast, one approach that has the potential to explain cultural variations in sex-typed toy preferences is derived from research on social learning (e.g., Fagot, 1985). Social learning theory posits that children's sex-typed behaviour is influenced by their social environment. Children in minority world cultures acquire information about sex-typed toys from parents (Idle, Wood, & Desmarais, 1993) peers (Martin et al., 2013), mass media (Aubrey & Harrison, 2004; Pike & Jennings, 2005), and the availability of mass-produced toys (Auster & Mansbach, 2012). Additionally, children are reinforced for behaviour that conforms to gender norms. For instance, parents and peers encourage sex-typed behaviour (Caldera, Huston, & O'Brien, 1989; Carter & McCloskey, 1984; Freeman, 2007; Martin et al., 2013).

Most theorising about social processes has considered immediate social influences such as parenting practices and sibling relationships, but typically has not considered more general cultural context (e.g., Fagot & Hagan, 1991; Martin et al., 2013). However, a logical prediction from social learning theory is that different cultural environments would produce different toy preferences. Children in majority world cultures have access to different information about sex-typed toys than do children in minority world cultures, especially in remote locations with no mass media or mass communication. In addition, children in majority world cultures would be less likely, or unlikely, to have a reinforcement history with mass-produced toys. Logically, according to social learning theory, these environmental differences could result in variations in sex-typed toy preferences between cultures (e.g., Witkin, 1979).

Cognitive Approaches

Other theories of sex and gender development consider children's internal cognitive processes to be important. In particular, cognitive theories posit that children selectively attend to and incorporate information about activities that are appropriate for their gender (Eaton, Von Bargen, & Keats, 1981; Weisgram et al., 2014). This information is incorporated into gender schemas, and these gender schemas are used to evaluate and contextualise new objects, such as toys (Martin et al., 2002).

Previous research demonstrates that children's internal gender schemas are important, over and above direct social learning. For example, children predict that their parents will approve of their play with sex-typed toys, and disapprove of their play with cross-sex-typed toys, even when their parents self-report a rejection of common gender stereotypes (Freeman, 2007). Another study measured children's development of gender schemas using a computerised test, and found that gender-schematic children were more likely than gender-aschematic children to change their play behaviour towards sex-typed toys when an adult observer was present (Wilansky-Traynor & Lobel, 2008).

Cognitive processes are likely to be important to children's toy preferences, but they may be difficult to assess in a cross-cultural study. Systems of thinking and communicating about gender are likely to be different across cultures. In addition, children's sex-typed play behaviour is likely to vary in different cultures, and so are the types of toys that might be gender stereotyped. Some of these cultural variations in children's play behaviour and gender are reviewed in the following section.

Gender and Play in Cultural Context

Gender and play have been studied as aspects of culture. Gendered behaviours, including children's behaviours, may vary cross-culturally (Edwards, 2000; Richert et al., 2017; Roopnarine & Johnson, 1994). Children's play has also been studied in cultural context, typically from an anthropological perspective, using primarily ethnographic methods such as systematic observations (e.g., Boyette, 2016; Drewes, 2005; Edwards, 2000).

The Six Cultures Study (Whiting et al., 1966) provided the first systematic documentation of cross-cultural variations in children's play. Researchers observed the play of children aged 3 to 10 years, in small isolated communities in Kenya (LeVine & Lloyd, 1966), Mexico (Romney & Romney, 1966), the Philippines (Nydegger & Nydegger, 1966), Japan (Maretzki & Maretzki, 1966), and India (Minturn & Hitchcock, 1966), and in a large city in the United States (Fischer & Fischer, 1966), between 1954 and 1956. Although the sample size in each location was relatively small (16-24 children), researchers observed child behaviour over several sessions to collect a large dataset. The focus of the Six Cultures study was on child rearing practices, and not on toy preferences, but a later re-analysis focused on play and gender (Edwards, 2000). The re-analysis concluded that role play might help children to prepare for adult roles, not only through practice of adult work (e.g., using mud and sticks to imitate adult food preparation), but also adult leisure. For example, children were observed to most avidly imitate adult pleasure past-times like smoking and talking on the telephone (Edwards, 2000). Although the reanalysis suggested that girls participated in-role play more than boys did in most cultures, it did not look for gender differences in children's preferences for particular toys or objects.

A cross-cultural study of children's toy preferences has not yet been reported. The Six Cultures study provided cross-cultural, empirical information about sex differences in children's play, but it took an ethnographic approach rather than an experimental one. In contrast, experimental psychology studies of children's toy preferences have been limited to the specific cultural context of the minority world. This is a limitation of the field, because results from the minority world cannot be assumed to generalise to other cultural contexts (Tudge et al., 2010). Further, based on research into the cultural universality of other cognitive and social phenomena, results from the minority world may be outliers that are not similar to human experience in the majority world (Henrich et al., 2010c, 2010b).

Some experimental studies have examined play in both industrialised and majority world samples, but these studies do not use the same methods in each setting, and therefore their ability to generalise results over the samples and to the larger human population is limited. For example, Bloch examined children's play activities in the US by telephoning parents and asking them to report on their children's activities (Bloch, 1987). The same author also examined children's play activities in Senegal by observing children's activities directly, but did not compare results between the US and Senegal samples (Bloch & Adler, 1994). Methodological variations like these prevent direct comparison of results between cultures. In contrast, the current study aims to test whether results from minority world cultures can be generalised to other cultures, by using the same method in all the cultures. For this purpose, it was necessary to identify a method that would be identical across situations, so that the results could be directly compared between cultures.

Method

Procedure

Research procedures were standardised across locations, with the aim of generating results that could be compared between cultures. Children were videotaped in a single 5-minute play session in a private play area. The play area was either an empty classroom or a space in a public area of the village, delineated by sheets and a mat. The play space contained twelve toys arranged in a circle, as shown in Figure 4.1.

The four boy-type toys (BTT set) were: a racing car (propulsive BTT), a Hulk action figure with accessories (figure BTT), a pirate costume (dress-up BTT), and a plastic sword (role play BTT). The four girl-type toys (GTT set) were: a My Little Pony carriage (propulsive GTT), a Barbie doll with accessories (figure GTT), a princess costume (dress-up GTT), and a baby doll (role play GTT). The four neutral toys were: a dinosaur costume, a set of plastic African animals, pencils and paper, and a stuffed dog. Toys were arranged in a random order but so that no two BTT, GTT, or neutral toys were adjacent. The random order was changed each day.

Figure 4.1. Example toy arrangement for the play session in a Shipibo village. The fourth side is left open to the jungle, so that the child cannot be observed but also knows that he or she is always free to leave.

Children were brought into the centre of the circle individually and told: “You can play with the toys however you like”, in the appropriate language. The experimenter and translator left the play area, and the child was left to play for 5 minutes, unsupervised. Play was recorded by two GoPro video cameras, positioned above and to the side of the play area in order to capture all angles of play. At the end of the 5 minutes the experimenter and translator re-entered the play area and told the child that the play session was finished. Children who had completed the activity were asked not to talk about it with other children, and protocols were enacted in each location to separate children who had and had not completed the experiment, so that each child was encountering the toys for the first time.

Required Sample Size and Statistical Power

The sizes of sex differences in toy preferences are large (d = 1.22 to 3.38), according to a recent meta-analysis (Davis & Hines, forthcoming). Therefore, for mean comparisons, within each country, a power analysis suggests the required sample size to be at least 11 girls and 11 boys to detect a comparable effect at power 0.8. For regressions, assuming a medium effect of gender on toy preference, controlling for age and culture, the required sample size would be 76 children overall. To maximise the chances of detecting any existing effect, the target sample size for the current study was 100 children in each location, with an approximately even split between boys and girls.

Video Coding and Reliability

Videos were coded per an *a priori* protocol, and coders recorded the start and end times that children contacted each toy. This coding protocol allowed for play with no toys, and for play with more than one toy at once. Coders had excellent inter-rater reliability, with an average intra-class correlation coefficient of 0.98.

Start and end times were used to calculate the total number of seconds that children spent playing with each toy. These play times were divided by the total number of seconds children spent contacting any of the twelve toys, to create a percentage of time spent playing with each toy. Percentage times were multiplied by 100 to give a proportional play time for each toy.

Statistical Analysis

Mean Comparisons and Effect Sizes. Statistical differences between groups were evaluated using t-tests with alpha set at 0.05. The magnitude of these differences was described by an effect size, the standardised mean difference (d). Analyses tested for *sex differences* (e.g., that boys play with BTT more than girls do) as well as *sex-typed preferences* (e.g., that boys play with BTT more than they play with GTT). The first set of analyses tested children’s preferences for BTT, GTT, and neutral toys. In previous research, the most commonly used BTT are a toy vehicle and a doll, respectively. Therefore, the second set of analyses tested children’s preferences for the baby doll and the toy racing car.

Age as a Covariate. The present analysis did not include age as a covariate. Age was used to identify children who were eligible to participate in the study, but was not used in the statistical analysis, for three reasons. First, the toys were selected to appeal to a wide age range, so age differences in toy preferences were not expected *a priori*. Second, age was not reliably estimated in all locations, and so analysing age may have introduced errors. A follow-up analysis, conducted *post hoc*, confirmed that including age as a covariate did not alter the substantive results of this thesis.

Tests for Cross-Cultural Effects. The aim of this study was to compare sex differences in toy preferences between a minority world culture (Brisbane) and three majority world cultures. Pairwise comparisons were therefore established *a priori* to test whether the effect sizes in the minority world culture (Brisbane) were statistically different to the effect sizes in each of the majority world cultures (Shipibo, Lenakel and Navhal). Cross-cultural effects were tested using meta-analysis of the effect sizes between different cultures, following a procedure recommended by Gleser & Olkin (2009). These statistical procedures were selected to account for dependency in the data introduced by testing multiple cultural groups against a single control. These methods also focus on effect sizes, rather than on omnibus statistical tests. These decisions reduced the possibility of Type 1 error due to multiple testing, as is explained below.

The meta-analysis method is designed for studies that compare multiple participant groups against a common control (Gleser & Olkin, 2009). This method reduced the possibility of artefactual findings due to multiple testing, as it avoided making statistical comparisons between all the possible combinations of the study groups. An alternative, using a standard factorial approach, might have been to conduct a 2 (sex) by 4 (culture) omnibus test (e.g., ANOVA), with preference for each toy type (BTT, GTT, and neutral toys) as three dependent variables. The next step in the factorial approach would be to follow up any significant interaction term with multiple *post hoc* pairwise *t*-tests. For each significant interaction term, a total of 48 follow-up *t*-tests would be required for all possible combinations of the study groups (for example: boys in Brisbane compared to girls in Brisbane; boys in Brisbane compared to boys in Lenakel; boys in Brisbane compared to girls in Lenakel; and so on). Applied to the three toy groups (BTT, GTT, and neutral toys), the standard factorial approach could have resulted in up to 144 separate *t*-tests. Such a large number of *post hoc t*-tests would increase the risk of finding spurious differences among the study groups.

Instead, the present research identified *a priori* statistical tests that provided theoretically useful insights: comparisons of each of the three majority world cultures with the minority world culture; and tests for sex differences within each culture. Limiting the analysis to statistical tests that were identified *a priori* decreases the total number of statistical tests, and decreases the risk of spurious findings, compared to standard approaches. Additionally, statistical analyses in this thesis focus on effect sizes, to further reduce the risk of spurious findings.

Because the pairwise comparisons were identified *a priori*, this thesis reports unadjusted *p* values and reports the results of null tests alongside tests that returned a statistically significant result.

Interpretation. If the cross-cultural meta-analysis effect (*dc*) is statistically significant, then the omnibus effect is statistically different from zero. Further, the effect may have a culturally universal component if it is found in each culture, especially if it is found both in the Brisbane sample (representing a cultural context where these toys are familiar, and where these toys have been previously studied) and in the Navhal sample (representing a cultural context where these toys are completely novel).

In addition, the meta-analysis produces an estimate of the statistical heterogeneity between effect sizes in each culture. If the heterogeneity estimate (*Q*) is statistically significant, then the effect sizes may arise from different underlying distributions, and the effect may have a component that is not culturally universal. Further, the effect may have a component that is not culturally universal, if the effect sizes are significantly different between different cultures. *A priori* pairwise comparisons tested whether the effect sizes in the minority world culture (Brisbane) were statistically different to the effect sizes in each of the majority world samples (Shipibo, Lenakel and Navhal).

Results

Excluded Data

Toy play data were collected for 390 participants. Participants under 4 years of age were removed from the dataset. The total number of participants removed for being under 4 years of age was 28 (20 from the Brisbane sample, 0 from the Shipibo sample, 3 from the Lenakel sample, and 5 from the Navhal sample). In addition, participants who played for less than one minute in total were removed from the dataset. The total number of participants removed for playing less than one minute was 55 (2 from the Brisbane sample, 6 from the Shipibo sample, 19 from the Lenakel sample, and 28 from the Navhal sample).

Sample Demographics

The final sample contained 307 participants (161 boys and 146 girls). There was no difference in average age (in years) between boys (M = 6.26, SD = 1.81) and girls (M = 6.39, SD = 1.74), t (286.27) = 0.62, p = .536. The four locations were not significantly different by gender ratio, F (3,303) = 1.41, p = .240.

The four locations were significantly different in average age, F (3,268) = 4.90, p = .002. The average age of children in the Shipibo sample (M = 6.87, SD = 1.77) was significantly higher than the average age of children in the Brisbane sample (M = 5.94, SD = 1.61), t (170.46) = 3.63, p < .001, in the Lenakel sample (M = 6.00, SD = 1.53), t (28.81) = 2.20, p = .036, and in the Navhal sample (M = 6.14, SD = 1.87), t (163.91) = 2.63, p = .009. The average age of children in Brisbane was not significantly different to the average age of children in the Lenakel sample, t (28.30) = -0.16, p = .875, or in the Navhal sample, t (154.48) = -0.72, p = .470. The average age of children in the Lenakel sample was not significantly different to the average age of children in the Navhal sample, t (32.33) = -0.34, p = .739.

Brisbane Sample. The final Brisbane sample contained 80 participants (42 boys and 38 girls). There was no difference in average age between boys (M = 5.93, SD = 1.77) and girls (M = 5.95, SD = 1.43), t (77.06) = 0.05, p = .958.

Shipibo Sample. The final Shipibo sample contained 94 participants (43 boys and 51 girls). There was no difference in average age between boys (M = 6.91, SD = 1.76) and girls (M = 6.84, SD = 1.80), t (89.52) = -0.18, p = .857.

Lenakel Sample. The final Lenakel sample contained 53 participants (27 boys and 26 girls). There was no difference in average reported age between boys (M = 5.70, SD = 1.01) and girls (M = 6.33, SD = 1.94), t (12.11) = 0.87, p = .401.

Navhal Sample. The final Navhal sample contained 80 participants (49 boys and 31 girls). There was no difference in average reported age between boys (M = 6.08, SD = 1.90) and girls (M = 6.23, SD = 1.86), t (65.12) = 0.34, p = .739.

Sex Differences

Effect sizes for sex differences in children's play with GTT, BTT, and neutral toys, are summarised in Figure 4.2. Mean proportion play times with each toy, effect sizes, and significance levels are also reported in Appendix 2.

Brisbane Sample. In the Brisbane sample, the BTT set (action figure, racing car, pirate costume, and sword) was played with by boys ($M = 79.91$, $SD = 26.44$) more than girls ($M = 17.16$, $SD = 20.11$), and this difference was statistically significant, $t\ (75.83) = -12.01$, $p < .001$. The GTT set (baby doll, barbie doll, princess costume, and pony carriage) was played with by boys ($M = 6.53$, $SD = 17.54$) less than girls ($M = 43.53$, $SD = 34.57$), and this difference was statistically significant, $t\ (53.61) = 5.94$, $p < .001$. The set of neutral toys (pencils, dinosaur costume, stuffed dog, and toy animals) was played with by boys ($M = 13.56$, $SD = 20.49$) less than girls ($M = 39.32$, $SD = 37.26$), and this difference was statistically significant, $t\ (56.21) = 3.78$, $p < .001$.

Shipibo Sample. In the Shipibo sample, the BTT set was played with by boys ($M = 76.49$, $SD = 21.62$) more than girls ($M = 14.72$, $SD = 21.49$), and this difference was statistically significant, $t\ (89.16) = -13.84$, $p < .001$. The GTT set was played with by boys ($M = 4.55$, $SD = 9.89$) less than girls ($M = 75.87$, $SD = 26.12$), and this difference was statistically significant, $t\ (66.17) = 18.03$, $p < .001$. The set of neutral toys was played with by boys ($M = 18.96$, $SD = 19.61$) more than girls ($M = 9.41$, $SD = 13.40$), and this difference was statistically significant, $t\ (72.20) = -2.71$, $p = .008$, and in the opposite direction to the Brisbane sample.

Lenakel Sample. In the Lenakel sample, the BTT set was played with by boys ($M = 55.87$, $SD = 26.87$) more than girls ($M = 29.19$, $SD = 24.29$), and this difference was statistically significant, $t\ (89.16) = -13.84$, $p < .001$. The GTT set was played with by boys ($M = 14.98$, $SD = 13.47$) less than girls ($M = 56.67$, $SD = 26.41$), and this difference was statistically significant, $t\ (66.17) = 18.03$, $p < .001$. The set of neutral toys was played with by boys ($M =29.16$, $SD = 25.95$) more than girls ($M = 14.14$, $SD = 14.77$), and this difference was statistically significant, $t\ (41.54) = -2.60$, $p = .013$, in the opposite direction to the Brisbane sample but in the same direction as the Shipibo sample.

Navhal Sample. In the Navhal sample, the BTT set was played with by boys ($M = 43.18$, $SD = 27.93$) more than girls ($M = 20.99$, $SD = 19.70$), and this difference was statistically significant, $t\ (76.99) = -4.16$, $p < .001$. The GTT set was played with by boys ($M = 22.49$, $SD = 22.43$) less than girls ($M = 39.59$, $SD = 27.92$), and this difference was statistically significant, $t\ (53.89) = 2.87$, $p = .006$. Play with the set of neutral toys was not significantly different between boys (*M* = 34.33, *SD* = 29.10) and girls (*M* = 39.42, *SD* = 32.63), *t* (58.52) = 0.71, *p* = .482.

Cross-Cultural Effects. When all samples were combined, boys played more with the BTT set than girls did *dc* = 1.67, *se* = 0.07, *p* < .001, and girls played more with the GTT set than boys did, *dc* = -1.55, *se* = 0.07, *p* < .001. There was no significant sex difference in children's play with neutral toys, *dc* = 0.04, *se* = 0.06, *p* = .472.

There was significant heterogeneity between cultures for the BTT set, *Q* (3) = 168.86, *p* < .001. Pairwise comparisons revealed that the sex difference in preference for BTT was smaller in the Brisbane sample (*d* = 2.65, *var*(*d*) = 0.11) than in the Shipibo sample (*d* = 2.87, *var*(*d*) = 0.10), *t* = -4.27, *p* < .001, but larger in the Brisbane sample than in the Lenakel sample (*d* = 1.04, *var*(*d*) = 0.09), *t* = 29.18, *p* < .001, and larger in the Brisbane sample than in the Navhal sample (*d* = 0.88, *var*(*d*) = 0.06), *t* = 35.18, *p* < .001.

There was significant heterogeneity between cultures for the GTT set, *Q* (3) = 193.46, *p* < .001. Pairwise comparisons revealed that the sex difference in preference for the GTT set was smaller in the Brisbane sample (*d* = -1.37, *var*(*d*) = 0.07) than in the Shipibo sample (*d* = -3.50, *var*(*d*) = 0.13), *t* = 44.85, *p* < .001, and smaller in the Brisbane sample than in the Lenakel sample (*d* = -2.00, *var*(*d*) = 0.13), *t* = 11.09, *p* < .001, but larger in the Brisbane sample than in the Navhal sample (*d* = -0.69, *var*(*d*) = 0.06), *t* = -17.32, *p* < .001.

There was significant heterogeneity between cultures for neutral toys, *Q* (3) = 123.20, *p* < .001. Pairwise comparisons revealed that the sex difference in preference for neutral toys was larger in the Brisbane sample (*d* = -0.87, *var*(*d*) = 0.06) than in the Shipibo sample (*d* = 0.58, *var*(*d*) = 0.05), *t* = -41.97, *p* < .001, and larger in the Brisbane sample than in the Lenakel sample (*d* = 0.71, *var*(*d*) = 0.08), *t* = -33.21, *p* < .001, and larger in the Brisbane sample than in the Navhal sample (*d* = -0.17, *var*(*d*) = 0.05), *t* = -18.98, *p* < .001. In terms of direction, the sex difference in the Brisbane and Navhal samples (girls played more than boys did with neutral toys) was opposite to the direction of the sex difference in the Shipibo and Lenakel samples (boys played more than girls did with neutral toys).

Sex-Typed Preferences

Brisbane Sample. In the Brisbane sample, boys played with the BTT set more than they played with the GTT set, and this difference was statistically significant, *t* (41) = 11.91, *p* < .001. Girls played with the GTT set more than they played with the BTT set, and this difference was statistically significant, *t* (37) = -3.82, *p* < .001.

Shipibo Sample. In the Shipibo sample, boys played with the BTT set more than they played with the GTT set, and this difference was statistically significant, t (42) = 17.27, p < .001. Girls played with the GTT set more than they played with the BTT set, and this difference was statistically significant, t (50) = -9.51, p < .001.

Lenakel Sample. In the Lenakel sample, boys played with the BTT set more than they played with the GTT set, and this difference was statistically significant, t (26) = 6.31, p < .001. Girls played with the GTT set more than they played with the BTT set, and this difference was statistically significant, t (25) = -2.89, p = .008.

Navhal Sample. In the Navhal sample, boys played with the BTT set more than they played with the GTT set, and this difference was statistically significant, t (48) = 3.49, p = .001. Girls played with the GTT set more than they played with the BTT set, and this difference was statistically significant, t (30) = -2.91, p = .007.

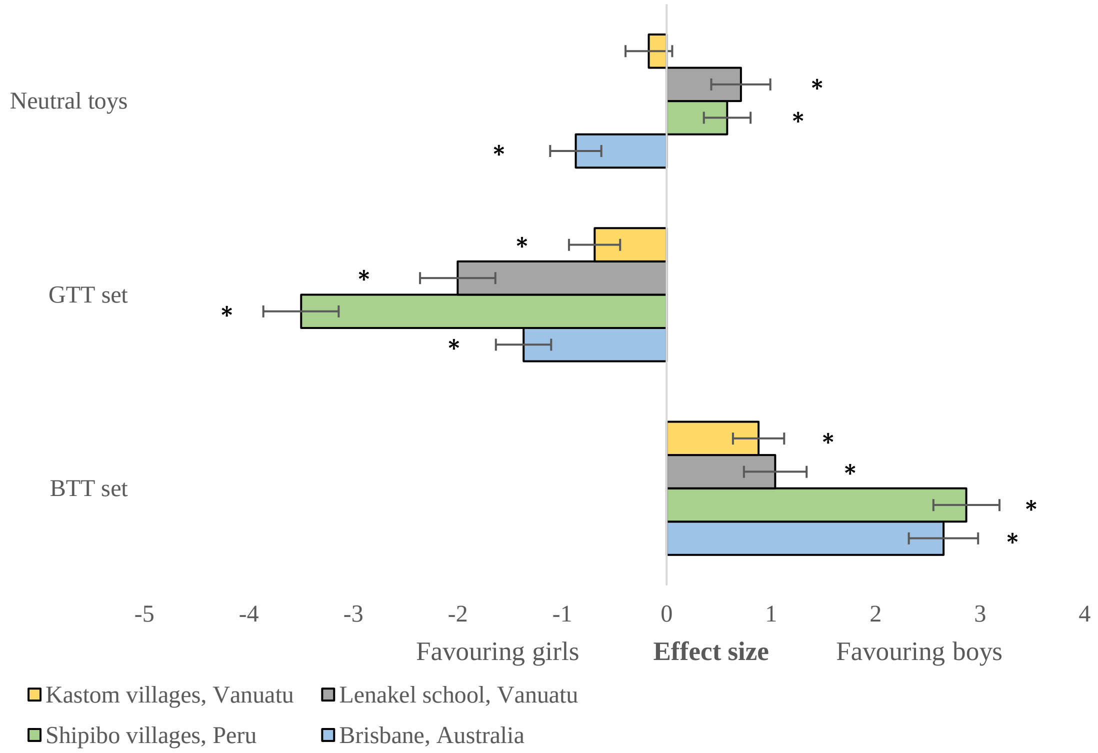


Figure 4.2. Sex differences in preferences for sex-typed toys across cultures. Bars show effect sizes: standardised mean differences between boys' and girls' proportions of play time with sex-typed toys. Error bars show standard errors of the effect size. Asterisks indicate statistically significant effects (at p < .05).

Doll and Vehicle

Effect sizes for the doll (baby doll) and vehicle (racing car) are summarised in Figure 4.3.

Brisbane Sample. In the Brisbane sample, the racing car was played with more by boys (*M* = 8.73, *SD* = 13.48) than by girls (*M* =0.85, *SD* = 2.31), and this difference was statistically significant *t* (43.65) = -3.73, *p* < .001. The baby doll was played with less by boys (*M* = 0.40, *SD* = 1.20) than by girls (*M* = 2.17, *SD* = 6.43), but this difference was not statistically significant, *t* (39.35) = 1.69, *p* = .099.

Boys played more with the racing car than they did with the baby doll, and this difference was statistically significant, *t* (41), = 4.04, *p* < .001. Girls played more with the baby doll than they did with the racing car, but this difference was not statistically significant, *t* (37) = -1.19, *p* = .242.

Shipibo Sample. In the Shipibo sample, the racing car was played with more by boys (*M* = 38.70, *SD* = 31.45) than by girls (*M* = 5.46, *SD* = 14.89), and this difference was statistically significant, *t* (57.64) = -6.36, *p* < .001. The baby doll was played with less by boys (*M* = 0.47, *SD* = 1.65) than by girls (*M* = 29.87, *SD* = 29.52), and this difference was statistically significant, *t* (50.37) = 7.10, *p* < .001.

Boys played more with the racing car than they did with the baby doll, and this difference was statistically significant, *t* (42) = 7.92, *p* < .001. Girls played more with the baby doll than they did with the racing car, and this difference was statistically significant, *t* (50) = -4.93, *p* < .001.

Lenakel Sample. In the Lenakel sample, the racing car was played with more by boys (*M* = 12.16, *SD* = 13.43) than by girls (*M* = 10.34, *SD* = 17.44), but this difference was not statistically significant, *t* (46.96) = -0.42, *p* =. 674.. The baby doll was played with less by boys (*M* = 2.13, *SD* = 6.19) than by girls (*M* = 3.17, *SD* = 4.01), but this difference was not statistically significant, *t* (44.78) = 0.73, *p* .467.

Boys played more with the racing car than they did with the baby doll, and this difference was statistically significant, *t* (26) = 3.37, *p* = .002. Girls played with the baby doll less than they played with the racing car, but this difference was not statistically significant, *t* (25) = 1.96, *p* = .061.

Navhal Sample. In the Navhal sample, the racing car was played with more by boys (*M* = 14.39, *SD* = 19.60) than by girls (*M* = 12.66, *SD* = 19.04), but this difference was not statistically significant, t (65.35) = -0.39, p = .698. The baby doll was played with less by boys (M = 6.26, SD = 13.53) than by girls (M = 10.93, SD = 19.80), but this difference was not statistically significant t (47.76) = 1.15, p = .255.

Boys played more with the racing car than they did with the baby doll, and this difference was statistically significant, t (48) = 2.21, p = .032. Girls played less with the baby doll than they did with the racing car, but this difference was not statistically significant, t (31) = 0.31, p = .755.

Cross-Cultural Effects. When all samples were combined, boys played more with the racing car than girls did dc = 0.62, se = 0.06, p < .001, and girls played more with the baby doll than boys did, dc = -0.57, se = 0.06, p < .001.

There was significant heterogeneity between cultures for the racing car, Q (3) = 88.57, p < .001. Pairwise comparisons revealed that the sex difference in preference for the racing car was smaller in the Brisbane sample (d = 0.80, $var(d)$ = 0.06) than in the Shipibo sample (d = 1.39, $var(d)$ = 0.06), t = -16.44 p < .001, but larger in the Brisbane sample than in the Lenakel sample (d = 0.12, $var(d)$ = 0.08), t = 14.75, p < .001, and larger in the Brisbane sample than in the Navhal sample (d = 0.09, $var(d)$ = 0.05), t = 19.23, p < .001.

There was significant heterogeneity between cultures for the baby doll, Q (3) = 66.19, p < .001. Pairwise comparisons revealed that the sex difference in preference for the baby doll was smaller in the Brisbane sample (d = -0.40, $var(d)$ = 0.05) than in the Shipibo sample (d = -1.35, $var(d)$ = 0.06), t = 27.03, p < .001, but larger in the Brisbane sample than in the Lenakel sample (d = -0.20, $var(d)$ = 0.08), t = -4.31, p < .001, and larger in the Brisbane sample than in the Navhal sample (d = -0.29, $var(d)$ = 0.05), t = -3.01, p = .003.

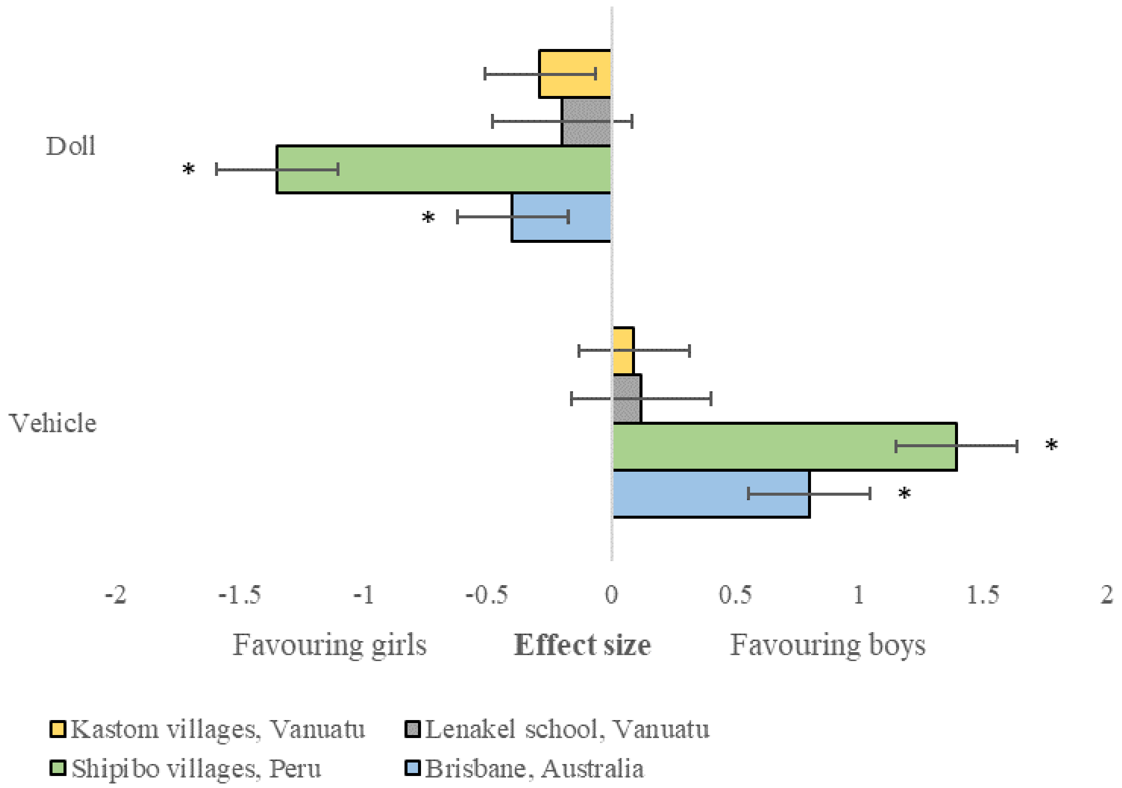


Figure 4.3. Sex differences in preference for doll (baby doll) and vehicle (racing car) across cultures. Bars show effect sizes: standardised mean differences between boys' and girls' proportions of play times with the doll and vehicle. Error bars show standard errors of the effect size. Asterisks indicate statistically significant effects (at $p < .05$).

Discussion

The current study presented a cross-cultural comparison of children's toy preferences, including in a cultural context where these toys were gender typed, and in cultural contexts where these toys were less familiar or completely novel. The results of the current study suggest that, broadly, sex differences in children's toy preferences can replicate in different cultures, including in cultures where these toys might be completely novel. This finding supports a culturally universal process underlying sex differences in children's toy preferences, and potentially, a culturally universal process underlying some sex differences in human behaviour more broadly.

Additionally, the results of the current study indicate that culture plays a substantial role in children's sex-typed toy preferences. The sizes of the sex differences in toy preferences were significantly different between cultures. Furthermore, the current study took toys that were sex-stereotyped in one culture, and transported these toys to a culture in which they were less familiar or novel. In two of the cultures where the toys were less familiar or completely novel (Lenakel and Navhal), children still showed sex differences in their toy preferences, but these sex differences were smaller than in the culture where the toys were sex-stereotyped. Additionally, sex differences were not found, in the most remote culture, for dolls or vehicles; in this culture, boys did not play with a toy vehicle more than girls did, and girls did not play with a baby doll more than boys did. These findings suggest that culture affects children's toy preferences, perhaps specifically via cultural stereotypes about toys.

Future theoretical work on sex and gender development may usefully consider the role of culture. Research on sex and gender development has often used toy preferences to indicate children's sex-typicality, because toy preferences show large sex differences. However, existing theories of sex and gender development do not make explicit predictions about how children's toy preferences might differ in cultures where the toys are unfamiliar. The current research found that sex differences in toy preferences can be partially replicated in cultures where the toys are unfamiliar, and therefore, that some aspects of theories of sex and gender development may also be generalisable across cultures. However, the results of the current study also suggest that cultural context may affect the sizes of the sex differences in children's toy preferences. It may therefore be useful for theories of sex and gender development to give more consideration to the role of culture, so that future theories may be able to address both culturally universal and culturally variable effects.

Study Limitations

In general, these findings should not be interpreted as evidence of children's sex typicality, or gender typicality, in different cultural contexts. The present study did not incorporate culturally relevant toys from each setting, but instead, used toys from the minority world. The goal of the present study was not to measure gender typicality in each culture, but instead, to measure whether culture plays a role in children's preference for these specific toys.

The present study used methods that were designed to be used in the same way across different cultural contexts, and to see if the procedures would reproduce the behavioural sex differences found in previous research in minority world cultures. The present study did not include protocols for gathering culturally specific information, such as gender stereotypes, or cultural information about the toys. For example, children were not asked to identify the gender stereotypes of the toys, in part because it was difficult to elicit information like this without implying a value judgement. Children in the Shipibo, Lenakel, and Navhal samples were not accustomed to participating in research or other regular, non-evaluative questioning. As a result, participants would have been likely to misinterpret researchers' requests for their opinions about the toys (e.g., *do you think this toy is for boys or for girls*) as evaluative tests of their knowledge about the toys (*do you know who this toy is for?*). An evaluative test of knowledge about foreign toys is likely to be stressful, and previous field experience indicated that both adults and children may be reluctant to answer questions when they felt they might not know the correct answer.

Additionally, the play situation may have had different meanings for children in Brisbane, who may have been accustomed to playing alone, compared to children in the other samples, who may be typically surrounded by siblings and other children. A relatively large number of children in the Navhal *kastom* villages were excluded from the study because they did not play with the toys for more than one minute. This low play time may have reflected a lack of comfort with the study setting, although this possibility was considered during study design and more children were recruited in the Navhal villages to account for it. This limitation, and others, are discussed more fully in the General Discussion chapter of the thesis.

Relevance to Theories of Sex and Gender Development

The main theories of sex and gender development do not explicitly predict how children's sex-typed toy preferences might vary in different cultural contexts, particularly where the toys are unfamiliar. However, it is possible to extrapolate some predictions about culturally universal and culturally variable processes from these theories.

The results of the present study suggest that some aspects of children's sex-typed preferences may be replicated in cultures where the toys are relatively novel, and therefore have little cultural context. Early hormone exposure could explain some preferences that are not dependent on cultural context. Boys and girls have different levels of prenatal and early postnatal exposure to hormones, particularly androgens, due to genetic differences encoded on the sex chromosomes (Hines, 2005). These early differences in hormone exposure may affect later sex-typed behaviour, as indicated by studies in non-human mammals (Arnold, 1985; Arnold, 2009; McCarthy, 2010) and in human children with genetic conditions that affect early hormone exposure (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Pasterski et al., 2005; Servin et al., 2003). Furthermore, comparative research indicates that sex-typed toy preferences may also exist in non-human primates (Alexander & Hines, 2002; Hassett et al., 2008). These converging lines of evidence suggest that differences between girls and boys in early hormone exposure may contribute to children's sex-typed toy preferences. The General Discussion chapter of this thesis integrates these results with the rest of the thesis, and discusses why sex differences might be evident in cultures where the toys were completely novel.

Additionally, the results of the present study indicate that culture influences children's toy preferences. Cultural variations in the social environment could influence children's toy preferences via processes described by social approaches to gender development. Previous research exploring the role of social learning has found that children's sex-typed preferences are influenced by reinforcement (Caldera et al., 1989; Fagot & Patterson, 1969; Idle et al., 1993), and by observing the preferences of others (Barkley, Ullman, Otto, & Brecht, 1977; Slaby & Frey, 1975; Wilansky-Traynor & Lobel, 2008). In the current study, the finding that children's sex differences in toy preferences were smaller in the most remote cultural group than in the other cultural groups could therefore be partially due to these children having no reinforcement history with sex-typed toys, and no previous observations of others playing with the toys.

Furthermore, cognitive processes may influence cultural variations in children's toy preferences. Children in remote cultures may not have encountered the toys before, and additionally, may not have encountered information about the toys' gender stereotypes through mass media or mass communication. If the children's gender schemas and gender stereotypes do not include any information about the toys, then the children would be expected to show reduced sex-typed preferences (Bem, 1981; Martin et al., 2002; Weisgram et al., 2014).

More generally, cultural context may influence children's toy preferences through an adaptive pressure to conform to, and seek status within, social groups in general. Naturalistic studies of children's toy requests suggest that sex differences in toy preferences tend to be greater for toys that are associated with eventual adult social status (Freeman et al., 1995; Richardson & Simpson, 1982). Children can seek status within any social group, but gender is a particularly salient social identity because in most cases it is physically marked, unchangeable, dichotomous, and ubiquitous (Duveen & Lloyd, 1986). Sex-typed toy preferences may therefore be partly related to children's general preferences for culturally relevant adult gender roles.

Conclusion

The present study found that sex differences in children's toy preferences were partially replicated in multiple cultures, including in a culture where these toys were completely novel. Furthermore, the present study found that culture influenced children's preferences for sex-typed toys. Future theories could usefully address the role of culture in sex and gender development.

Chapter 5

Children's Preferences for Toys with Faces, Propulsion, Dress-Up, and Role Play Toys in Four Cultures

Children's toy preferences are an extensively studied indicator of early sex-typed behaviour. Studies of children's sex-typed toy preferences typically include a range of toys, because different toys may allow for different types of play. Play affordance is the range of play behaviours that are afforded by a toy (Mori, Nakamoto, Mizuochi, Ikudome, & Gabbard, 2013). For example, a toy vehicle affords propulsion: it can be vigorously moved around in space, partly due to its moving wheels. In contrast, a toy doll does not afford propulsion: it has no wheels, and so is not intended to move around in space in the same way.

Sex differences in children's toy preferences may be partly due to sex differences in children's preferences for play affordances. For example, toy vehicles are typically preferred by boys, and dolls are typically preferred by girls. The types of play afforded by toy vehicles and by dolls are different; toy vehicles can be raced around with propulsive motion, while dolls have faces and therefore may afford social play. Some theorists have proposed that play affordances may be one of the key functional features of a toy that makes it preferred by boys or by girls (e.g., Alexander, 2003; Benenson, Liroff, Pascal, & Cioppa, 1997; Connellan, Baron-Cohen, Wheelwright, Batki, & Ahluwalia, 2000). That is, sex differences in children's preferences for certain types of play may be partly responsible for sex differences in children's toy preferences. For example, a male preference for propulsion may be partly responsible for boys' preference for toy vehicles, whereas a female preference for social play may be partly responsible for girls' preferences for dolls. Similar reasoning also has been used to explain proposed differences between girls and boys in preferences for mechanical toys and toys with faces (Connellan et al., 2000).

Background Literature

Studies of sex-typed toy preferences typically include toys with propulsive motion affordances, particularly toy vehicles, as boy-type toys, and toys with social affordances, particularly dolls, as girl-type toys. Indeed, toy vehicles and dolls are the most commonly used toys in studies of sex-typed toy preferences. Free play studies of toy preferences typically include at least one vehicle as a boys' toy, and at least one doll as a girls' toy (e.g., Berenbaum & Hines, 1992; Pasterski et al., 2005; Serbin, Connor, Burchardt, & Citron, 1979), and typically find sex differences in children's toy preferences.

However, studies do not always agree that propulsion and social play are the affordances that might drive boys' and girls' preferences for sex-typed toys. Another perspective was provided in an early study by Rekers and Yates (1976). They separated toys into "maternal nurturance" toys (a baby doll and a Barbie doll) and "masculine aggression" toys (dart guns, a rubber knife, plastic handcuffs, and a set of cowboys and Indians). Rekers and Yates also predicted that children's gender role behaviour more broadly could be predicted from their play with these categories of toys. More recently, Connellan and colleagues (Connellan et al., 2000) conceptualized girl-preferred toys as having faces, and boy-preferred toys as being mechanical. To examine their perspective, they tested whether day-old infants fixated more on an adult face, or on a moving mobile. Female infants fixated more on the face than on the mobile, while male infants fixated more on the mobile than on the face. The authors concluded that female infants had a general preference for faces, while male infants had a general preference for mechanical motion. Similarly, Escudero and colleagues (Escudero et al., 2013) tested whether 6-month-old infants fixated more on pictures of faces, or on pictures of cars, and expected to see a female preference for faces and a male preference for cars, but their results did not support this preference.

In most studies of sex-typed toy preferences, it is difficult to separate the play affordances from the gender stereotypes of the toys. Children can identify the gender stereotypes of toys, and prefer to play with toys that they know are stereotyped for their gender (Stern & Karraker, 1989; Weisgram et al., 2014; Zosuls et al., 2009). Therefore, in a test of whether girls and boys prefer toys with different play affordances, it may be useful to include toys that have similar affordances but different gender stereotypes.

The present study aims to explore girls' and boys' preferences for toys with different play affordances, in different cultures. Although previous research is suggestive of certain patterns, such as a male preference for propulsion and a female preference for faces, there is not enough previous research to make confident predictions about the direction of effects. The present study therefore takes an exploratory approach.

Method

This chapter is a reanalysis of data reported in Chapter 4 of this thesis.

Materials

The present study aimed to account for the gender stereotypes of the toys, by including a girl-type toy and a boy-type toy for each affordance. The two toys with faces were: a Hulk action figure (boy-type toy) and a Barbie doll (girl-type toy). The two propulsion toys were: a racing car (boy-type toy) and a pony carriage (girl-type toy). The two dress-up toys were: a pirate costume (boy-type toy) and a princess costume (girl-type toy). The two role play toys were: a sword (boy-type toy) and a baby doll (girl-type toy). In addition to these eight toys, there were four neutral toys: a dinosaur costume, a set of plastic African animals, pencils and paper, and a stuffed dog.

Procedure

Toy preferences were measured in a free play session. The procedure has been described previously (Chapter 3). In brief, children were videotaped individually in a single 5-minute play session. Children's physical contact with the toys was recorded and converted into a score representing the proportion of the child's total play time that was spent contacting each toy. Statistical methods were identical to those used in Chapter 3.

Results

Effect sizes for sex differences in children's preferences for toys with faces, propulsion, dress-up, and role play affordances are summarised in Figure 5.1.

Toys with Faces

In the Brisbane sample, toys with faces were played with more by boys (*M* = 53.87, *SD* = 36.62) than by girls (*M* = 29.54, *SD* = 30.87), and this difference was statistically significant, *t* (77.63) = 3.22, *p* = .002. In the Shipibo sample, toys with faces were played with more by boys (*M* = 34.05, *SD* = 31.59) than by girls (*M* = 33.83, *SD* = 29.83), but this difference was not statistically significant, *t* (87.39) = 0.04, *p* = .972. In the Lenakel sample, toys with faces were played with less by boys ($M = 40.38$, $SD = 27.96$) than by girls ($M = 43.46$, $SD = 27.74$), but this difference was not statistically significant, t (50.65) = -0.42, $p = .673$. In the Navhal sample, toys with faces were played with more by boys ($M = 26.08$, $SD = 27.37$) than by girls ($M = 19.63$, $SD = 23.51$), but this difference was not statistically significant, t (70.92) = 1.12, $p = .266$.

When all samples were combined, boys played more with toys with faces than girls did, $dc = 0.22$, $se = 0.05$, $p < .001$. There was also significant heterogeneity between cultures for toys with faces, Q (3) = 31.98, $p < .001$. Pairwise comparisons revealed that the sex difference in preference for toys with faces was larger in the Brisbane sample ($d = 0.71$, $var(d) = 0.05$) than in the Shipibo sample ($d = 0.01$, $var(d) = 0.04$), $t = 21.01$, $p < .001$, and larger in the Brisbane sample than in the Lenakel sample ($d = -0.12$, $var(d) = 0.08$), $t = 1.81$, $p < .001$, and larger in the Brisbane sample than in the Navhal sample ($d = 0.25$, $var(d) = 0.05$), $t = 12.73$, $p < .001$.

Propulsion

In the Brisbane sample, propulsion toys were played with more by boys ($M = 10.64$, $SD = 13.97$) than by girls ($M = 7.03$, $SD = 14.12$), but this difference was not statistically significant, t (77.03) = 1.15, $p = .255$. In the Shipibo sample, propulsion toys were played with more by boys ($M = 42.16$, $SD = 31.47$) than by girls ($M = 16.58$, $SD = 21.39$), and this difference was statistically significant, t (71.92) = 4.52, $p < .001$. In the Lenakel sample, propulsion toys were played with less by boys ($M = 18.43$, $SD = 16.94$) than by girls ($M = 18.66$, $SD = 20.78$), but this difference was not statistically significant, t (48.24) = -0.04, $p = .966$. In the Navhal sample, propulsion toys were played with more by boys ($M = 24.46$, $SD = 26.56$) than by girls ($M = 18.42$, $SD = 25.10$), but this difference was not statistically significant, t (66.65) = 1.03, $p = .309$.

When all samples were combined, boys played more with propulsion toys than girls did, $dc = 0.40$, $se = 0.06$, $p < .001$. There was also significant heterogeneity between cultures for propulsion toys, Q (3) = 43.49, $p < .001$. Pairwise comparisons revealed that the sex difference in preference for propulsion toys was smaller in the Brisbane sample ($d = 0.26$, $var(d) = 0.05$) than in the Shipibo sample ($d = 0.97$, $var(d) = 0.05$), $t = -20.80$, $p < .001$, but larger in the Brisbane sample than in the Lenakel sample ($d = -0.01$, $var(d) = 0.08$), $t = 5.93$, $p < .001$, and not statistically different in the Brisbane sample and in the Navhal sample ($d = 0.23$, $var(d) = 0.05$), $t = 0.69$, $p = .492$.

Dress-up

In the Brisbane sample, dress-up toys were played with less by boys (*M* = 15.37, *SD* = 23.39) than by girls (*M* = 19.66, *SD* = 23.97), but this difference was not statistically significant, *t* (76.78) = -0.81, *p* = .421. In the Shipibo sample, dress-up toys were played with less by boys (*M* = 1.17, *SD* = 2.14) than by girls (*M* = 9.86, *SD* = 10.70), and this difference was statistically significant, *t* (32.45) = -4.32, *p* < .001. In the Lenakel sample, dress-up toys were played with less by boys (*M* = 4.21, *SD* = 5.78) than by girls (*M* = 17.46, *SD* = 14.56), and this difference was statistically significant, *t* (32.45) = -4.32, *p* < .001. In the Navhal sample, dress-up toys were played with less by boys (*M* = 4.81, *SD* = 7.80) than by girls (*M* = 11.09, *SD* = 17.67), but this difference was not statistically significant, *t* (37.50) = -1.87, *p* = .070.

When all samples were combined, girls played more than boys did with dress-up toys, *dc* = -0.68, *se* = 0.06, *p* < .001. There was also significant heterogeneity between cultures for dress-up toys, *Q* (3) = 52.48, *p* < .001. Pairwise comparisons revealed that the sex difference in preference for dress-up toys was smaller in the Brisbane sample (*d* = -0.18, *var*(*d*) = 0.05) than in the Shipibo sample (*d* = -1.08, *var*(*d*) = 0.05), *t* = 26.30, *p* < .001, and smaller in the Brisbane sample than in the Lenakel sample (*d* = -1.20, *var*(*d*) = 0.09), *t* = 20.85, *p* < .001, and smaller in the Brisbane sample than in the Navhal sample (*d* = -0.50, *var*(*d*) = 0.05), *t* = 8.80, *p* < .001.

Role play

In the Brisbane sample, role play toys were played with more by boys (*M* = 6.54, *SD* = 10.17) than by girls (*M* = 4.45, *SD* = 9.72), but this difference was not statistically significant, *t* (77.75) = 0.94, *p* = .349. In the Shipibo sample, role play toys were played with less by boys (*M* = 3.65, *SD* = 7.48) than by girls (*M* = 30.33, *SD* = 29.53), and this difference was statistically significant, *t* (57.50) = -6.22, *p* < .001. In the Lenakel sample, role play toys were played with more by boys (*M* = 7.82, *SD* = 9.35) than by girls (*M* = 6.29, *SD* = 6.14), but this difference was not statistically significant, *t* (45.11) = 0.71, *p* = .484. In the Navhal sample, role play toys were played with less by boys (*M* = 10.32, *SD* = 14.71) than by girls (*M* = 11.45, *SD* = 19.90), but this difference was not statistically significant, *t* (50.55) = -0.27, *p* = .787.

When all samples were combined, girls played more than boys did with role play toys, *dc* = -0.24, *se* = 0.06, *p* < .001. There was also significant heterogeneity between cultures for role play toys, *Q* (3) = 108.66, *p* < .001. Pairwise comparisons revealed that the sex difference in preference for role play toys was smaller in the Brisbane sample ($d = 0.21$, $var(d) = 0.05$) than in the Shipibo sample ($d = -1.19$, $var(d) = 0.05$), $t = 40.52$, $p < .001$, but not significantly different in the Brisbane sample and in the Lenakel sample ($d = 0.19$, $var(d) = 0.08$), $t = 0.40$, $p = .690$, and larger in the Brisbane sample than in the Navhal sample ($d = -0.07$, $var(d) = 0.05$), $t = 7.72$, $p < .001$.

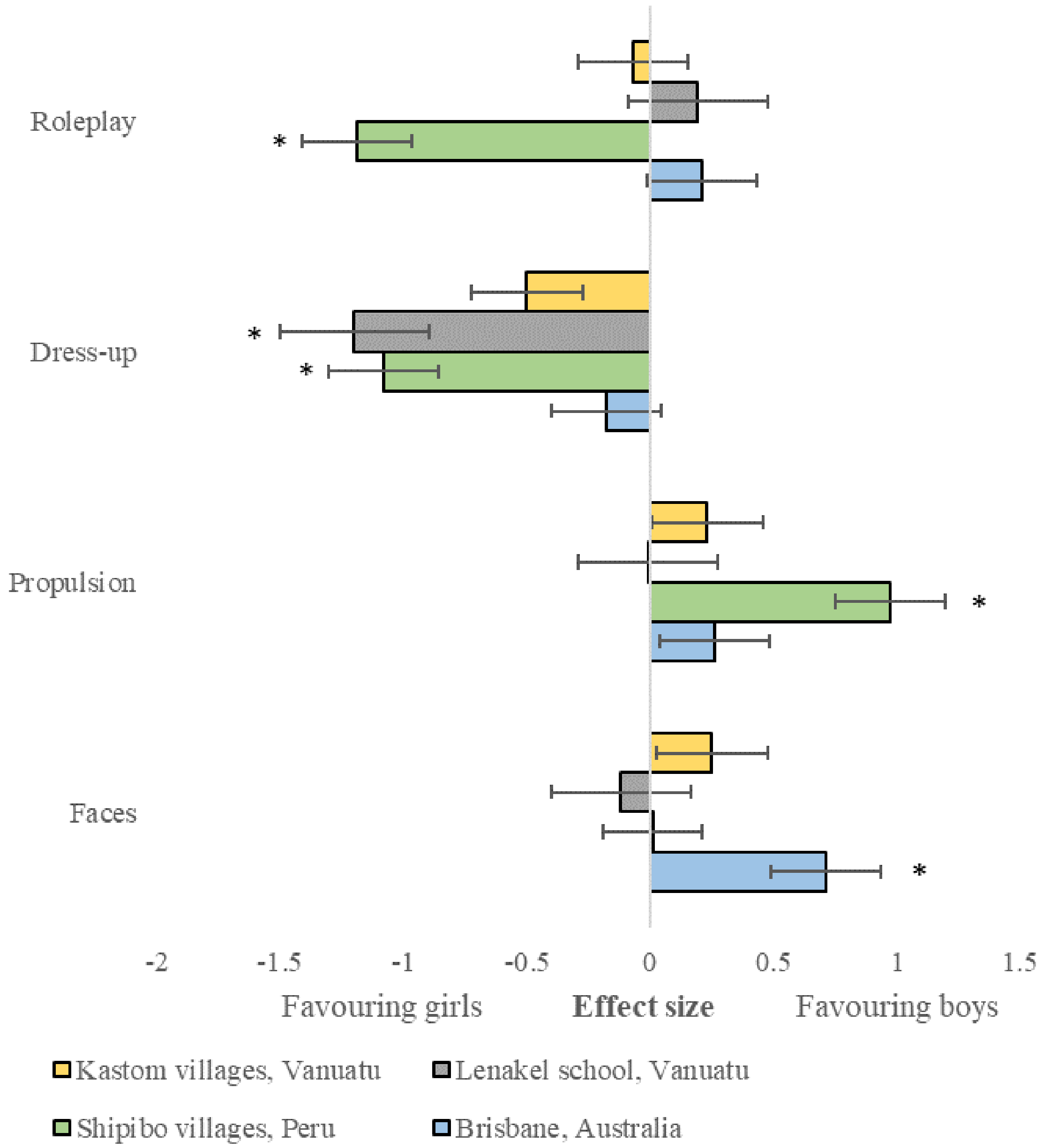


Figure 5.1. Sex differences in preference for role play, dress-up, propulsion, and toys with faces, across four cultures. Bars show effect sizes: standardised mean differences between boys' and girls' proportions of play time with affordance pairs. Error bars show standard errors of the effect size. Asterisks indicate statistically significant effects (at $p < .05$).

Discussion

The current study explored sex differences in children's preferences for toys with different play affordances, such as toys with faces, propulsion, dress-up, and role play affordances. In all four cultures, boys played with propulsion toys more than girls did, and girls played with dress-up toys more than boys did. Sex differences also appeared in children's preferences for toys with faces and for role play toys, but in some cultures, the sex difference favoured girls, while in other cultures, the sex difference favoured boys. Additionally, the results indicated that culture may play a role in children's preferences for toys with different play affordances.

Summary of Results

Toys with Faces. The present study found gender differences in children's play with toys with faces, but in some cultures the gender difference favoured boys, while in other cultures the gender difference favoured girls. The face toys used in the present study included a boy-type toy with a face (an action figure) as well as a girl-type toy with a face (a Barbie doll). In contrast to the results of the present study, researchers have previously asserted that sex differences in children's preferences for dolls may be explained by a female preference for faces (Connellan et al., 2000; Escudero et al., 2013). However, these previous studies typically include only female-type toys with faces (dolls), and do not include male-type toys with faces. The few studies that have included a boy-type toy with a face, such as an action figure or toy soldiers, have found a male preference for these toys (Cherney & London, 2006). The theory that girls prefer dolls due to a female preference for faces may therefore require some revision.

Propulsion. Boys played more with propulsion toys than girls did in each culture, although the sex difference did not reach statistical significance in every culture. The propulsion toys used in the present study included a boy-type propulsion toy (a racing car) as well as a girl-type propulsion toy (a pony carriage). Consistent with the present study, some previous studies have also found a male preference for toys with propulsion affordances in infants (Alexander, Wilcox, & Woods, 2009; Jadva et al., 2010) and in non-human primates (Alexander & Hines, 2002), even where the propulsion toys included toys that were stereotyped for girls as well as toys that were stereotyped for boys (Hassett et al., 2008). However, in contrast to the present study, a previous report did not find a male preference for toys with propulsion affordances, where the available propulsion toys included toys that were stereotyped for girls as well as toys that were stereotyped for boys (Dinella et al., 2016). The present study found a male preference for propulsion toys in cultures where the toys were novel, and therefore did not carry stereotypes.

Dress-up. Girls played more with dress-up toys than boys did in each culture, although the sex difference did not reach statistical significance in every culture. One possible explanation is that girls may have felt comfortable playing dress-up with both the boy-type costume (the pirate costume) and the girl-type costume (the princess costume), while the boys may only have felt comfortable playing dress-up with the boy-type costume. Previous research has found that teachers and peers give significantly more criticism to boys than to girls for engaging in cross-gender dress-up (Fagot, 1977), and therefore, boys may have avoided playing with the princess costume due to past negative experiences with girl-type costumes. However, the present study also included cultural contexts where the princess costume and pirate costume were novel, and therefore, where boys would not have had previous experiences with these toys. In these cultures, girls still played with dress-up toys more than boys did.

Role play. The present study found that in some cultures, girls played more with role play toys than boys did, while in other cultures, boys played more with role play toys than girls did. The present study included a baby doll as a girl-type role play toy, and a sword as a boy-type role play toy. Previous studies of role play have suggested that girls may engage in more role play than boys, either because they own more role play toys (Pomerleau et al., 1990) or because more role play toys are available to represent adult female roles than male roles (Jones & Glenn, 1990). The Six Cultures study also found that girls engaged in role play more than boys did (Edwards, 2000), although their analysis was not focused on toys. The present study also found cross-cultural variation in children's preferences for sex-typed role play toys. It is possible that these toys were not interesting across all cultural contexts; for example, the toy sword may have been considered a poor representation of a grown man's machete by boys in the Lenakel and Navhal samples, who were accustomed to wielding real machetes themselves.

Possible Alternative Explanations

Toy affordances were not consistently found to influence sex differences in children's toy preferences in each culture. The null findings could be due to limitations in the study design (as discussed below). An alternative explanation is that affordances other than propulsion, faces, dress-up, or roleplay may be more important. One affordance that was not explored in the present study, but that may be important to sex differences in development, is the aggressive and competitive play afforded by the toy. The association of boy-type toys with aggression and competition has been demonstrated in minority world cultures (e.g., Blakemore & Centers, 2005; Miller, 1987), and boys may use toys to practice aggression and competition (Hellendoorn & Harinck, 1997; Humphreys & Smith, 1987). In majority world cultures, then, boys may use toys for practice contests, especially in cultural environments were aggression and competition are important parts of adult masculinity. The General Discussion chapter further discusses this possibility, integrated with the results of other thesis chapters.

Limitations

The present study included only two toys for each play affordance, and only one example of each sex-typed affordance toy. For example, propulsion was afforded by only one boy-type toy (the racing car) and one girl-type toy (the pony carriage). The use of only a single toy for each category introduces the possibility that specific features of the toys, other than affordances, may have been affecting children's preferences. These features may have been inherent to the design of the toys: for example, the racing car may have afforded better propulsion than the pony carriage, due to its more streamlined design and lack of a passenger.

Additionally, features of the toys, other than their play affordances, may have interacted with elements of the cultural context. For example, the purple feathers on the princess costume may have been attractive to both boys and girls in the Navhal-speaking *kastom* villages, because of a similarity to the brightly painted feathers that are used in ceremonies, such as traditional marriage celebrations (Farran, 2004). Also, as noted above, the toy sword may have been unappealing to the boys in the Lenakel and Navhal cultures, because of their use of real machetes.

Conclusion

The present study explored the children's preferences for toys with faces, propulsion toys, dress-up toys, and roleplay toys, in four cultures. Toys with faces, and role play toys, were preferred more by boys in some cultures, and more by girls in others. Propulsion toys showed a sex difference favouring boys in each culture, but this was not always significant. Dress-up toys showed a sex difference favouring girls in each culture, but this was not always significant. These results suggest that further research is required to discover how play affordances might affect children’s toy preferences. Future work could include variations on the current methodology, such as including a greater variety of toys with the same affordances.

Chapter 6

A Machine Learning Model of Sex, Culture, and Affordance in Children's Toy Choices

Theories about sex-typed toy preferences, and gender-related behaviour more broadly, acknowledge that these are probably determined by some combination of genetic sex, social and cognitive processes that could relate to culture, and features of the toy or behaviour (e.g., Bandura & Bussey, 2004; Hines, 2013; Liben & Bigler, 2002; Pasterski et al., 2005). However, empirical studies of toy preferences do not typically include measurements of more than one or two possible influences. The cross-cultural study of children's toy preferences (Chapters 2, 3, and 4 of this thesis) found that sex, cultural context, and the play affordances of toys, may have each independently affected children's toy preferences. The present study re-analyses this data to discover how sex, cultural context, and play affordance might interact.

Theorists have identified complex and dynamic systems theory as a way of integrating explanations for children's sex-typed behaviour (e.g., Hines, 2013; Martin & Ruble, 2010). However, systems theory has had limited practical application so far, partly because of a perceived need for large datasets with which to parameterise systems models, and partly because there are few worked examples available to demonstrate how systems models could be used in studies of sex and gender development (e.g., Didonato, England, Martin, & Amazeen, 2013). This chapter aims to demonstrate how a systems model can be applied to toy play data, of the type that is commonly collected in free play studies of children's sex-typed toy preferences. Specifically, this chapter demonstrates a machine learning model, applied to the cross-cultural dataset of children's toy choices.

Background Literature

Sex-related development involves a multidimensional system based on genetic information coded in sex chromosomes, and it is partially dependent on aspects of the developmental environment, especially prenatal and early postnatal exposure to androgenic hormones. Early hormone environments in humans have been linked to sex-typed behaviour, including toy preferences, in studies of children exposed to atypical early hormone environments (Berenbaum & Hines, 1992; Nordenström et al., 2002; Pasterski et al., 2005;

Servin et al., 2003), and in studies of typically developing children (Hines et al., 2002; Knickmeyer et al., 2005; Lamminmäki et al., 2012; but see Constantinescu & Hines, 2012). These hormonal processes are mapped to biological sex because they depend primarily on genetic information coded on the sex chromosomes.

Additionally, cultural context influences behaviours and preferences, especially through social and cognitive mechanisms. Children observe sex differences in behaviour and reproduce them (Barkley et al., 1977), and are likely to be reinforced for reproducing sex-typed behaviour (Fagot, 1978; Fagot & Patterson, 1969; Langlois & Downs, 1980). Children also remember information about sex-typed behaviour, and retain new information about sex and gender based on how well this new information matches dominant cultural norms (Atkin, 1975; Stangor & Ruble, 1989). These processes map to cultural context because they depend primarily on cultural definitions of sex-appropriate behaviour and preferences.

Finally, the function of the sex-typed behaviour plays a role. Toy affordances are the range of functional play behaviours that are afforded by a toy; for example, a toy vehicle affords the function of propulsive motion, and a toy doll affords the function of social play (Mori et al., 2013). Children’s sex-typed preferences for certain toys are often described in terms of toy affordance, and sex differences in preference for propulsion, faces, dress-up, or role play may explain sex differences in preference for toys that afford these types of play, such as toy vehicles, dolls, costumes, or roleplay props (Benenson et al., 1997, 2011; Connellan et al., 2000; Dinella et al., 2016; Freeman, 2007; Pomerleau et al., 1990), at least in some cultures.

Overall, evidence supports the separate influences of sex, cultural context, and toy affordance on children’s toy preferences. However, the complex interactions of these factors have not been empirically described in previous research. Complex interactions between predictive factors can be described using systems models, such as regression-based models. Theorists have previously identified systems models as potentially useful to theories of sex-typed behaviours, including toy preferences (e.g., Hines, 2013; Martin & Ruble, 2010), but systems models have not yet been applied to toy choice data.

Traditional statistical approaches may miss valuable theoretical insights hidden in complex datasets. As an example, the toy preference dataset, collected for this thesis, includes multiple predictor variables: child gender, toy sex-type, cultural context, and toy affordance. These predictor variables each affect the dependent variable, toy preference. But the predictor variables are not independent, and their interactions may be very complex. For example, the relationship between child gender and toy preference may depend on toy affordance, but only in some cultural contexts, and only for GTT. Traditional statistical approaches are not well-suited to uncovering complex patterns like this. Omnibus tests, such as ANOVA, provide a single statistical test that may mask underlying interactions where the effects are in different directions. Mean comparisons, such as *t*-tests, would require a very large number of comparisons to test all possible interactions, thereby inflating the potential for false positives. Standard regression modelling would require a large number of high-order interaction terms, making interpretation difficult. Dimension reduction techniques, such as discriminant analysis, may be used to identify the most important variables in a complex dataset, but would have the same challenges as regression models in addressing the large number of possible interactions. Thus, traditional statistical analyses could fail to capitalise on the potential for complex datasets to answer questions of theoretical importance.

In contrast, there are statistical methods, based on systems theory, that are designed to make statistical inferences from complex datasets. This chapter introduces boosted regression trees (BRTs) and classification and regression trees (CARTs) as methods for analysing complex toy choice data. The purpose of this chapter is to demonstrate the use of BRTs and CARTs for analysing interactions in the toy play dataset; and, additionally, to discover what additional insights might be offered by using these techniques, compared with standard approaches.

Method

Dataset: Cross-Cultural Study of Toy Preferences

Chapter 3 of this thesis provided details of the design and conduct of the study on which the current analysis is based. In brief, the study measured free play in boys and girls across four cultural contexts, with a design that included toys with different play affordances. Participants were recruited from four cultural contexts: a large industrialised city in Brisbane, Australia (N = 42 boys, 51 girls); a set of remote villages of Shipibo people in the Lake Imiria region of the Peruvian Amazon Basin (N = 43 boys, 51 girls); a school population in a town on Tanna Island, Vanuatu (N = 27 boys, 26 girls); and a set of kastom villages on Tanna Island (N = 49 boys, 31 girls).

Children were given access to twelve toys for a free play period of five minutes. The twelve toys represented three categories of possible sex-typed toy choices: boy-type toys

(BTT), girl-type toys (GTT), and neutral toys. The BTT set included a toy car, an action figure, a pirate costume, and a sword. The GTT set included a pony carriage, a Barbie doll, a princess costume, and a baby doll. The neutral set included pencils and paper, a dinosaur costume, a stuffed dog, and toy African animals. Toys were selected in matched pairs for play affordance within the sex-typed groups, so that children had an option of a BTT or a GTT for each play affordance. Matched affordance pairs were: propulsion toys (a toy car and a pony carriage); toys with faces (an action figure and a Barbie doll); dress-up toys (a pirate costume and a princess costume); and roleplay toys (a sword and a baby doll). Neutral toys were not included in the affordance pairs design.

The play session was recorded on video and coded later. The dependent variable for the current paper – toy choice – was operationally defined as a single event representing a child's selection of a toy for play. A child was scored as "choosing" a toy when he or she touched it for the first time. For this reason, the results section of this paper refers to toy *choices*, rather than toy *preferences*, because the word *preference* has been used previously to reflect the average duration of time playing with a toy (as in Chapter 3). If a child put down the toy, and then picked it up again later in the play session, a second choice of that toy was recorded. The focus on toy choices allowed the systems model to represent ecologically valid aspects of the play session, such as children having a variable number of toy choices, switching between toys, and playing with more than one toy at a time. Children who did not play with any toys over the free play period were not included in the analyses.

The final toy choice dataset contained a total of 3448 observed toy choices over play sessions from 344 children, giving an average of 10.02 toy choices per child in the 5-minute play session. One toy choice represented one observation, or one data point, in the dataset. The toy choices could take one of three values: play with BTT (B), play with GTT (G), or play with neutral toys (N). Each toy choice also had a time, t, at which the toy choice was observed. For example, if a child picked up the toy car, five seconds after the start of the play session, the data point representing this observation would be a toy choice B, with time $t = 5$.

In the context of systems theory, recording each toy choice had three advantages over recording play duration. First, children make multiple toy choices in a single play session, and therefore the total number of observations is greater for toy choices than for play duration. This is useful because systems models often require a large number of data points, in order to correctly describe the data. Additionally, in contrast to statistical methods typically used for play duration, such as t-tests or regression, the machine learning model does not assume independence of observations, and therefore can include more than one observation per child. Second, the type of statistical model used in the current paper (see below) is designed for use with observations of single events (like toy choices), rather than for use with aggregates or averages (like mean play times) (Elith, Leathwick, & Hastie, 2008). Finally, using toy choices allowed the model to represent ecologically valid aspects of the play session, such as children having a variable number of toy choices, switching between toys, and playing with more than one toy at a time.

The final dataset for the analysis of children's choices of sex-typed toys contained a total of 3448 observed toy choices over play sessions from 344 children, giving an average of 10.02 toy choices per child in the 5-minute play session. One toy choice represented one observation, or one data point, in the dataset. The toy choices could take one of two values: play with BTT (*B*), or play with GTT (*G*)[3].

Statistical Methods

Complex interactions between child features (sex), environment features (cultural context), and toy features (affordance), were modelled using boosted regression trees (BRT) and classification and regression trees (CART). In a CART, the response variable (in this case, choice of a BTT or GTT) is described by a cascading series of binary splits of the explanatory variables. This process is represented as a tree-like structure, with the final nodes representing subsets of the responses (the toy choices). The tree is built by an iterative process: first, the model identifies the single variable that best splits the data into two groups. The data are separated, and then this splitting process is applied separately to each subgroup, and so on recursively until a predefined stopping rule is reached, such as a minimum subgroup size, or until the splits are no longer statistically improving the model's predictive performance. The selection of variables, the placement of the variables in the tree, and the location of the binary splits are all determined statistically by the model and based on the dataset.

A boosted regression tree (BRT) is a sequential collection of CARTs, in which the misclassified observations from the first CART are collected and fitted with a second CART, and so on, until again a pre-defined stopping rule is reached. This combination of CART and boosting improves predictive performance, especially for small groups that may otherwise be

[3] To test the full model including sex, cultural context, and toy affordance, the analysis focused on children's choices of sex-typed toys (BTT or GTT). This focus was necessary because neutral-type toys were not included in the matched affordances design, and so choices of neutral-type toys could not be estimated in terms of toy affordance.

misclassified. The two approaches together facilitate strong behavioural inference, since BRTs are optimised for predictive performance and CARTs are more easily interpretable. BRTs and CARTs allow for multiple complex interactions between predictor variables, and do not assume independent observations or linear relationships. The statistical analysis focused on a single dependent variable, choice of BTT or GTT. The dependent variable was operationalised as a binary continuous variable, ranging from 0 (BTT) to 1 (GTT). The model included three explanatory variables: (i) sex of the child, which was a binary variable ranging from 0 (male) to 1 (female); (ii) cultural setting, which was a categorical factor with four levels: Brisbane, Shipibo, Lenakel, or Navhal; and (iii) toy affordance, which was a categorical factor with four levels: toys with faces, propulsion, dress-up, and role play toys.

The BRT models were fit to the data using a gradient boosted model with Bernoulli response and 5000 trees, an interaction depth of 3 (allowing for 3-way interactions), and a stopping rule based on predictive performance as evaluated by 2-fold cross-validation, implemented in the R statistical software package “dismo”. The CART models were fit to the data using a maximum depth of 10 and 10-fold cross-validation, implemented in the R statistical software package “rpart”. The fit of the models was evaluated using estimates of deviance and correct classification of predicted compared with observed responses.

Results

Predictors of Toy Choice

The BRT model correctly classified 70.27% of toy choices (BTT or GTT). Based on the BRT model, the primary indicator of toy choices in a free play session was sex, followed by cultural context and then toy affordance (Figure 6.1).

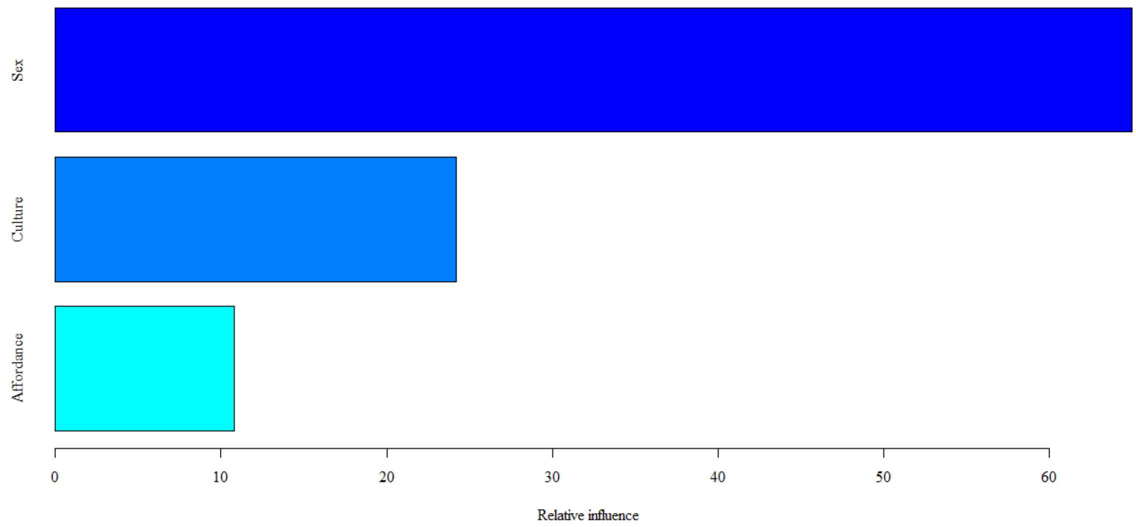


Figure 6.1. Relative influence of sex, cultural group, and toy affordance on choice of sex-typed toys across the sample of 3448 toy choices in 344 children.

As seen in Figures 5.1 and 5.2, the BRT revealed differential effects of sex, cultural context, and toy affordance on children's toy choices, expressed as log odds of choosing a GTT. Sex was the most important predictor of children's toy choice (Figure 6.1), with a relative importance of 65%, and girls were predicted to be more likely to choose a GTT than boys were (Figure 6.2). Culture was the second most important predictor (Figure 6.1), with a relative importance of 24.2%, and children from Navhal-speaking *kastom* villages in Vanuatu were predicted to be most likely to choose a GTT (Figure 6.2). Affordance was the least important predictor (Figure 6.1), with a relative importance of 10.8%, and GTT with dress-up and roleplay affordances were predicted to be more likely to be selected than GTT with figure and propulsion affordances (Figure 6.2). These marginal relationships are not tested (for statistical significance) in the BRT: instead, the next step is to examine interactions between sex, culture, and affordance, in the CART.

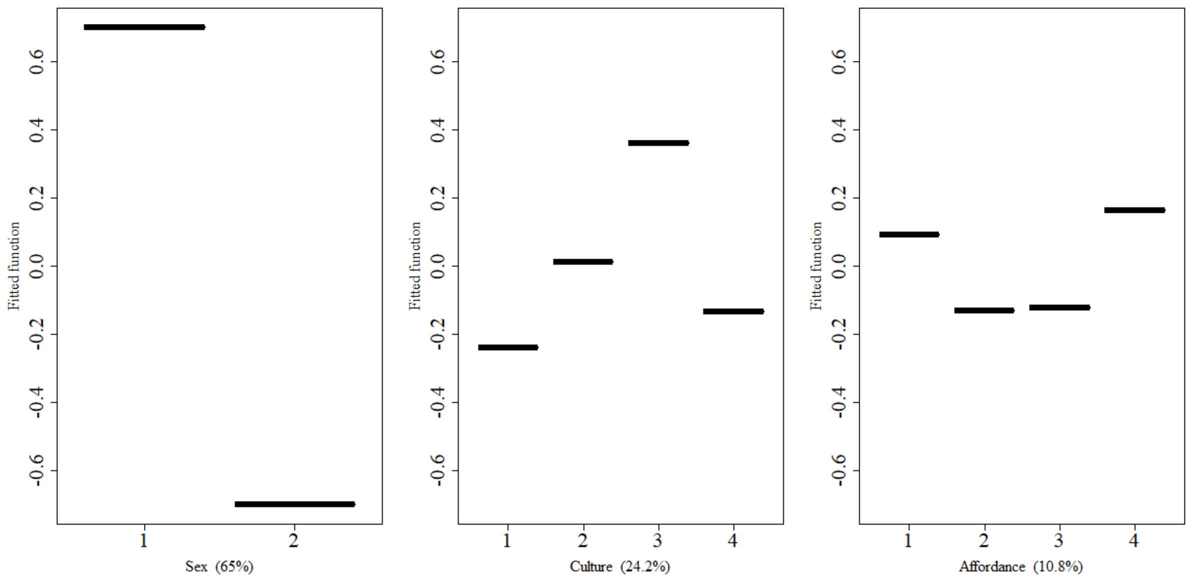


Figure 6.2. Predicted nonlinear marginal relationships between predictor variables and sex-typed toy choices in the BRT. Vertical axis is log-odds of choosing a female-type toy. Sex (1) = female, (2) = male. Culture (1) = Brisbane, Australia, (2) = Lenakel school, Vanuatu, (3) = Kastom villages, Vanuatu, (4) = Shipibo villages, Peru. Affordance (1) = dress-up, (2) = toys with faces, (3) = propulsion, (4) = roleplay. The number in parentheses is the relative importance of the variable in the BRT model.

Interactions between Sex, Culture, and Affordance

The CART analysis revealed the interactions between sex, culture, and toy affordance in the dataset. Figure 6.3 displays the final tree produced by the CART. The tree is interpreted in more detail below.

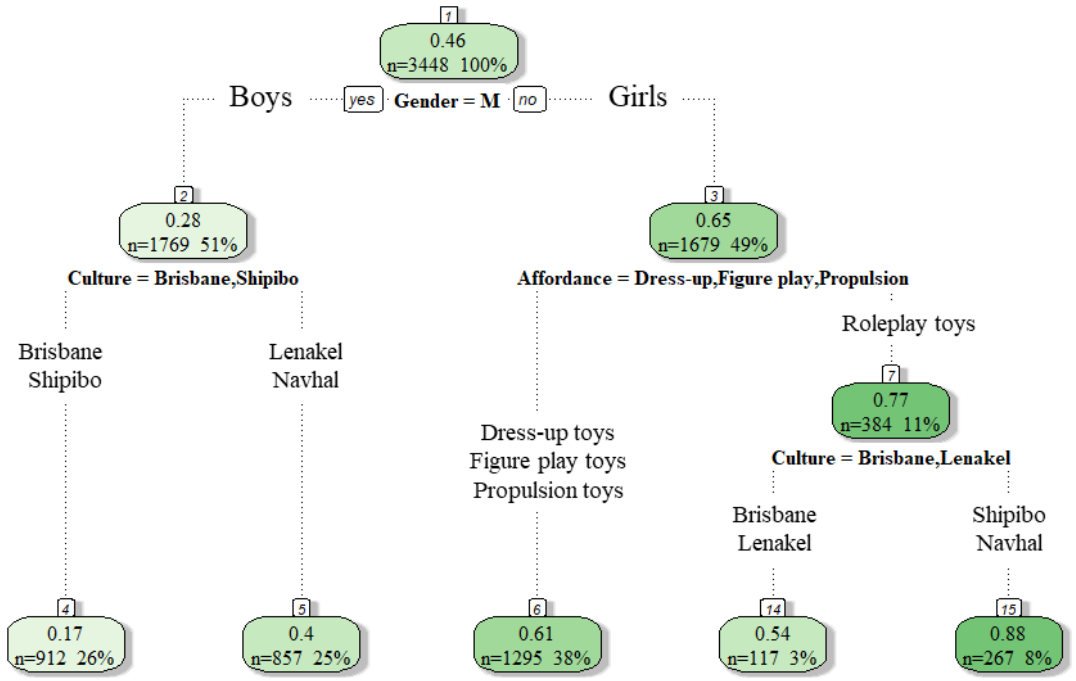


Figure 6.3. CART for sex-typed toy choices. Splits are selected from the data by a statistical algorithm that optimises for prediction of final states: choice of BTT or GTT. Nodes for the CART's classification decisions are presented from the top down, with the most important predictors at the top of the tree, and the final classified states at the bottom of the tree. The "yes" condition is always on the left. The first number in the node represents the probability of choosing a GTT over a BTT, for all the toy choices that were classified within that node. n is the number of toy choices that were classified into that node, and the percentage is n as a percentage of the total dataset. For example, toy choices in boys from either Brisbane or Shipibo samples are represented by node 4, they make up 912 observations or 26% of the total dataset, and their predicted toy choice probability is 0.17, representing a 17% chance of choosing a GTT over a BTT.

The CART analysis, like the BRT analysis, suggested that sex was the primary indicator of children's toy choices in the free play session. As seen in Figure 6.3, the CART model selected sex as the first node on which to split the data. In addition, in boys, sex-typed toy choices were predicted by culture. Boys in the samples taken from Brisbane, Australia and Shipibo villages in Peru were less likely (17%) to choose a GTT over a BTT than boys in the samples taken from schools or *kastom* villages in Vanuatu (40%). In boys, toy affordance did not influence toy choices. In girls, however, sex-typed toy choices were predicted first by toy affordance, and then by culture. Girls were likely to choose a GTT over a BTT if it afforded role play, and if they lived in rural villages in either Navhal-speaking *kastom* regions of Tanna Island, Vanuatu or in Shipibo villages in Peru (88%). Girls in Brisbane and in Lenakel, however, were only slightly more likely to choose a GTT than a BTT with role play affordance (54%). For toys with faces, propulsion toys, and dress-up toys, girls' sex-typed toy choices did not depend on cultural context, and girls were more likely to choose a GTT than a BTT (61%).

Groups of Children Likely to Make Sex-Typed and Non-Sex-Typed Toy Choices

Under no conditions did the model predict a high likelihood (>50%) of boys choosing a GTT over a BTT. Boys were least likely to choose a GTT over a BTT when the boys were from Brisbane, Australia, or from Shipibo villages in Peru (17%). Boys were most likely to choose a GTT over a BTT when the boys were from Navhal-speaking *kastom* villages, or attending school in Lenakel (40%), but boys were still likely to choose a BTT over a GTT in these conditions.

The model predicted that girls were most likely to choose a GTT over a BTT when the toy had a role play affordance, and when the girls were from Shipibo villages in Peru or from *kastom* villages in Tanna, Vanuatu (88%). Girls were least likely to choose a GTT over a BTT when the toy had a roleplay affordance, and when the girls were from the school in Lenakel town, Vanuatu or Brisbane, Australia (54%), but girls were still likely to choose a GTT over a BTT in these conditions, although to a lesser extent than boys.

Discussion

The primary goal of the present study was to demonstrate how a machine learning systems model (BRT and CART) could be applied to a dataset of children's toy choices. The systems model correctly predicted, based on sex, culture, and toy affordance, 70 per cent of children's toy choices. The systems model produced several insights about the dataset that would not have been produced by standard analysis methods, such as *t* tests or regression models. These insights relate to the relative importance of different predictors of toy choice; the interactions between different levels of these predictors; and which groups of children are most likely to choose sex-typed toys and non-sex-typed toys.

Predictors of Toy Choice

Sex appeared to be an important predictor of children's sex-typed toy choices, and more important than cultural context or toy affordance. The BRT model selected sex as the most important predictor, based entirely on the observed pattern of children's toy choices in the data. This result was despite the BRT model not having been instructed or coerced to select sex over cultural context or toy affordance.

The effect of sex, however, was not simply a sex difference in toy preferences. According to the model, sex also affected the interaction between culture and toy affordance, and their impact on toy choice. These interactions are discussed in the following section.

Interactions between Sex, Culture, and Affordance

The CART model found that sex affected children's toy preferences not only directly, but also indirectly, through sex differences in the effects of culture and toy affordance. In other words, boys and girls showed different patterns of toy choices in response to variations in cultural context and toy affordance.

Boys' sex-typed toy choices appeared to be influenced more by cultural context than by toy affordance. Specifically, boys showed different toy choices in Brisbane and Shipibo samples than in Navhal and Lenakel samples. Boys in the combined Brisbane-Shipibo group were more likely to choose BTT over GTT than were boys in the combined Navhal-Lenakel group.

In contrast, girls' sex-typed toy choices appeared to be influenced more by toy affordance than by cultural context. For toys with faces, propulsion toys, and dress-up toys, girls were likely to choose a GTT over a BTT, regardless of cultural context. For roleplay toys, however, girls showed different choices in Shipibo and Navhal samples than in Brisbane and Lenakel samples. Girls in the combined Shipibo-Navhal group were more likely to choose a GTT roleplay toy over a BTT roleplay toy than were girls in the combined Brisbane-Lenakel group. The GTT roleplay toy was a baby doll, and the BTT roleplay toy was a sword. One possible explanation for this effect may be that the baby doll represents an adult activity that is salient for girls in the Shipibo villages in Peru, and girls in the Navhal *kastom* villages in Vanuatu, who are commonly responsible for caring for younger siblings. Alternatively, girls in Brisbane, Australia, and girls attending school in Lenakel, might have less experience caring for real children than girls in the Navhal and Shipibo cultures, and might therefore see the baby doll as less important for adult roleplay.

According to the analysis presented in this chapter, roleplay toys were particularly important to children's cross-gender play. This finding supports the results of earlier naturalistic studies of gender differences in children's play. In a reanalysis of data from the Six Cultures study, Edwards (2000) found that boys and girls differed in their roleplay activities, based on culturally relevant adult gender roles, and that girls engaged in roleplay more than boys did. The results of the current chapter provide further support for Edwards' conclusions, particularly that culturally-appropriate roleplay may be important for girls. These insights align with prior theory that suggests toys function as symbolic representations of adult gender roles (e.g., Richardson & Simpson, 1982; Serbin, Poulin-Dubois, Colburne, Sen, & Eichstedt, 2001).

Groups of Children Likely to Make Sex-Typed and Non-Sex-Typed Toy Choices

Regardless of culture or toy affordance, no groups of boys were likely to choose a girl-type toy over a boy-type toy. The most likely group of children to make a non-sex-typed toy choice was girls playing with role play toys in Brisbane, Australia, or in Lenakel, Vanuatu. These findings are similar to those of previous research, suggesting that boys avoid cross-gender toys more than girls do (Lobel & Menashri, 1993; Wong & Hines, 2014). Additionally, the systems model builds on previous research, by suggesting that girls' interest in cross-gender toys may be most relevant to role play toys. In the current dataset, girls were sometimes likely to role play in male roles, using the sword. In general, the results of the systems model suggest that role play may be a useful direction for future research on children's sex-typed toy preferences.

Limitations

The model was parameterised with a single cross-cultural dataset of children's sex-typed toy preferences. Results of the model are only applicable within the parameters of this dataset, until they can be tested on other large cross-cultural datasets. Adding more toys, or more cultures, to the dataset may change the results of the model. Furthermore, many covariates that are theoretically important to toy preferences – for example, parental opinions about the toys, or children's early hormone environments – were not measured in the dataset and are not captured in the model. Future work could measure additional theoretically important covariates, and these could be included in the machine learning model.

Additionally, the substantive findings of this chapter are subject to the same limitations as previous chapters, due to study design. The analysis in this chapter is based on a dataset collected during a brief, solo play session with a limited selection of toys. Results of this analysis should be interpreted in this context. In future, this method could be expanded to larger datasets and more generalised inferences.

Conclusions

Systems models, including those trained by machine learning algorithms, can account for complex interactions between predictor variables in behavioural studies. The current study demonstrated how systems models, specifically BRTs and CARTs, could be applied to a cross-cultural dataset of children's toy choices in free play. Results revealed that children's toy choices depended on complex interactions between children's sex, cultural context, and toy affordance. Sex was the most important predictor of children's toy preferences. The model predicted that boys' toy choices depended on cultural context but not on toy affordance, but that girls' toy choices depended on cultural context and on toy affordance. Girls were likely to engage in same-gender and cross-gender roleplay, while boys were not. The machine learning algorithm integrated sex, cultural context, and toy affordance together in a complex system. This method allowed for extra insights into the data, over and above the results of previous approaches that tested each predictor separately (Chapters 3 and 4 of this thesis). Complex and dynamic systems methods, like the approach demonstrated in the present study, have the potential to integrate multiple explanations of children's sex-typed behaviour in a single model.

Chapter 7

A Multistate Dynamic Model of Sex and Culture in Children's Toy Choices

Children's play, including toy choices, may be regarded as an example of a behaviour that results from a complex and dynamic system. According to complex systems theory, any observed behaviour is the product of a system of underlying processes. It may not be possible to isolate and observe these underlying processes directly, but they can be described mathematically. Additionally, the underlying processes may be numerous and inter-related, and therefore, they may produce a variety of observed behaviours in the larger system, including behaviours that may seem random or complicated. This complexity property means that even random-seeming or complicated behaviours may be empirically describable, using simple rules (Bar-Yam, 1997).

Complex systems have been used to analyse phenomena that are highly stochastic, highly variable, or difficult to predict, from the action of vaccines (Haken & Jumarie, 2006), to shifts in the global economy e, for example. Systems approaches are increasingly influential in developmental psychology, particularly in studies of social development (van Geert, 2011) and learning (Tenenbaum et al., 2006). An integrated systems-based perspective on developmental science (Overton & Lerner, 2014), has shifted recent research on child development away from two-dimensional Cartesian-based thinking, and toward a more systems-based approach. The current chapter aims to extend some of this systems-based thinking to research on toy preferences.

Dynamic systems, also called dynamical systems, are the subset of complex systems that deal with change over time. The system is described in terms of mathematical rules that include time explicitly, rather than focusing on one time point, or averaging over a time period. This chapter aims to demonstrate how a dynamic systems model can be applied to toy play data, of the type that is commonly collected in free play studies of children's sex-typed toy preferences. Specifically, this chapter demonstrates a theory-driven transitional Markov model, applied to the cross-cultural dataset of children's toy choices.

Background Literature

Complexity and Dynamics in Children's Toy Choices

Children's toy choices can be described as a complex system. Consider a child, choosing a toy from a set of available options. The child's toy choice is the product of a set of underlying processes; fundamental processes that could apply to any type of choice, processes that might depend on the child's sex, cognitive and social processes that might depend on the child's culture, or any number of other possible processes. These underlying processes interact in a complex system that can be recursive, nonlinear, and stochastic. Each underlying process cannot be isolated and observed, because the removal of any one process would change the entire system. All that we can observe is the final product of these underlying processes: the system behaviour, that emerges as the child making a toy choice. We cannot directly observe the underlying complex system of processes that emerge as the child's toy choice, but we can theorise about this system, using mathematical logic.

Studies of children's toy choices typically average the amount of time that children spend playing with a toy, over a play session. However, children's play, including toy choices, can change over time. When total play times are averaged across an entire session, some of the information about children's toy choices may be lost.

Researchers have identified this problem, and have proposed theoretical applications of dynamic systems to sex-typed behaviours, including toy choices (DiDonato et al., 2013; DiDonato et al., 2012; Fausto-Sterling et al., 2012; Martin & Ruble, 2010; Yoshikawa & Hsueh, 2001). However, dynamic systems have not yet been empirically applied to sex-typed toy choices.

Standard approaches, that average children's toy choices over a single time frame, may mask behavioural dynamics that are important to theory. For example, to explain why some research finds that boys are more strongly sex-typed in their toy choices than girls are, some theorists have suggested that boys may intentionally avoid GTT, as part of a more general avoidance of anything related to female gender (e.g., Banerjee & Lintern, 2000; Golombok et al., 2008; Huston, 1985; Serbin et al., 1979). But an alternative explanation might be that boys try playing with GTT and dislike them, and consequently discard them. Current statistical methods, based on averaging toy preference over time, cannot distinguish these two explanations. A dynamic model, however, may be able to distinguish between children's toy choices based on avoidance, and toy choices based on dislike. The following section provides a worked example, applied to the toy preference data, to demonstrate how a dynamic systems model might distinguish the two.

A Worked Example: A Multistate Markov Model

The current chapter demonstrates a possible method for applying dynamic systems analysis to empirical research on sex-typed toy choices. The primary goal of the chapter is to provide a worked example of how a dynamic systems model could, in practice, be applied to a dataset of sex-typed toy choices. As a secondary goal, the chapter aims to provide a substantive interpretation of the dynamic systems model results.

The dynamic systems model demonstrated in this chapter is a multistate transitional state-space model: specifically, a Markov model. Markov models are mathematical models that deal specifically with change over time. Consequently, Markov models have been widely used in behavioural ecology, to describe the dynamics of animal behaviour (e.g., Franke, Caelli, & Hudson, 2004; Patterson, Basson, Bravington, & Gunn, 2009; Patterson, Thomas, Wilcox, Ovaskainen, & Matthiopoulos, 2008; Schliehe-Diecks, Kappeler, & Langrock, 2012). A Markov model may, therefore, also be suitable for research on the dynamics of human behaviour, including sex-typed toy choices. A Markov model can handle complex bidirectional relationships, allowing children to change from any one category of toys, to any other, at any time. Therefore, it can flexibly represent the dynamics of a free play session, such as children switching from one toy to another toy, and then back again. A Markov model can also account for covariates, such as the sex of the child, or the culture of the study, and therefore a Markov model may be appropriate for a cross-cultural toy choice dataset.

Applying the Markov model to the toy choice dataset follows two steps. In step 1, children's toy choices in a free play session are represented as a set of transitional states over time. In step 2, this representation is adjusted for covariates, such as sex and culture. The Method section explains these two steps further.

Method

Dataset: Cross-Cultural Study of Toy Choices

The methods used to construct the dataset are identical to Chapter 5 and are not repeated here.

The final toy choice dataset contained a total of 3448 observed toy choices over play sessions from 344 children, giving an average of 10.02 toy choices per child in the 5-minute play session. One toy choice represented one observation, or one data point, in the dataset.

The toy choices could take one of three values: play with BTT (*B*), play with GTT (*G*), or play with neutral toys (*N*). Each toy choice also had a time, *t*, at which the toy choice was observed. For example, if a child picked up the toy car, five seconds after the start of the play session, the data point representing this observation would be a toy choice *B*, with time $t = 5$.

The following sections describe the implementation of the Markov model, with reference to the toy choice dataset. The Methods sections describe the two theoretical steps of the Markov model. The Results section describes how the Markov model is applied to the toy choice dataset, using the R statistical software package "msm" for multi-state Markov models (Jackson, 2011).

Step 1. Toy Choice as Transitional Dynamic States in a Markov Model

First, toy choice is represented as a set of possible states. There are three possible states: play with BTT (B), play with GTT (G), and play with neutral toys (N). This set of possible states is referred to as the model *state space*. At any time in the free play session, children are assumed to be in one of these three states[4]. Children can also change states: for example, they can change from play with a BTT, to play with a neutral toy. These changes in play state have an attached probability: theoretically, children might be more likely to change between some toys than others. The process of changing states is referred to as a *transition*, and the probability of changing states is referred to as a *transition probability*. The data are structured as a matrix of possible states, with a probability assigned to each transition in the state space. This *transition matrix* is the foundation of a Markov model.

Specifying the initial state space. The state space of the play session can be described as a set of states, transitions, and transition probabilities. The state space can be represented by a 3-dimensional transition intensity matrix: 3-dimensional because there are 3 possible states, and transition intensity because each position in the matrix represents the size (intensity) of a transition probability.

An annotated version of the initial state space matrix is given in Figure 7.1, and is explained below.

[4] For parsimony, play with no toy is not included as a possible state in the current demonstration, but the Markov model could be extended to include it.

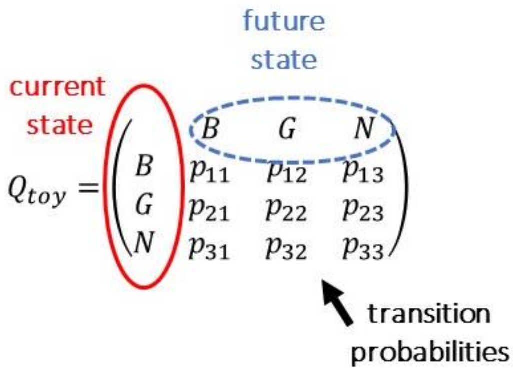


Figure 7.1. Annotated state space matrix, showing the initial state space and transition probabilities of the play session.

As seen in Figure 7.1, the state space matrix shows the possible states, and the transition probabilities, of the play session. The possible states are play with BTT (*B*), play with GTT (*G*) and play with neutral toys (*N*). The transition probabilities are represented by the *p* values in each cell of the matrix, and each probability has a subscript to indicate its position in the matrix. Each transition probability represents the probability of switching to the state on the column, given that the child is presently in the state on the row. For example, p_{21} is located in the second row, and in the first column, and represents the probability of switching to a BTT next, given that the child is currently playing with a GTT. This initial transition matrix is a theoretical representation of the starting state space of the play session, and it does not yet contain any data.

System state vectors. The transition matrix is theoretical, and it gives the *predicted* probabilities for transitioning between each of the three states. Empirically, however, each *observed* toy choice can only represent one state. If we take a picture of a child playing, we do not observe a distribution of probabilities; we observe that the child is playing with BTT, GTT, or neutral toys. In mathematical terms, this picture is the *system state* at time *t*.

The system state at time *t* is represented as a column vector (a one-dimensional matrix) with 3 possible states – play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). An annotated version of the system state vector is given in Figure 7.2.

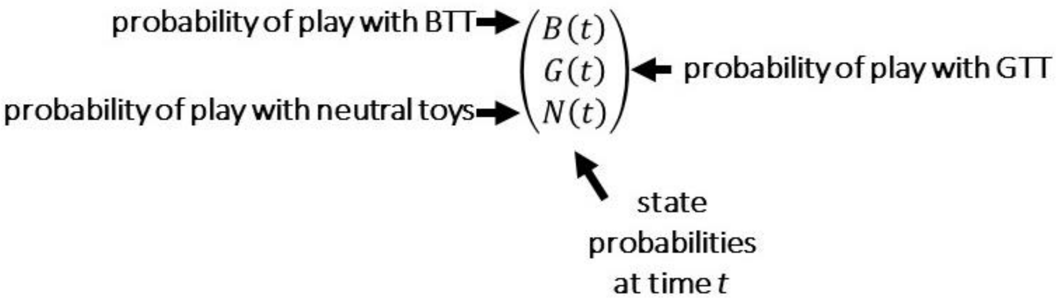


Figure 7.2. Annotated system state vector, showing the possible toy choices at time *t*.

The vector values (Figure 7.2) represent the predicted probability that a system is in state *B* (play with BTT), state *G* (play with GTT), or state *N* (play with neutral toys) at time *t*. When the system is observed – for example, when we take a picture of the child playing – the probabilities of each state become either 100% (the system is in that state), or 0% (the system is not in that state). These are represented in the system state vector as values of 0 and 1, as shown in Figure 7.3.

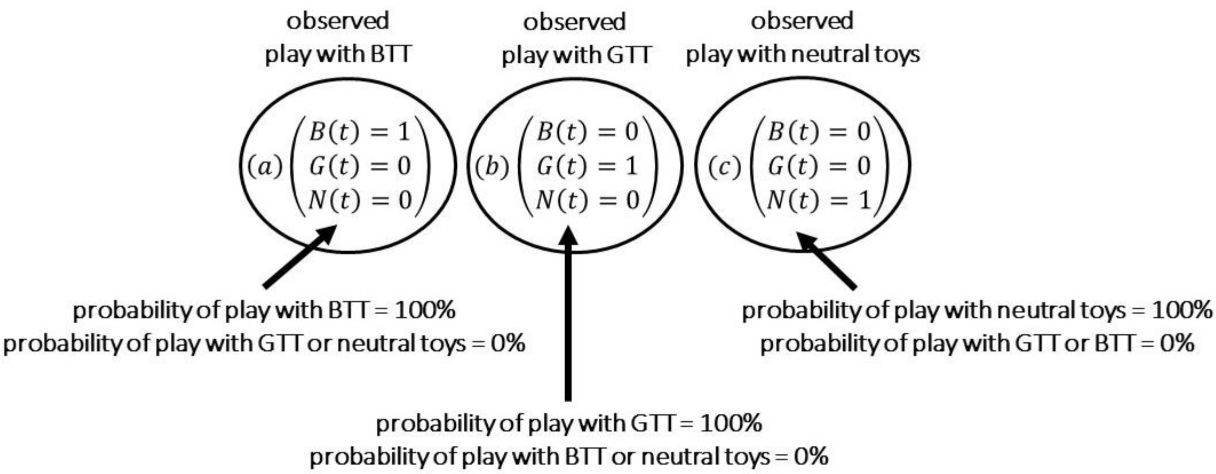


Figure 7.3. Annotated system state vectors for the observed toy choices: play with BTT (a), play with GTT (b), and play with neutral toys (c).

Vectors (a), (b), and (c), in Figure 7.3, are examples of observed system state vectors. These vectors represent observed system states that correspond to play with BTT (a), play with GTT (b), and play with neutral toys (c), at time *t*. For example, if the child is observed playing with BTT at time t = 30 seconds, then the probability of play with BTT at t = 30 seconds is 100%, and the probability of play with GTT and neutral toys at t = 30 seconds is 0. This observation is reflected in the system state vector values, which in this example for t = 30 seconds will have a value of (B (30) = 1, G (30) = 0, N (30) = 0), or vector (a).

Thus, each time a child makes a toy choice, the data can be represented by a system state vector: play with BTT (vector a), play with GTT (vector b), or play with neutral toys (vector c), and the time *t* at which the toy choice happened. Every toy choice in the dataset can be represented by these system state vectors.

Transition matrices and stable states. Next, the system state vectors (the toy choices) in the dataset are used to calculate the probabilities in the transition matrix. The play session can be thought of as a series of points in continuous time, and ranges from the beginning of the play session (t = 0) to the end of the play session ($t = t_i$). For each time point (t), the model multiplies the observed state space vector for the previous time point (t – 1) by the transition matrix, and it updates the transition matrix probabilities.

The initial transition matrix, representing a hypothetical starting space, is iteratively multiplied with the observed system state vectors over the time points, to give a probability of switching to each state. These probabilities are recorded in a transition matrix that has the same dimensions as the initial state space matrix, but it has transition probabilities (the *p* values) that are based on the data.

For example, to calculate the transition matrix for the third time point (t = 3), the model would multiply the system state vector for the second time point (t = 2) by the 3-dimensional transition matrix. In the toy choice example, the child's toy choice at the third time point (t = 3) would be predicted by multiplying the child's toy choice at the second time point (t = 2) by the probability of switching toys. The probability of switching toys would be updated, to include the data from the child's toy choice at the second time point (t = 2). The model would then proceed to the next time point: it would predict the child's toy choice at the fourth time point (t = 4) by multiplying the child's toy choice at the third time point (t = 3) by the updated probability of switching toys. This process is represented in Figure 7.4.

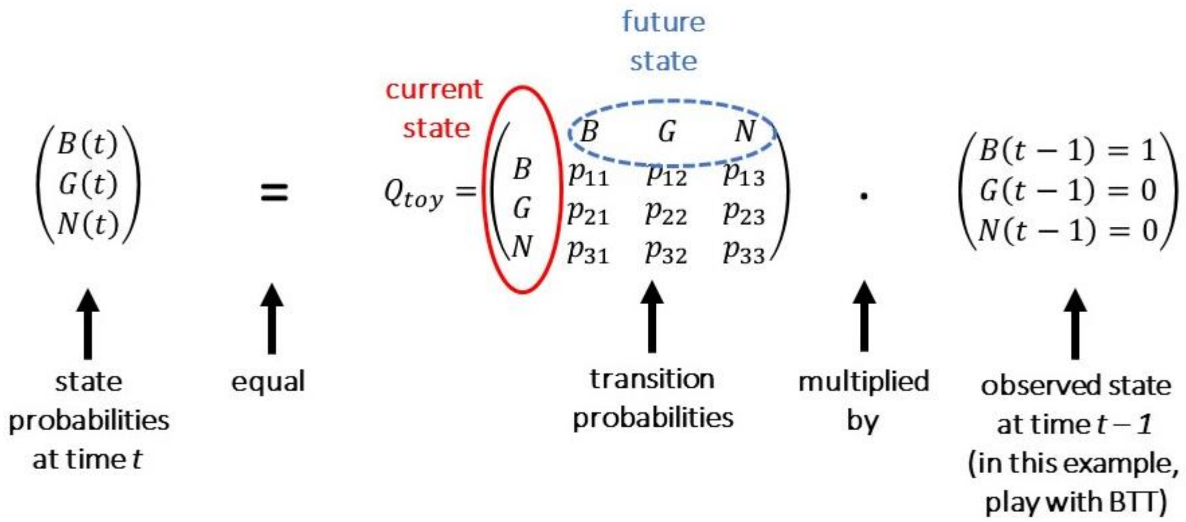


Figure 7.4. Dynamic process of updating the transition matrices. The transition probabilities at each time point are iteratively multiplied by the system state at the preceding time point.

As seen in Figure 7.4, the transition matrix has three possible states: play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). The values $B(t)$, $G(t)$, and $N(t)$ represent the observed state of the play session at time *t*. These are iteratively calculated by multiplying the observed state of the play session at time *t*-1 (the time immediately preceding time *t*) with the probability of switching states. The probability of switching states is taken from the three-dimensional transition matrix, which is then updated to reflect the total probability of switching states in the next iteration, given the total number of switches between each of the states in the play session so far.

Eventually, these iterations converge on a set of values that represent a stable state of the system; that is, the proportion of time that children spend with each category of toys in a hypothetical equilibrium. This *stable state* does not represent an average state of the system (like mean play time), but instead, is a transition matrix that represents the dynamic system of children's toy choices.

Step 2. Environmental Covariates: Sex and Culture

In step 2, covariates are added into the model, to test how the dynamic system of children's toy choices might be predicted by sex and culture. Mathematically, the transition matrix is modelled as a function of the covariates using proportional hazards methods (Marshall & Jones, 1995). The proportional hazards model can be used to calculate transition matrices that are adjusted for different levels of the covariates. These *covariate-adjusted transition matrices* give the probability of switching states, given specific levels of the covariates; for example, the probability of switching from play with BTT to play with GTT, given that a child is a boy from Brisbane.

Model selection. In analyses focused on mean time, the statistical significance of covariates would be tested, for example using *t*-tests. In contrast, in the Markov model, the covariates can be tested using *model selection*, based on the model "fit" to the data. The model "fit" is a statistical description of how closely the values predicted by the mathematical model match the values observed in the data. Covariates are tested using nested models: that is, the fit of the model including the covariate is compared to the fit of the model without the covariate. If the model, that includes the covariates, fits the data better than a null model, that does not include the covariates, then the model that includes the covariates is selected.

Covariate matrices. In analyses focused on mean time, the effect of covariates would be estimated, for example using regression coefficients. In contrast, in the Markov model, the effect of covariates is also estimated; but for each covariate, instead of providing a single estimate, the Markov model produces a matrix of the same dimensions as the transition matrix. In these *covariate matrices*, the numbers (the *p* values) do not represent the transition probabilities (the probability of switching from one toy to another). Instead, the numbers in the covariate matrices are hazard ratios that represent the effect of the covariate on each probability. Hazard ratios may be interpreted as the change in the transition probability for a one-unit change in the covariate (for continuous covariates), or as the difference in the transition probability between the covariate level and the baseline category (for factor covariates). An annotated example is given in Figure 7.5.

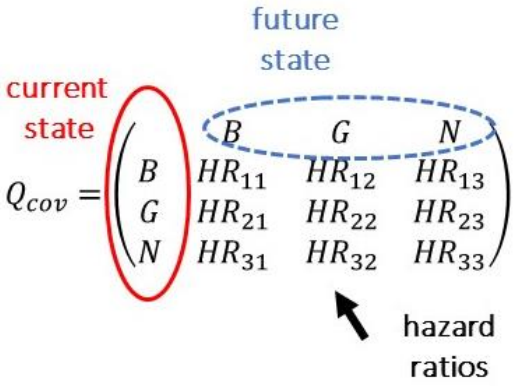


Figure 7.5. Annotated covariance matrix, showing the effect of a covariate on the transition matrix for the play session.

The Markov model may be best understood through a worked example. The following section presents the results of the Markov model for the cross-cultural toy choice dataset.

Results

Step 1. Toy Choice as Transitional Dynamic States: Interpreting the Markov Model

State space. The toy choice dataset was represented as a 3-dimensional transition intensity matrix, with states corresponding to play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). The matrix had no constraints: that is, children could choose any toy, and could transition from any state to any other.

System state vectors. Toy choices were represented as system state vectors. For each time point in the play session, the Markov model calculated a predicted system state vector (the probability of states *B*, *G*, and *N* at time *t*). In addition, for each time point in the play session, the dataset provided an observed system state vector (the observed state of either *B*, *G*, or *N* at time *t*). To check the Markov model's accuracy, the predicted system state vectors can be compared to the observed system state vectors at each time point.

Figure 7.6 shows how the predicted system state vectors can be compared to the observed system state vectors for each time point. Figure 7.6 is based on the data for boys from Brisbane, Australia. The plot revealed that the Markov model fit the general shape of the data well, because the predicted and observed values were similar. However, the predicted values diverged from observed values, indicating decreased model accuracy, towards the end of the play session.

It is not practical to present separate plots for each group of children (boys or girls, in Brisbane, Shipibo, Lenakel, or Navhal samples). Instead, model fit statistics are given in the below section on model selection.

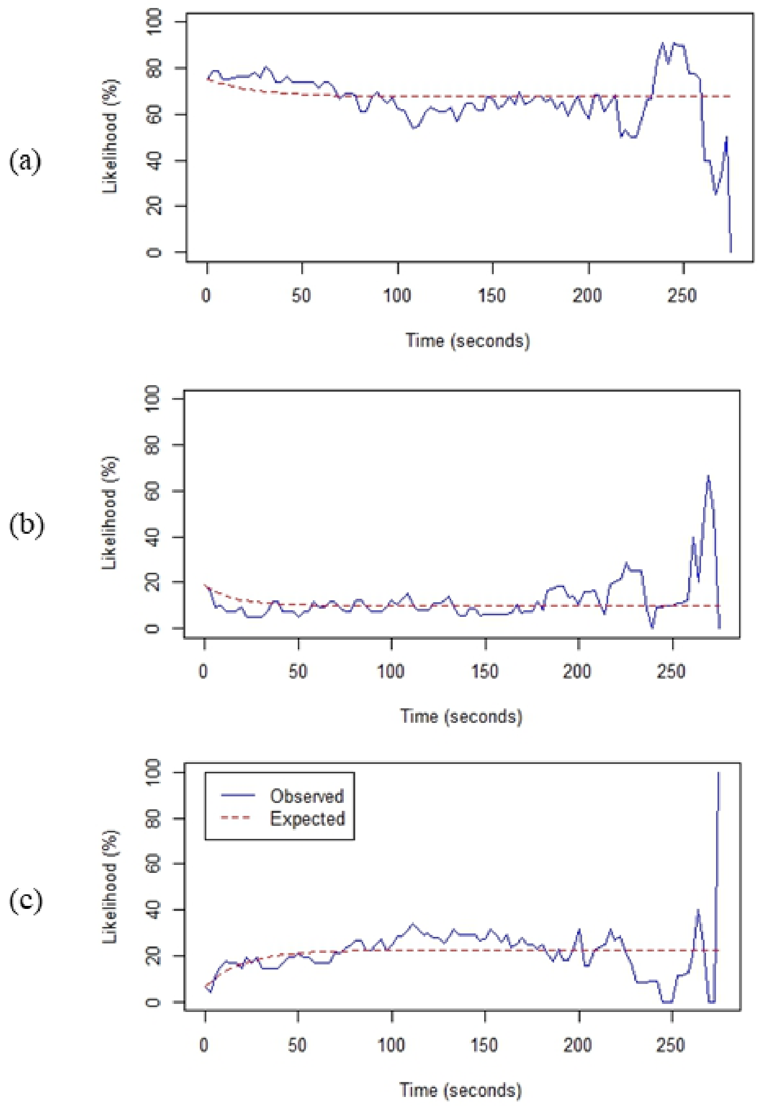


Figure 7.6. Dynamics of gender-typed toy play. Toy choices are represented as three possible states, one state per panel in the figure: (a) play with BTT, (b) play with GTT, and (c) play with neutral toys. The x-axis is the elapsed time in the free play session, in seconds. The y-axis is the likelihood of being in each state (a, b, or c). The dotted red line represents the predicted state of the system, based on the multistate Markov model. The solid blue line represents the observed state of the system, based on the dataset. The model predictions in Figure 7.6 represent the covariate-adjusted estimates for boys in Brisbane, Australia.

Transition matrices. The transition matrices represent the probability of switching from any system state (*B*, *G*, or *N*) to any other system state (*B*, *G* or *N*) at each time point. The Markov model dynamically updated the transition matrices over all the time points in the dataset, and finally calculated a stable state matrix. The stable state matrix represented the probability of switching from any system state (*B*, *G*, or *N*) to any other system state (*B*, *G*, or *N*) when the system was in a hypothetical equilibrium. There are multiple stable state matrices, one for each group of participants (boys or girls in Brisbane, Shipibo, Lenakel, or Navhal samples). These stable state matrices can be represented using tables. Table 7.1 presents the stable state matrix for boys in Brisbane, Australia.

		Future state (*t*)		
		B	*G*	*N*
Current state (*t*-1)	*B*	-0.03 (-0.04, -0.03)	0.02 (0.02, 0.02)	0.02 (0.02, 0.02)
	G	0.02 (0.02, 0.03)	-0.04 (-0.04, -0.04)	0.02 (0.01, 0.02)
	N	0.02 (0.02, 0.03)	0.02 (0.01, 0.02)	-0.04 (-0.04, -0.04)

Table 7.1. Stable state matrix for boys in Brisbane, Australia. States are play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). The number in each cell is the probability of switching to the state in the column (at time *t*) given the immediately previous state in the row (at time $t - 1$), represented as a hazard ratio. Numbers in brackets are 95% confidence intervals for the hazard ratio. A confidence interval crossing 1 represents an effect that is not statistically different from null.

Step 2. Environmental Covariates: Sex and Culture

Model selection. The fully specified model (including covariates for children's sex, culture, and the interaction between sex and culture) was tested against nested models that included fewer covariates. The results for model selection are shown in Table 7.2. Lower model fit statistics indicate better model "fit" to the data, and significant model comparisons indicate significantly better fit. The fully specified model, which included sex, culture, and the interaction between sex and culture, provided the best fit to the data.

Model	**Covariates**	**Interactions**	**Model fit statistics**	**Model comparison**
Null (1)	None	None	24620.25	
Partially specified (2)	Sex	None	24232.05	$\chi^2(6) = 388.20$ $p < .001$
Partially specified (3)	Culture	None	24310.05	$\chi^2(18) = 310.21$ $p < .001$
Partially specified (4)	Sex Culture	None	23956.44	$\chi^2(6) = 353.61$ $p < .001$
Fully specified (5)	Sex Culture	Sex * Culture	23779.09	$\chi^2(18) = 177.35$ $p < .001$

Table 7.2. Model fit estimates for null (1), partially specified (2), and fully specified (3) Markov models for toy choice. Model fit statistics are -2 log likelihood, and lower numbers indicate better fit. Significant model comparison values indicate significantly better fit to the data than the previous model. To preserve the nested structure necessary for model comparison, Models (2) and (3) were each tested against Model (1).

Covariate matrices. The covariates of the best-fitting model were sex, culture, and the interaction of sex and culture. The effects of the covariates are represented in the following tables. The numbers in each cell of the following tables are hazard ratios. For each covariate, the covariate matrix is presented, with an interpretation.

Sex. Table 7.3 presents the covariate matrix for sex. The numbers in each cell of Table 6.3 represent hazard ratios for the transition probabilities in girls, relative to boys. A hazard ratio of 1.00, or 100%, represents equal transition probabilities between boys and girls. If the hazard ratio is greater than 1.00, then the transition probability is greater in girls than in boys. In contrast, if the hazard ratio is less than 1.00, then the transition probability is smaller in girls than in boys. The magnitude of the covariate effect is the hazard ratio minus 1.00.

The hazard ratios are calculated with corresponding confidence intervals. If the confidence interval does not cross zero, then the covariate effect for that transition may be statistically significant. A statistically significant hazard ratio indicates that the transition probability is statistically significantly different for boys and for girls, and is shown with an asterisk in the table. The statistically significant hazard ratios are interpreted below the table.

		Future state (t)		
		B	*G*	*N*
Current state (t-1)	*B*		2.49* (1.60, 3.86)	1.81* (1.05, 3.14)
	G	0.29* (0.18, 0.47)		0.18* (0.10, 0.32)
	N	0.30* (0.17, 0.50)	3.10* (1.26, 7.66)	

Table 7.3. Hazard ratios for the sex covariate in the Markov model. States are play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). Numbers are hazard ratios of the transition intensities for girls compared with boys and represent the change in probability of switching to the state in the column (at time t) given the immediately previous state in the row (at time $t - 1$). Numbers in brackets are 95% confidence intervals for the hazard ratio. A confidence interval that does not cross 1 represents a statistically significant difference between the transition probability for boys and the transition probability for girls and is indicated with *.

As seen in Table 7.3, in the toy choice dataset, all transition probabilities were significantly larger or smaller in girls than in boys. To aid interpretation, the statistically significant transitions are presented in the following order: transitions to BTT, transitions to GTT, and transitions to neutral toys.

Transitions to BTT. Girls were 71% less likely than boys were to choose a BTT after a GTT (hr_{21}=0.29; 1.00-hr_{21}=0.71). Similarly, girls were 70% less likely than boys were to choose a BTT after a neutral toy (hr_{31}=0.30; 1.00-hr_{31}=0.70).

Transitions to GTT. Girls were 149% more likely than boys were to choose a GTT after a BTT (hr_{12}=2.49). Similarly, girls were 210% more likely than boys were to choose a GTT after a neutral toy (hr_{32}=3.10).

Transitions to NTT. Girls were 81% more likely than boys were to choose a neutral toy after a BTT (hr_{13}=1.81). In contrast, girls were 82% less likely than boys were to choose a neutral toy after a GTT (hr_{23}=0.18; 1.00-hr_{23}=0.82).

Culture. Table 7.4 presents the covariate matrix for culture. The numbers in each cell of Table 7.4 represent hazard ratios for the transition probabilities in children from the Shipibo, Lenakel, and Navhal samples, relative to the Brisbane sample. There are three covariates, one for each of the three named cultures: Shipibo, Lenakel, and Navhal. A hazard ratio of 1.00, or 100%, represents equal transition probabilities between children in the named culture (either Shipibo, Lenakel, or Navhal) and children in Brisbane. If the hazard ratio is greater than 1.00, then the transition probability is greater in the named culture than in Brisbane. In contrast, if the hazard ratio is less than 1.00, then the transition probability is smaller in the named culture than in Brisbane. The magnitude of the covariate effect is the hazard ratio minus 1.00.

The hazard ratios are calculated with corresponding confidence intervals. If the confidence interval does not cross zero, then the covariate effect for that transition may be statistically significant. A statistically significant hazard ratio indicates that the transition probability is statistically significantly different for the named culture (either Shipibo, Lenakel, or Navhal) and for Brisbane, and is shown with an asterisk in the table. The statistically significant hazard ratios are interpreted below the table.

			Future state (*t*)		
			B	*G*	*N*
Shipibo (school in Peru)	Current state (*t*-1)	*B*		0.78 (0.50, 1.15)	2.82* (1.94, 4.09)
		G	1.99* (1.22, 3.26)		0.88 (0.45, 1.70)
		N	2.35* (1.70, 3.25)	1.47 (0.52, 4.12)	
Lenakel (school in Vanuatu)	Current state (*t*-1)	*B*		2.59* (1.78, 3.80)	3.48* (2.32, 5.22)
		G	1.24 (0.79, 1.93)		0.68 (0.40 – 1.18)
		N	1.17 (0.82, 1.68)	4.27* (1.77, 10.29	
Navhal (kastom villages in Vanuatu)	Current state (*t*-1)	*B*		2.49* (1.74, 3.57)	2.20* (1.45, 3.33)
		G	0.67 (0.43, 1.03)		0.63* (0.39 – 0.82)
		N	0.57* (0.39, 0.82)	4.47* (1.93, 10.36)	

Table 7.4. Hazard ratios for the location covariates in the Markov model. States are play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). Numbers are hazard ratios of the transition intensities for girls compared with boys and represent the change in probability of switching to the state in the column (at time *t*) given the immediately previous state in the row (at time *t* – 1). Numbers in brackets are 95% confidence intervals for the hazard ratio. A confidence interval that does not cross 1 represents a statistically significant difference between the transition probability for boys and the transition probability for girls and is indicated with *.

As seen in Table 7.4, in the toy choice dataset, some of the transitions were statistically larger or smaller in the three cultures (Shipibo, Lenakel, or Navhal) than in

Brisbane. To aid interpretation, the statistically significant transitions are presented in the following order: transitions to BTT, transitions to GTT, and transitions to neutral toys.

Transitions to BTT. Children in the Shipibo villages were 99% more likely than children in Brisbane were to choose a BTT after a GTT (hr_{21}=1.99). Similarly, children in the Navhal villages were 43% less likely than children in Brisbane were to choose a BTT after a neutral toy (hr_{91}=0.57; 1.00-hr_{91}=0.43).

Transitions to GTT. Children in Lenakel were 159% more likely than children in Brisbane were to choose a GTT after a BTT (hr_{42}=2.59). Furthermore, children in Lenakel were also 327% more likely than children in Brisbane were to choose a GTT after a neutral toy (hr_{62}=4.27). Similarly, children in the Navhal villages were 149% more likely than children in Brisbane were to choose a GTT after a BTT (hr_{72}=2.49). Furthermore, children in the Navhal villages were also 347% more likely than children in Brisbane were to choose a GTT after a neutral toy (hr_{92}=4.47).

Transitions to neutral toys. Children in the Shipibo villages were 182% more likely than children in Brisbane were to choose a neutral toy after a BTT (hr_{13}=2.82). Similarly, children in Lenakel were 248% more likely than children in Brisbane were to choose a neutral toy after a BTT (hr_{43}=3.48). Similarly, children in the Navhal villages were 120% more likely than children in Brisbane were to choose a neutral toy after a BTT (hr_{73}=2.20). In contrast, children in the Navhal villages were 37% less likely than children in Brisbane were to choose a neutral toy after a GTT (hr_{83}=0.63; 1.00-hr_{91}=0.37).

Sex and Culture Interaction. Table 7.5 presents the covariate matrix for the interaction between sex and culture. The numbers in each cell of Table 7.5 represent interaction hazard ratios for the transition probabilities in girls, relative to boys, for children in each of the three cultures (Shipibo, Lenakel, or Navhal), relative to children in Brisbane. An interaction hazard ratio of 1.00, or 100%, represents equal sizes of the sex differences (the differences between boys' and girls' transition probabilities) in the named culture (Shipibo, Lenakel, or Navhal) and the sex differences (the differences between boys' and girls' transition probabilities) in Brisbane. If the interaction hazard ratio is greater than 1.00, then the sex difference is larger in the named culture (Shipibo, Lenakel, or Navhal) than the sex difference is in Brisbane. In contrast, if the interaction hazard ratio is less than 1.00, then the sex difference is smaller in the named culture (Shipibo, Lenakel, or Navhal) than the sex difference is in Brisbane. The magnitude of the covariate effect is the hazard ratio minus 1.00.

The hazard ratios are calculated with corresponding confidence intervals. If the confidence interval does not cross zero, then the covariate effect for that transition may be statistically significant. A statistically significant hazard ratio indicates that the sex difference in the transition probabilities is, statistically, significantly different in the named culture (either Shipibo, Lenakel, or Navhal) than in Brisbane, and is shown with an asterisk in the table. The statistically significant hazard ratios are interpreted below the table.

			Next state (*t*)		
			G	*B*	*N*
Sex differences in Shipibo Children	Previous state (*t*-1)	*G*		3.62* (2.05, 6.41)	0.59 (0.30, 1.15)
		B	0.59 (0.32, 1.09)		1.97 (0.87, 4.47)
		N	0.93 (0.48, 1.80)	2.77 (0.89, 8.65)	
Sex differences in Lenakel children	Previous state (*t*-1)	*G*		0.78 (0.45, 1.37)	0.63 (0.32, 1.23)
		B	1.84* (1.03, 3.30)		3.02* (1.42, 6.43)
		N	3.39* (1.74, 6.64)	1.04 (0.38, 2.85)	
Sex differences in Navhal children	Previous state (*t*-1)	*G*		0.56* (0.32, 0.98)	0.94 (0.48, 1.84)
		B	2.39* (1.32, 4.32)		4.89* (3.46, 9.72)
		N	2.35* (1.19, 4.64)	0.44 (0.17, 1.16)	

Table 7.5. Hazard ratios for the gender*location interaction covariates in the Markov model. States are play with BTT (*B*), play with GTT (*G*), and play with neutral toys (*N*). Numbers are interaction hazard ratios of the transition intensities for the sex differences (girls compared with boys) in each culture (Shipibo, Lenakel, or Navhal), relative to the sex differences in Brisbane. The interaction hazard ratios represent the change in probability of switching to the state in the column (at time *t*) given the immediately previous state in the row (at time *t* – 1). Numbers in brackets are 95% confidence intervals for the hazard ratio. A confidence interval that does not cross 1 represents a statistically significant difference between the sex differences in each culture and the sex differences in Brisbane and is indicated with *.

As seen in Table 7.5, in the toy choice dataset, some of the transitions showed statistically larger or smaller sex differences in the three cultures (Shipibo, Lenakel, or Navhal), compared to Brisbane. To aid interpretation, the statistically significant transitions are presented in the following order: transitions to BTT, transitions to GTT, and transitions to neutral toys.

Transitions to BTT. Overall, boys were more likely than girls were to choose a BTT after a GTT, but this difference was 84% larger in Lenakel than in Brisbane (hr_{51}=1.84). Furthermore, in Lenakel, the difference between boys' and girls' probability of choosing a BTT after a neutral toy was 239% larger than in Brisbane (hr_{61}=3.39). Similarly, in the Navhal villages, the difference between boys' and girls' probability of choosing a BTT after a GTT was 139% larger than in Brisbane (hr_{81}=2.39). Furthermore, in the Navhal villages, the difference between boys' and girls' probability of choosing a BTT after a neutral toy was 135% larger than in Brisbane (hr_{91}=1.35).

Transitions to GTT. Overall, girls were more likely than boys were to choose a GTT after a BTT, but this difference was 262% larger in the Shipibo villages than in Brisbane (hr_{12}=3.62). In contrast, in the Navhal villages, the difference between boys' and girls' probability of choosing a GTT after a BTT was 44% lower than in Brisbane (hr_{72}=0.56; 1.00-hr_{72}=0.44).

Transitions to neutral toys. In Lenakel, the difference between boys' and girls' probability of choosing a neutral toy after a GTT was 202% larger than in Brisbane (hr_{53}=3.02). Similarly, in the Navhal villages, the difference between boys' and girls' probability of choosing a neutral toy after a GTT was 389% larger than in Brisbane (hr_{83}=4.89).

Discussion

This chapter demonstrated how a dynamic multistate Markov model could be applied to the cross-cultural toy choice dataset. The primary goal of the chapter was to provide a worked example of how to apply a dynamic systems model to toy choice data. As a secondary goal, the chapter aimed to provide a substantive interpretation of the dynamic systems model results.

Applying Dynamic Systems to Toy Choice Data

Children's free play was represented as a set of transitional states over time, using a Markov model. The Markov model assumes the Markov property: that the system state in the immediate future depends only on the system state at the present time, and not on the sequence of past system states. The ability of the Markov model to predict free play suggests that children's toy choices may meet this assumption, and therefore, that children may choose future toys to play with based primarily on the toy they are currently using.

The Markov model fit the general shape of the data well. However, the predictive accuracy of the Markov model decreased towards the end of the 5-minute play session. This decrease in predictive accuracy may have occurred because children were less likely to choose new toys to play with near the end of the play session. By the end of the play session, children may have settled into play with a preferred toy. In technical terms, children's transition probabilities may have changed towards the end of the play session. The Markov model could theoretically be extended to account for this possibility, as discussed in the below section on matrix algebra and simulation.

New Insights from Dynamic Systems

The Markov model revealed several novel insights into the toy choice data. The covariate analysis revealed that transition probabilities were different for boys and girls, for every type of toy. These findings indicate that boys and girls may have used different transition rules when choosing toys in the free play session. Specifically, both boys and girls avoided other-gender toys, but boys avoided playing with GTT more than girls avoided playing with BTT. This result suggests that boys may be more sex-typed than girls are, throughout the free play session.

Furthermore, the results of the Markov model suggest that the underlying process influencing the sex difference in toy choices was, specifically, boys' avoidance of GTT. Sex differences in transition probabilities were largest for transitions toward play with GTT, and smaller for transitions away from GTT. Previous research has also found boys to be more sex-typed than girls are, on average (Banerjee & Lintern, 2000; Golombok et al., 2008; Huston, 1985; Serbin et al., 1979). However, the dynamics of sex typing are not clear from previous research. For example, in the case of toy preferences, are boys more sex-typed because they try playing with GTT and don't like them, or because they avoid playing with GTT in the first place? The results of the Markov model suggest that in the present data, boys avoided playing with GTT in the first place, although boys in Brisbane may have already learned to avoid GTT from previous experience. This finding aligns with previous assertions that boys may avoid GTT because of their status as a physical representation of female gender (e.g., Fagot, Leinbach, & Hagan, 1986; Roopnarine & Johnson, 1994).

The Markov model analyses behavioural transitions dynamically over the play session, and so could distinguish when boys tried a GTT and disliked it, from when a boy was avoiding playing with GTT at all. Traditional statistical methods could similarly have shown that boys made more sex-typed choices than girls did, but they could not have distinguished between dislike and avoidance.

The covariate analyses found that sex affected every transition probability, but that culture affected only some transition probabilities. Additionally, there were some significant interactions between culture and sex. Together, these findings indicate that in the present data, sex affected the entire dynamic system of children's toy choices, while cultural context had a more variable effect on parts of the system, and that sex and culture interacted to influence toy choices.

Limitations

For the Markov model, toy choices had to be defined as discrete and mutually exclusive (BTT, GTT, or neutral toys). However, toy choices in the real world may not fall neatly into discrete states: for example, children in a free play session may engage in simultaneous play with toys of more than one gender type. This limitation could be addressed by adding more states to the model – for example, a state representing simultaneous play with toys of more than one gender type, or a state representing no play – but these adjustments would increase model complexity, and so would have to be balanced with the model's predictive ability, to ensure that the results remained easy to interpret. Even with this limitation, the Markov model predicted children's toy choices with reasonable accuracy, although the accuracy declined towards the end of the play session.

Matrix Algebra and Simulation

The current chapter used matrix algebra to show how the Markov model was calculated and evaluated. In practice, the Markov model may also be calculated and evaluated using simulated data. Simulation methods can be extended to cover a wide range of scenarios, including scenarios that may not be possible with the matrix algebra procedures. For example, in the Markov model, children's transition probabilities were static; the model assumed that children were equally likely to make transitions at the beginning and at the end of the play session. However, in reality, children's transition probabilities could change over the duration of a play session; children may be more likely to switch between toys at the beginning of a play session than at the end of the play session, for example. Simulation methods could address this limitation by allowing transition matrices to be represented as variable probability distributions, instead of static estimates. A wide variety of simulation methods are available, including Monte Carlo simulations, as sometimes used in Bayesian statistics (see, for examples, Gamerman & Lopes, 2006; Gilks, Richardson, & Spiegelhalter, 1995; Kass, Carlin, Gelman, & Neal, 1998).

Practical Suggestions for Dynamic Systems Modelling in Toy Play Studies

Dynamic systems methodology is useful for psychological theory, because it requires theorists to specify the ultimate goals and fundamental mechanisms of the system (Guastello et al., 2009). For this reason, it has been used as a framework for theorising about gender development (e.g., Martin & Ruble, 2010). However, the statistical application of this methodology has been limited so far in gender development, partly due to a perceived lack of large datasets with which to validate the models (Martin & Ruble, 2010). The number of participants typically included in a toy play study is relatively small, with sample sizes usually less than 100 participants. When data points are calculated as average play time for each participant, then sample sizes are indeed too small for most dynamic systems methods. However, approaching the dataset as a dynamic system rather than as a series of static points may reveal more data, especially where these data were actually collected over a continuous time period. For example, in a free play session of 5 minutes with a single participant, an average play time approach will result in a single data point for that participant. A dynamic systems approach, in contrast, may result in many data points for that participant, because now the behaviour of interest is demonstrated over many slices of continuous time. In the current chapter, my dataset of 344 participants expanded to a full set of 3448 observed transitions between toys. It is hoped that this demonstration will encourage researchers who are interested in applying dynamic systems to their own work, to consider a dynamic systems approach, not only to their theoretical models, but also to the structure of their data.

Conclusions

The current chapter demonstrated how a dynamic systems method could be applied to a real-world dataset of cross-cultural toy choices. The demonstration showed that children's toy choices could be modelled as a set of transitional states over continuous time. A multistate Markov model represented the dynamics of toy choices in a free play session and described variations in toy choices across boys and girls in different cultures. The Markov model gave several new insights into the toy choice data. One insight was that boys appeared to actively avoid toys that were stereotyped for girls, rather than trying these toys and disliking them. Another insight was that sex affected all of the dynamic transitions in children's sex-typed toy preferences, while cultural context affected only some transitions. Finally, the interactive effects indicated that some cultural contexts affected the size of the difference between boys' and girls' toy choice transitions, but that some cultural contexts did not. Although the current chapter focused on toy choices, the method presented is flexible, and could be generalised to other research contexts. Markov models, like the one demonstrated here, may be usefully applied to research that focuses on human behaviour as a dynamic system.

Chapter 8

General Discussion

The present thesis aimed to examine the role of culture in children's sex-typed preferences for colours, toys, and affordances. It addressed this aim through two objectives: a cross-cultural study of children's preferences for colours, toys, and affordances; and a demonstration of how systems theory could be applied to the cross-cultural dataset. Overall, the results of the thesis suggested that some sex-typed preferences, such as some toy preferences, may be observed in multiple cultures. Additionally, the thesis found substantial cross-cultural variation in children's toy preferences. Finally, the thesis did not find evidence of culturally universal colour preferences or affordance preferences. Together, these results suggested that some aspects of children's sex and gender development may be culturally universal, and furthermore, that culture plays a role in children's sex-typed preferences. These findings may suggest that, in future, theories of the development of sex and gender could usefully consider the role of culture.

The chapters presented in this thesis each included a discussion section, which contextualised each chapter's specific findings with previous research. Therefore, in the current discussion chapter, these chapter-specific discussions are not repeated. Instead, the conclusions of each chapter are summarised in the context of the overall programme of research. Then, limitations of the overall research are presented, with caveats to the interpretation of the research findings. Finally, the thesis's novel findings are reviewed, with a focus on their relevance to theories at the intersection of gender, culture, and child development.

Summary of Chapters

Chapter 1 presented a review of theoretical approaches to the study of gender differences in toy preferences, and argued that a cross-cultural study of children's preferences for colours, toys, and affordances could address a gap in the existing literature. It further identified that a systems approach might offer a way to contextualise the results of the cross-cultural study. Chapter 2 gave an overview of the thesis aims, rationale, scope, and general methods.

Chapter 3 tested children's preferences for pink and blue in three cultural contexts. In a colour curve analysis, culture was found to be the most important determinant of children's colour preferences. The results of Chapter 3 suggested that children's sex-typed preferences for pink and blue may not replicate across cultures.

Chapter 4 tested children's preferences for sex-typed toys in four cultural contexts. Sex-typed toy preferences were found in each of the four cultures, including in a culture where the toys were completely novel. The results of Chapter 4 suggested that children's sex-typed preferences for toys may replicate across cultures, and that there was significant cultural variation in sex-typed toy preferences.

Chapter 5 reanalysed the data from Chapter 4, and tested children's preferences for play affordances in four cultural contexts. The pattern of sex-typed affordance preferences across cultures was not clear, perhaps due in part to limitations of the study design. The results of Chapter 5 suggested that the role of culture in children's sex-typed preferences for play affordances may require further study.

Chapter 6 demonstrated a systems approach to the cross-cultural dataset. A boosted regression tree (BRT) and classification and regression tree (CART) were applied to the cross-cultural dataset. The results of Chapter 6 suggested that children's sex affected their toy preferences, over culture and toy affordances.

Chapter 7 demonstrated an alternative, theory-driven systems approach to the cross-cultural dataset. A Markov model was applied to the cross-cultural dataset. The results of Chapter 7 suggested that boys and girls might apply different transition rules in free play, and therefore, that sex might affect the underlying dynamic system of children's toy choices.

Taken together, these chapters make substantive and methodological contributions to research on children's sex-typed preferences. However, before these contributions are discussed, this chapter first presents some limitations and caveats to the interpretation of this research.

Limitations and Caveats

The research presented in this thesis has several limitations, and associated caveats to the interpretation of the research results. Limitations specific to each chapter have been discussed previously. Here, limitations are presented that are relevant to the total set of results.

First, the results presented here are drawn from a limited selection of cultures. These results are drawn from four cultures: one minority world culture, similar to those that have already been studied extensively (Brisbane, Australia); two closely related ethnic and cultural groups (*kastom* villages on Tanna Island, Vanuatu; and school children in Lenakel town, on Tanna Island, Vanuatu); and a group of Shipibo villages in the Peruvian Amazon basin. This limited selection of cultures does not provide a comprehensive dataset across modern humans, and so results from this dataset should be interpreted as an initial attempt, not a definitive solution, to the role of culture in children's sex-typed preferences. In comparison, the Ethnographic Atlas compares information about basic features of life in 863 cultural groups across very diverse geographic regions (Murdock, 1967). More research is required to understand the role of culture in children's sex-typed preferences. At a minimum, future studies could test sex-typed preferences in populations that are of particular theoretical importance to human evolutionary history; for example, in modern populations who obtain most of their food by foraging.

Second, this research was designed and carried out with the assumption that play affordances, such as toys with faces, propulsion, dress-up, and role play, were an important determinant of children's preferences for sex-typed toys. This assumption was not consistently supported by the research findings. There are several possible reasons for this finding, and these are discussed in Chapter 5. In general, however, the focus on these affordances came at the cost of a focus on other affordances of the toys that might also have varied by sex-type, such as their potential for aggressive or competitive play (Blakemore & Centers, 2005; Miller, 1987). Furthermore, the affordances may not have been equally represented in both sex-typed toys; for example, the racing car may have better afforded propulsive motion than did the pony carriage. This limitation might be possible to address in part in future research, by retroactively coding the toys used in the present study according to these additional criteria, but this limitation would be best addressed by additional studies. For example, future studies could select a set of toys that vary in their potential for aggression or competition. Alternatively, future studies could select boy-type and girl-type toys that have equal aggressive or competitive properties, although girl-type aggressive toys may be difficult to find.

Third, the results are based on the amount of time that children spent physically in contact with the toys, and contain no information about how children used the toys. The analyses assumed that children were using the toys for their associated play affordance; for example, that children were using the toy vehicles for propulsion. This assumption is common in free play studies, because toy choices, as a behavioural measurement, are well-defined, transparent, and replicable. However, personal impressions, based on viewing the videos, suggest that the children almost always played with the toys in the expected way.

Fourth, there are general caveats to the interpretation of behavioural research that applies a single, standardised empirical paradigm in variable cultural contexts. The tasks in this thesis were the same, and the colours and toys were the same, in every culture studied. Behavioural tasks may be interpreted differently in different cultures, and it is important to consider the meaning of each task in a local context, to ensure that the task really is measuring something comparable in each place. Additionally, the data were gathered over a single, brief, solitary play session, focused on acquiring comparable data in each place. The toy preference dataset thus has low ecological validity, and it may not reflect what children would do in a more naturalistic setting. The research protocol attempted to address this issue through the recruitment and consultation of collaborators in each location. These collaborators had grown up, and had family members in, the target villages, but had also lived in cities and could therefore provide a comparative perspective. These collaborators provided advice on the local meaning of the tasks, as well as contacts, ethnographic data, information on ethical issues, and translation. The aim of this collaboration was to ensure that the study methods were, as far as possible, relevant to all of the cultures studied.

Fifth, the research presented in this thesis should not be interpreted as a test of how sex-typed children were in each culture. For example, the sex differences in children's toy preferences in the Navhal *kastom* villages were smaller than the sex differences in children's toy preferences in Brisbane, but this result should not be interpreted to mean that children in the Navhal *kastom* villages were less sex-typed in general than were children in Brisbane. The point of the current research was not to test how sex-typed the children in each culture were, or to test for sex differences in culturally relevant sex-typed behaviours and preferences. Instead, the point of the current research was to test whether the sex differences, that have been seen in children's toy preferences in minority world cultures, were still apparent when the toys were removed from that cultural context. To achieve this goal, the same toys were used in each cultural context. Therefore, some of the toys were not culturally relevant in some of the samples, and this lack of cultural relevance was an intended part of the research design. If future researchers wanted to test how sex-typed children were in each culture, they would need to select toys or other objects that are culturally relevant, and probably different in each location.

Finally, the data for this experiment may be more “fuzzy” than data collected in the university-based, controlled setting that is typical of other toy preference studies. The research sites were unusual, remote, and in the cases of Vanuatu and Peru, had little infrastructure for communications or transport. These characteristics introduced a range of logistical issues. The toys selected for study had to be light enough to carry on planes, in boats, and through the jungle. The cameras occasionally failed due to lack of electricity with which to charge them, or because they ran out of recording space. Although the testing area was standardised as much as possible across sites, minor variations were unavoidable. These extra sources of variation in the data are not expected to have biased the results, since there is no reason to believe that they would have affected boys and girls differently. They would, however, have introduced extra statistical noise into the dataset. This possibility was anticipated during the research design, and is part of the reason for recruiting larger numbers of participants than is typical of existing free play studies of toy preferences, or than the power analysis, based on these existing studies, suggested were needed. In addition, the tasks were made as simple and straightforward as possible, so that some data could be collected even under difficult and changing conditions.

Synthesis of Study Results

Even with the above limitations, the research presented in this thesis represents the first large, empirical, cross-cultural dataset on sex differences in children’s preferences for colours, toys, and affordances. The following sections synthesise some of the more surprising or interesting findings from the thesis as a whole.

Gender Differences in Colour Preferences Were Not Found in All Cultures

Previous research has disagreed about the role of culture in adults’ colour preferences. One previous study found that sex differences in adult colour preferences replicated across majority world and minority world cultures (Sorokowski et al., 2014), but another study did not (Taylor, Clifford, et al., 2013). Furthermore, children’s colour preferences had not previously been tested in majority world cultures, and different theories made contrasting predictions. Cone-contrast theory (Hurlbert & Ling, 2007) predicted that children in majority world cultures should show sex differences in their preferences for pink and blue (Hurlbert & Owen, 2015). Other theories predicted that children in majority world cultures should not show sex differences in colour preferences. For example, ecological valence theory predicted that children should only prefer pink when they had direct personal experiences with pink objects, and that boys and girls in majority world cultures would not have had different experiences with pink objects (Palmer & Schloss, 2010). Alternatively, cognitive theories of gender development predicted that children should not show sex differences in colour preferences in majority world cultures, where the colour pink might not act as a gender marker in children's personal gender schemas or in cultural gender stereotypes.

In the present thesis, children's preferences for pink and blue did not show significant sex differences in either of the two majority world cultures. Furthermore, a colour curve analysis did not show sex differences in children's preferences for reddish hues, or for colour saturation, in either of the two majority world cultures. These results suggest that sex differences in children's preferences for pink are influenced by culture, and are unlikely to be culturally universal. Additionally, these results provide some support for ecological valence theory, and for cognitive theories of gender development in regard to colour preferences.

Some Gender Differences in Toy Preferences Were Found in Multiple Cultures

The findings of the present thesis support the view that children's toy preferences are influenced by a complex and dynamic system of interacting factors, including sex and culture. Toy preferences have been used as an indicator of children's broader gender development since the 1960s (e.g., Anastasiow, 1965; DeLucia, 1963). Theories of gender development have changed over that time, from theories that emphasised single influences (Anastasiow, 1965; Kohlberg, 1966; Bem, 1981; Berenbaum & Hines, 1992), to integrated theories that acknowledge the interactive relationships between an individual's biology, environment, and cognition (Eagly & Wood, 2013; Green et al., 2004; Hines, 2005; Liben & Bigler, 2002; Martin & Ruble, 2010). The cross-cultural dataset that was collected for this thesis could provide several contributions to these debates.

Sex differences in children's toy preferences appeared in each cultural context, suggesting that there may be some common processes that underpin sex or gender differences in children's toy preferences. These possibilities might be usefully explored in future. There was also significant variation between cultures, suggesting that children's toy preferences were affected by cultural context. However, preferences for some specific sex-typed toys were not consistent cross-culturally. For example, girls did not prefer the baby doll over the toy vehicle in every culture. It is not possible to know if this occurred because single toys provide relatively unreliable measures of children's preferences, or if specifics of the cultural contexts, for example differences in girls' childcare responsibilities between cultures, contributed to these results.

Functional Explanations: Play Affordance

This thesis aimed to identify the functional features of toys, specifically the play affordances, that might drive sex differences in children's toy preferences. As noted above, the results of the cross-cultural study suggested that some aspects of children's sex-typed toy preferences may be replicated in different cultures, even where the toys are relatively novel. It was hypothesised that play affordances might be the functional feature of toys that was responsible for these cross-cultural similarities. However, children's preferences for toy affordances, such as toys with faces, propulsion, dress-up, and role play, showed inconsistent results in each of the cultures studied. In each culture, boys preferred propulsion toys, and girls preferred dress-up toys, but these sex differences were not always statistically significant. In addition, even the directions of the sex differences were inconsistent for toys with faces and for roleplay toys (Chapter 5). These results suggest that these particular affordances may not be the functional feature responsible for the cross-cultural sex differences in toy preferences (but see the above limitations and caveats).

The results for affordances may be partly due to limitations in how affordances were defined and measured. Affordances are a relatively new area of toy preference research, and not many previous studies have analysed affordances as a functional feature of sex-typed toys. In one recent attempt, a set of studies by Dinella and colleagues (2016) tested children's preferences for sex-typed and neutral toys, with and without propulsion affordances. Like the current research, the studies by Dinella et al. (2016) found no sex difference in children's preferences for propulsion toys, when both boy-type propulsion toys and girl-type propulsion toys were available. However, like the current thesis, the studies by Dinella et al. (2016) did not measure children's actual propulsive motion, and instead only measured play with toys that were assumed to afford propulsion. Therefore, both the current thesis and the Dinella et al. (2016) study may have failed to find a sex difference in children's preference for propulsion, due to flaws in their operational definition of propulsion.

Overall, the results of the current thesis did not support propulsion, toys with faces, dress-up, or role play affordances as functional explanations for the cross-cultural sex difference in children's toy preferences. The next section therefore explores possible alternatives.

Possible Alternative Affordances

Alternative affordances, other than toys with faces, propulsion, dress-up, or role play affordances, may be candidates for the functional features of toys that explain sex differences in children's toy preferences. Specifically, male-type toys are commonly associated with aggression and competition, and female-type toys are commonly associated with appearance and caretaking (Blakemore & Centers, 2005; Miller, 1987). The current section examines the possibility that these might be the functional features of sex-typed toys that cause them to differentially appeal to boys and girls. The following discussion focuses on these alternative affordances of male-type toys, especially aggression and competition.

Could aggression and competition affordances explain the cross-cultural results of the current thesis? In each culture, there were statistically significant sex differences in children's preferences for boy-type toys (BTT) and girl-type toys (GTT). These sex differences were large in children from Shipibo villages in Peru, and small in samples of *kastom* children on Tanna Island in Vanuatu. This result was unexpected, because the Shipibo children and the *kastom* children were both assumed to be unexperienced with the toys. Toy affordances such as propulsion may not explain these cultural variations well, because the same toys were used in each culture, and the toys had faces, propulsion, dress-up, and roleplay affordances wherever the toys were located. In contrast, aggression and competition affordances could potentially explain these cultural variations better. Specifically, it may be useful to consider a possible relationship between the aggressive and competitive affordances of the BTT, and the importance of aggression and competition as masculine cultural stereotypes in the cultures studied.

Participants in the Shipibo sample may have placed higher importance on aggression and competition as masculine cultural stereotypes, compared to the other cultural settings. In the Shipibo villages, interviews with adults consistently elicited one adult activity that was performed only by males, and never by females: fishing. When asked why only males could fish, Shipibo adults explained that fishing was dangerous; fishing in the villages was done via rod and pole, bow and arrow, or harpoon, from a wooden canoe, on the tributary rivers and streams of the Ucayali River. These rivers and streams are home to many predatory species, including several species of crocodile, piranhas, and anacondas, all of which are attracted to the fishing site by the potential for food in the form of used bait and rejected fish. Field interviews in another Shipibo town (Behrens, 1986) suggest that hunting is also an exclusively masculine activity, potentially for similar reasons. Hunted game and fish are valued more highly than domesticated meats, and men who can regularly supply hunted meat and fish are likely to achieve social standing through their ability to distribute excess food to other households, and to sell the meat for money (Behrens, 1992). Masculinity in Shipibo culture may therefore be strongly associated with aggression and competition in a dangerous environment.

In contrast, masculinity in *kastom* culture on Tanna may not be associated with combat or dangerous environments. Meat is sourced from domesticated animals, particularly pigs and chickens, and is not hunted. Tanna Island has no predators, no dangerous large animals, no snakes, and no carnivorous fish, so there is no inherent danger in male (or female) food production. Social roles are segregated by gender, but the authority of men on Tanna is traditionally grounded in social and spiritual responsibility, not in violence or danger. Although it is the responsibility of men to resolve disputes, conflict resolution in *kastom* culture is explicitly focused on balance and reciprocity (Gregory & Gregory, 2002). Due to the linguistic diversity of *kastom* tribes, conflicts are resolved very formally, with careful attention to the right of every person involved to have an equal say in the outcome (Lindstrom, 1983). Thus, there may be a low association between masculinity, and aggression and competition, in *kastom* culture on Tanna.

If toy affordances such as aggression and competition were responsible for gender differences in children's preferences for sex-typed toys, then cultures where masculinity was strongly related to aggression and competition (like, potentially, the Shipibo) might also be expected to show a strong preference for BTT, which are characterised by aggression and competition. The present cross-cultural study included some BTT with aggressive/competitive affordances, such as the toy sword and the action figure (Benenson, 2014). No GTT in the present study, however, had aggressive/competitive affordances. In contrast, cultures where masculinity is related to power structures other than danger and combat (like, potentially, the *kastom* villages on Tanna) might be expected to show a weaker preference for BTT. These expected patterns match the observed results of the cross-cultural study, and therefore, may provide some support for the hypothesis that aggression and competition may be evolutionarily adaptive functional features of the toys. Future research might address this possibility more explicitly.

Accordingly, BTT may be used by boys to practice aggression and competition (Hellendoorn & Harinck, 1997; Humphreys & Smith, 1987), especially in cultural environments where aggression and competition are important parts of adult masculinity. Evolutionary analyses suggest that in humans, sexually dimorphic traits favouring men, such as size and muscularity, are functionally adapted to contest with other men (Eagly & Wood, 1999; Puts, 2010). However, children have not yet developed adult size and muscularity, and so may display competitive traits through toys instead. Consequently, the same adaptive pressures that favour adult size and muscularity in men, may influence boys to seek out aggressive and competitive toys, especially in cultural contexts where aggression and competition are stereotyped as masculine ideals.

This idea, that cultural variations in aggression and competition may influence children's sex-typed toy preferences, gives rise to specific and testable predictions. For example, according to this idea, sex differences in toy preferences might be higher in social settings where cultural representations of masculinity are closely tied to aggression and competition. Exploratory areas of research are also indicated: for example, this perspective would be enriched by observations of the relationship between children's sex-typed toy preferences, and sex differences in children's preferences for games with themes of competition, dominance, and ranking.

Complex and Dynamic Systems in Toy Preferences

Existing theories of sex and gender development do not make explicit predictions about children's toy preferences in cultures where the toys are unfamiliar. Although some biological perspectives may emphasise universality (e.g., Alexander, 2003), and some social and cognitive perspectives may emphasise variability (e.g., Carter & Patterson, 1982), it is generally acknowledged that these perspectives work together in a complex and dynamic system (Bandura & Bussey, 2004; Hines, 2005, 2013; Liben & Bigler, 2002; Martin & Ruble, 2010) . Therefore, this thesis attempted to integrate the findings using statistical methods based on systems theory.

Systems theory takes an integrative approach to theories of gender development. It focuses on the description of complex and dynamic interactions between explanatory factors (Gopnik et al., 2004; Hines, 2013; Martin & Ruble, 2010; Yoshikawa & Hsueh, 2001). The current thesis provided the first demonstrations of how systems theory could be empirically applied to research on children's toy choices. These demonstrations provided empirical confirmation that integrative theoretical models, like those that had previously been proposed by many investigators (e.g., Bussey & Bandura, 1999; Hines, 2013; Liben & Bigler, 2002; Martin & Ruble, 2010; Pasterski et al., 2005; Wallen, 1996), might be possible to analyse in real data.

In previous research, complex and dynamic systems theory has been used to explain how sex-typed behaviour, such as toy choices, might arise from an underlying system of biological and environmental processes (e.g., DiDonato et al., 2012; Fausto-Sterling et al., 2012). Similarly, in the present thesis, the results of the systems chapters suggest that sex may predispose an individual to a certain set of sex-typed behaviours, and that culture may influence how this predisposition translates into behaviours. The present studies therefore provide further support for the specific position that genetic sex provides the starting point for sex-typed development, and that development is modified by environmental processes (Eagly & Wood, 2013; Fausto-Sterling, 2012a; Hines, 2013; LeVay, 2011; Wallen, 2009).

The systems chapters demonstrated that complex theoretical processes could be mathematically recreated and applied to children's real-world behaviour. Consequently, the specific models demonstrated here could be used in a range of other applications. For example, the Markov model could be adapted to provide dynamically updating predictions of children's toy preferences in real time, and this might be useful for intervention studies. Additionally, the present thesis provided step-by-step demonstrations of the systems methods, in the hope that these demonstrations might help researchers in related fields to add these methods to their analytic toolsets. These general methods could be applied to other examples of sex-typed behaviour where complex systems are theorised, such as human sexuality (Fausto-Sterling, 2012b).

Gender, Toys, and Culture: Conclusions

Culture in this thesis was considered as the context in which children develop toy preferences, and it includes social structures, adult gender roles, and symbols. Previous toy preference research had considered culture in this broad sense. The present work extended this previous broad view of culture, to include a specific viewpoint on how cultural explanations and functional explanations might interact to explain children's sex-typed toy preferences.

Cultural Explanations. By adding work in majority world cultures, this thesis confirmed and extended previous work in the minority world, on toys as cultural objects. In a majority world culture where sex-typed toys were unfamiliar, children showed smaller sex differences in toy preference than they did in a minority world culture where the toys were familiar. This result strengthens prior findings that toys' gender categories are culturally defined in the minority world (Cherney & Dempsey, 2010; Edwards, 2000; Schau et al., 1980; Serbin et al., 1979). Additionally, children's preferences for sex-typed toys varied among the three majority world cultures, and this variation may relate to differences among the cultures in adult gender roles. This finding confirms earlier work that positioned toys as symbolic representations of gendered social structures (Eisenberg et al., 1982; Fagot & Patterson, 1969; Richardson & Simpson, 1982) and adult gender roles (Edwards, 2000; Serbin et al., 2001).

Functional Explanations. The present work also added to previous work on functional explanations for children's sex-typed toy preferences. Previous research had suggested that toys may have functional features, such as colour or affordance, that made them differentially attractive to boys and girls (Campbell et al., 2000; Jadva et al., 2010). Further, prior work suggested that sex differences in children's preferences for these functional features may relate to children's early androgen exposure, and may therefore arise independent of cultural context (Alexander, Wilcox, & Farmer, 2009; Hassett et al., 2008; Wong et al., 2013). Supporting this view, the present thesis found sex differences in children's toy preferences in a majority world culture where the toys were unfamiliar. Since these children had little cultural context for the toys, this finding supports prior assertions that there may be an inborn component to children's toy preferences, possibly related to functional features of the toys.

Toys as Sex-Typed Objects. The specific functional features explored in this thesis were colour and play affordance. First, colour was investigated in a separate test that focused on children's preferences for pink and blue. Results indicated that in majority world cultures, children did not show sex differences in preference for pink, in contrast to prior results in the minority world (Jadva et al., 2010; Wong & Hines, 2015). This result contradicts earlier claims, based on minority world research, that sex differences in colour preference are inborn and explain children's sex-typed toy preferences (Alexander, 2003; Del Giudice, 2017). Second, this thesis explored play affordance in a reanalysis of the toy preference data. Results did not suggest that affordances such as propulsion, faces, dress-up, or roleplay, were preferred by boys or by girls. This result contrasts with earlier proposals that these affordances may explain sex differences in children's toy preferences (e.g., Alexander, 2003; Benenson, Liroff, Pascal, & Cioppa, 1997; Campbell, Shirley, Heywood, & Crook, 2000; Cherney & Dempsey, 2010). However, this result aligns with recent studies that fail to find sex differences in preferences for propulsion in the minority world (Dinella et al., 2016).

Alternatively, the affordance analysis may have failed to find a result due to methodological shortcomings, and it did not consider potentially important alternative affordances, such as aggressive/competitive play. The present thesis thus contributed two key insights to research on toys as sex-typed objects: first, that colour is unlikely to be a universal explanation for sex differences in children's toy preferences; and second, that aggressive and competitive affordances may be a fruitful area of future research that seeks functional explanations for sex differences in children's toy preferences.

Statistical Methods Based on Systems Theory. The methods chapters of this thesis demonstrated new statistical methods, based on systems theory, that could be applied to children's toy preferences. The machine learning method additionally found that roleplay toys were particularly important for cross-gender play, concurrent with prior results of cross-cultural naturalistic work on gender differences in children's play (Edwards, 2000), and supporting the view that toys function partly as symbols of culturally-appropriate adult gender roles (Edwards, 2000; Richardson & Simpson, 1982; Serbin et al., 2001) . The dynamic transition method further discovered that boys avoided approaching girl-type toys, supporting previous theory that boys may be influenced by toys as representations of social gender structures, more than girls are (Doering, Zucker, Bradley, & MacIntyre, 1989; Fagot et al., 1986; Roopnarine & Johnson, 1994). Together, the methods chapters confirmed that children's sex-typed toy preferences are the result of a complex and dynamic system, providing empirical support for the increasing use of systems theory to characterise gender development (DiDonato et al., 2012; Fausto-Sterling et al., 2012; Hines, 2013; Martin & Ruble, 2010), and providing additional support for existing work on dynamic systems in child development more broadly (Overton & Lerner, 2014; van Geert, 2011).

Summary of Contributions

The general aim of the current thesis was to examine the role of culture in children's sex-typed preferences for colours, toys, and affordances, and to explore possible uses of systems theory in analysing and understanding children's sex-typed toy choices. This general aim addressed two previous gaps in the literature: first, that the results of previous empirical studies of children's sex-typed colour and toy preferences had not been replicated in different cultures; and second, that systems theory had not been mathematically applied to data on children's sex-typed toy choices. These two research gaps were addressed through two research objectives: a substantive objective, and a methodological objective.

The substantive objective was to test whether the results of previous empirical studies of children's sex-typed colour and toy preferences could be replicated in majority world cultures. This first objective was met by a large, cross-cultural, empirical study of children's colour and toy preferences. The cross-cultural study found that sex differences in children's colour preferences replicated in a minority world culture, but did not replicate in two majority world cultures. The cross-cultural study further found that sex differences in children's toy preferences replicated in a minority world culture and in three majority world cultures, and additionally, that there was significant variation between cultures in children's toy preferences. The cross-cultural study also investigated a functional component of toy preferences, specifically, play affordance, and found mixed results. The cross-cultural study provided novel data concerning the role of culture in children's preferences for colours, toys, and affordances.

The methodological objective was to provide a demonstration of how systems theory could be mathematically applied to data in this field. This second objective was met by two demonstrations: first, a machine learning method, and second, a multistate dynamic systems model. These two statistical applications of systems theory provided new insights into the complex and dynamic systems underlying children's toy choices. The statistical methods addressed two limitations in traditional statistical approaches: multiple non-independent predictor variables with complex interactions; and averaging dynamic behaviour changes over a single time frame. The systems chapters provided novel methodological contributions that could be applied to wider research on children's sex-typed toy choices.

Overall, the results of the current thesis suggest that culture plays a role in children's sex-typed preferences for colours, toys, and affordances, and it is possible that culture may play a role in sex or gender development more broadly. The results of the current thesis suggest that theories of sex and gender development could be usefully extended to include majority world cultures. Additionally, future research could move towards an integrated systems perspective, that sees children's sex-typed preferences as functional responses to, and influences on, their cultural environment.

References

Adams, R. J. (1987). An evaluation of color preference in early infancy. *Infant Behavior and Development, 10*(2), 143–150.

Alexander, G. M. (2003). An evolutionary perspective of sex-typed toy preferences: Pink, blue, and the brain. *Archives of Sexual Behavior*, *32*(1), 7–14.

Alexander, G. M., & Hines, M. (1994). Gender labels and play styles: Their relative contribution to children's selection of playmates. *Child Development*, *65*(3), 869–879. doi: 10.2307/1131424

Alexander, G. M., & Hines, M. (2002). Sex differences in response to children's toys in nonhuman primates (*Cercopithecus aethiops sabaeus*). *Evolution and Human Behavior, 23*(6), 467–479.

Alexander, G. M., & Saenz, J. (2012). Early androgens, activity levels and toy choices of children in the second year of life. *Hormones and Behavior, 62*(4), 500–504. doi: 10.1016/j.yhbeh.2012.08.008

Alexander, G. M., Wilcox, T., & Farmer, M. E. (2009). Hormone–behavior associations in early infancy. *Hormones and Behavior*, *56*(5), 498–502. doi: 10.1016/j.yhbeh.2009.08.003

Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex differences in infants' visual interest in toys. *Archives of Sexual Behavior*, *38*(3), 427–433. doi: 10.1007/s10508-008-9430-1

Anastasiow, N. J. (1965). Success in school and boys' sex-role patterns. *Child Development*, *36*(4), 1053–1066. doi: 10.2307/1126944

Arnold, A. (1985). Organizational and activational effects of sex steroids on brain and behavior: A reanalysis. *Hormones and Behavior, 19*(4), 469–498. doi: 10.1016/0018-506X(85)90042-X

Arnold, A. P. (2009). The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and Behavior*, *55*(5), 570–578. doi: 10.1016/j.yhbeh.2009.03.011

Arthur, W. B., Durlauf, S. N., & Lane, D. A. (1997). *The economy as an evolving complex system II* (Vol. 27). Addison-Wesley Reading, MA.

Atkin, C. K. (1975). *The effects of television advertising on children. Report no. 2: Second year experimental evidence. Final report*. Retrieved from https://eric.ed.gov/?q=%22The+Effects+of+Television+Advertising+on+Children%22&id=ED116784

Aubrey, J. S., & Harrison, K. (2004). The gender-role content of children's favorite television programs and its links to their gender-related perceptions. *Media Psychology*, *6*(2), 111–146. doi: 10.1207/s1532785xmep0602_1

Auster, C. J., & Mansbach, C. S. (2012). The gender marketing of toys: An analysis of color and type of toy on the Disney store website. *Sex Roles*, *67*(7–8), 375–388.

Bandura, A. (1969). Social learning of moral judgments. *Journal of Personality and Social Psychology*, *11*, 275-279.

Bandura, A., & Bussey, K. (2004). On broadening the cognitive, motivational, and sociostructural scope of theorizing about gender development and functioning: Comment on Martin, Ruble, and Szkrybalo (2002). *Psychological Bulletin*, *130*(5), 691–701. doi: 10.1037/0033-2909.130.5.691

Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397–408. doi: 10.1111/1467-9507.00133

Barkley, R. A., Ullman, D. G., Otto, L., & Brecht, J. M. (1977). The effects of sex typing and sex appropriateness of modeled behavior on children's imitation. *Child Development*, *48*(2), 721–725. doi: 10.2307/1128683

Bar-Yam, Y. (1997). *Dynamics of complex systems* (Vol. 213). Addison-Wesley Reading, MA.

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: an appreciation and an update. *Trends in Ecology & Evolution*, *28*(12), 712–718. doi: 10.1016/j.tree.2013.09.013

Behrens, C. A. (1986). The cultural ecology of dietary change accompanying changing activity patterns among the Shipibo. *Human Ecology*, *14*(4), 367–396.

Behrens, C. A. (1992). Labor specialization and the formation of markets for food in a Shipibo subsistence economy. *Human Ecology*, *20*(4), 435–462.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*(4), 354–364. doi: 10.1037/0033-295X.88.4.354

Benenson, J. F. (2014). *Warriors and worriers: The survival of the sexes*. New York: Oxford University Press.

Benenson, J. F., Liroff, E. R., Pascal, S. J., & Cioppa, G. D. (1997). Propulsion: A behavioural expression of masculinity. *British Journal of Developmental Psychology*, *15*(1), 37–50.

Benenson, J. F., Tennyson, R., & Wrangham, R. W. (2011). Male more than female infants imitate propulsive motion. *Cognition*, *121*(2), 262–267. doi: 10.1016/j.cognition.2011.07.006

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science*, *3*(3), 203–206.

Berenbaum, S. A., & Snyder, E. (1995). Early hormonal influences on childhood sex-typed activity and playmate preferences: Implications for the development of sexual orientation. *Developmental Psychology*, *31*(1), 31–42. doi: 10.1037/0012-1649.31.1.31

Berl, R. E. W., & 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. doi: 10.1371/journal.pone.0120180

Blakemore, J. E., LaRue, A. A., & Olejnik, A. B. (1979). Sex-appropriate toy preference and the ability to conceptualize toys as sex-role related. *Developmental Psychology*, *15*(3), 339–340. doi: 10.1037/0012-1649.15.3.339

Blakemore, J. E. O., & Centers, R. E. (2005). Characteristics of boys' and girls' toys. *Sex Roles*, *53*(9–10), 619–633. doi: 10.1007/s11199-005-7729-0

Bloch, M. N. (1987). The development of sex differences in young children's activities at home: The effect of the social context. *Sex Roles*, *16*(5–6), 279–301. doi: 10.1007/BF00289955

Bloch, M. N. (1989). Young boys' and girls' play at home and in the community: A cultural ecological framework. In M. N. Bloch & A. D. Pellegrini (Eds.). *The ecological context of children's play* (pp. 120–154). Norwood, N. J.: Praeger.

Bloch, M. N., & Adler, S. M. (1994). African children's play and the emergence of the sexual division of labor. In J. L. Roopnarine, J. E. Johnson, & F. H. Hooper (Eds.), *Children's play in diverse cultures* (pp. 148–178). SUNY Press.

Bock, J. (2002). Learning, life history, and productivity: Children's lives in the Okavango Delta, Botswana. *Human Nature*, *13*(2), 161–197. doi: 10.1007/s12110-002-1007-4

Boyette, A. H. (2016a). Children's play and culture learning in an egalitarian foraging society. *Child Development*, *87*(3), 759–769. doi: 10.1111/cdev.12496

Boyette, A. H. (2016b). Children's play and the integration of social and individual learning: A cultural niche construction perspective. In H. Terashima & B. S. Hewlett (Eds.), *Social learning and innovation in contemporary hunter-gatherers* (pp. 159–169). Tokyo: Springer Japan. doi: 10.1007/978-4-431-55997-9_13

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*(1), 1–14. doi: 10.1017/S0140525X00023992

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676–713. doi: 10.1037/0033-295X.106.4.676

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*(6), 477–484. doi: 10.1038/nrn1909

Caldera, Y. M., Huston, A. C., & O'Brien, M. (1989). Social interactions and play patterns of parents and toddlers with feminine, masculine, and neutral toys. *Child Development*, *60*(1), 70–76. doi: 10.2307/1131072

Caldwell, C., & Whiten, A. (2002). Evolutionary perspectives on imitation: Is a comparative psychology of social learning possible? *Animal Cognition*, *5*(4), 193–208. doi: 10.1007/s10071-002-0151-x

Campbell, A., Shirley, L., Heywood, C., & Crook, C. (2000). Infants' visual preference for sex-congruent babies, children, toys and activities: A longitudinal study. *The British Journal of Developmental Psychology*, *18*, 479.

Caro, T. M., & Borgerhoff Mulder, M. (1987). The problem of adaptation in the study of human behavior. *Ethology and Sociobiology*, *8*(1), 61–72. doi: 10.1016/0162-3095(87)90058-6

Carter, D. B., & Levy, G. D. (1988). Cognitive aspects of early sex-role development: The influence of gender schemas on preschoolers' memories and preferences for sex-typed toys and activities. *Child Development*, *59*(3), 782–792. doi: 10.2307/1130576

Carter, D. B., & McCloskey, L. A. (1984). Peers and the maintenance of sex-typed behavior: The development of children's conceptions of cross-gender behavior in their peers. *Social Cognition*, *2*(4), 294–314. doi: 10.1521/soco.1984.2.4.294

Carter, D. B., & Patterson, C. J. (1982). Sex roles as social conventions: The development of children's conceptions of sex-role stereotypes. *Developmental Psychology*, *18*(6), 812–824. doi: 10.1037/0012-1649.18.6.812

Cha, E.-S., Kim, K. H., & Erlen, J. A. (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing*, *58*(4), 386–395. doi: 10.1111/j.1365-2648.2007.04242.x

Cherney, I. D., & Dempsey, J. (2010). Young children’s classification, stereotyping and play behaviour for gender neutral and ambiguous toys. *Educational Psychology*, *30*(6), 651–669. doi: 10.1080/01443410.2010.498416

Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles*, *54*(9–10), 717–726. doi: 10.1007/s11199-006-9037-8

Chiu, S. W., Gervan, S., Fairbrother, C., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., & Zucker, K. J. (2006). Sex-dimorphic color preference in children with Gender Identity Disorder: A comparison to clinical and community controls. *Sex Roles*, *55*(5–6), 385–395. doi: 10.1007/s11199-006-9089-9

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behavior and Development*, *23*(1), 113–118. doi: 10.1016/S0163-6383(00)00032-1

Constantinescu, M., & Hines, M. (2012). Relating prenatal testosterone exposure to postnatal behavior in typically developing children: Methods and findings. *Child Development Perspectives*, 407–413. doi: 10.1111/j.1750-8606.2012.00257.x

Davis, J., & Hines, M. (forthcoming). A meta-analysis of sex differences in children's toy preferences. Accepted for publication in *Archives of Sexual Behavior*.

De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Vision Research*, *33*(8), 1053–1065. doi: 10.1016/0042-6989(93)90240-W

Del Giudice, M. (2017). Pink, blue, and gender: An update. *Archives of Sexual Behavior*, *46*(6), 1–9. doi: 10.1007/s10508-017-1024-3

DeLucia, L. A. (1963). The toy preference test: A measure of sex-role identification. *Child Development*, *34*(1), 107–117. doi: 10.2307/1126831

DiDonato, Matthew D., England, D., Martin, C. L., & Amazeen, P. G. (2013). Dynamical analyses for developmental science: A primer for intrigued scientists. *Human Development*, *56*(1), 59–75. doi: 10.1159/000342936

DiDonato, M.D., Martin, C. L., Hessler, E. E., Amazeen, P. G., Hanish, L. D., & Fabes, R. A. (2012). Gender consistency and flexibility: Using dynamics to understand the relationship between gender and adjustment. *Nonlinear Dynamics, Psychology, and Life Sciences*, *16*(2), 159–184.

Dinella, L. M., Weisgram, E. S., & Fulcher, M. (2016). Children's gender-typed toy interests: Does propulsion matter? *Archives of Sexual Behavior*, *46*(5), 1295–1305. doi: 10.1007/s10508-016-0901-5

Doering, R. W., Zucker, K. J., Bradley, S. J., & MacIntyre, R. B. (1989). Effects of neutral toys on sex-typed play in children with gender identity disorder. *Journal of Abnormal Child Psychology*, *17*(5), 563–574. doi: 10.1007/BF00916514

Douglas, B. (2002). Christian citizens: Women and negotiations of modernity in Vanuatu. *The Contemporary Pacific*, *14*(1), 1-38.

Drewes, A. A. (2005). Play in selected cultures: Diversity and universality. *Cultural Issues in Play Therapy*, 26–71.

Duveen, G., & Lloyd, B. (1986). The significance of social identities. *British Journal of Social Psychology*, *25*(3), 219–230. doi: 10.1111/j.2044-8309.1986.tb00728.x

Eagly, A. H., & Wood, W. (1999). The origins of aggression sex differences: Evolved dispositions versus social roles. *Behavioral and Brain Sciences*, *22*(2), 223–224. doi: 10.1017/S0140525X99331812

Eagly, A. H., & Wood, W. (2013). The nature–nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science*, *8*(3), 340–357. doi: 10.1177/1745691613484767

Eaton, W. O., Von Bargen, D., & Keats, J. G. (1981). Gender understanding and dimensions of preschooler toy choice: Sex stereotype versus activity level. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *13*(3), 203.

Edwards, C. P. (2000). Children's play in cross-cultural perspective: A new look at the Six Cultures study. *Cross-Cultural Research*, *34*(4), 318–338.

Eisenberg, N., Murray, E., & Hite, T. (1982). Children's reasoning regarding sex-typed toy choices. *Child Development*, *53*(1), 81–86. doi: 10.2307/1129639

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Ellis, L., & Ficek, C. (2001). Color preferences according to gender and sexual orientation. *Personality and Individual Differences*, *31*(8), 1375–1379.

Escudero, P., Robbins, R. A., & Johnson, S. P. (2013). Sex-related preferences for real and doll faces versus real and toy objects in young infants and adults. *Journal of Experimental Child Psychology*, *116*(2), 367–379. doi: 10.1016/j.jecp.2013.07.001

Fagot, B. I. (1977). Consequences of moderate cross-gender behavior in preschool children. *Child Development*, *48*(3), 902–907. doi: 10.2307/1128339

Fagot, B. I. (1978). The influence of sex of child on parental reactions to toddler children. *Child Development*, *49*(2), 459–465. doi: 10.2307/1128711

Fagot, B. I. (1985). Changes in thinking about early sex role development. *Developmental Review*, *5*(1), 83–98.

Fagot, B. I., & Hagan, R. (1991). Observations of parent reactions to sex-stereotyped behaviors: Age and sex effects. *Child Development*, *62*(3), 617–628. doi: 10.2307/1131135

Fagot, B. I., Leinbach, M. D., & Hagan, R. (1986). Gender labeling and the adoption of sex-typed behaviors. *Developmental Psychology*, *22*(4), 440–443. doi: 10.1037/0012-1649.22.4.440

Fagot, B. I., Leinbach, M. D., & O'Boyle, C. (1992). Gender labeling, gender stereotyping, and parenting behaviors. *Developmental Psychology*, *28*(2), 225–230. doi: 10.1037/0012-1649.28.2.225

Fagot, B. I., & Patterson, G. R. (1969). An in vivo analysis of reinforcing contingencies for sex-role behaviors in the preschool child. *Developmental Psychology*, *1*(5), 563–568. doi: 10.1037/h0027965

Farran, S. (2004). Pigs, mats and feathers: Customary marriage in Vanuatu'. *Journal of Pacific Studies*, *27*(2), 245–276.

Fausto-Sterling, A. (2012a). Not your grandma's genetics: Some theoretical notes. *Psychology of Women Quarterly*, *36*(4), 411–418. doi: 10.1177/0361684312462016

Fausto-Sterling, A. (2012b). The dynamic development of gender variability. *Journal of Homosexuality*, *59*(3), 398–421.

Fausto-Sterling, A., Garcia Coll, C., & Lamarre, M. (2012). Sexing the baby: Part 2 applying dynamic systems theory to the emergences of sex-related differences in infants and toddlers. *Social Science & Medicine*, *74*(11), 1693–1702. doi: 10.1016/j.socscimed.2011.06.027

Fischer, J. L., & Fischer, A. (1966). *The New Englanders of Orchard Town, U.S.A.* New York: McGraw-Hill.

Franke, A., Caelli, T., & Hudson, R. J. (2004). Analysis of movements and behavior of caribou (Rangifer tarandus) using hidden Markov models. *Ecological Modelling*, *173*(2), 259–270.

Franklin, A., Gibbons, E., Chittenden, K., Alvarez, J., & Taylor, C. (2012). Infant color preference for red is not selectively context specific. *Emotion*, *12*(5), 1155–1160. doi: 10.1037/a0025333

Freeman, G., Sims, T., Kutsch, K., & Marcon, R. A. (1995). Linking gender-related toy preferences to social structure: Changes in children's letters to Santa since 1978. Presented at the *Southeastern Psychological Association*. Retrieved from http://eric.ed.gov/?id=ED383480

Freeman, N. K. (2007). Preschoolers' perceptions of gender appropriate toys and their parents' beliefs about genderized behaviors: Miscommunication, mixed messages, or hidden truths? *Early Childhood Education Journal*, *34*(5), 357–366. doi: 10.1007/s10643-006-0123-x

Fry, D. P. (1992). "Respect for the rights of others is peace": Learning aggression versus nonaggression among the Zapotec. *American Anthropologist*, *94*(3), 621–639.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. CRC Press.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.

Gilligan, C. (1982). *In a different voice*. Harvard University Press.

Gilligan, C. (1995). Hearing the difference: Theorizing connection. *Hypatia*, *10*(2), 120–127. doi: 10.1111/j.1527-2001.1995.tb01373.x

Gleser, L. J., & Olkin, I. O. (2007). *Stochastically dependent effect sizes*. Technical Report No. 2007-2, Department of Statistics, Stanford University.

Goldstein, J. H., Buckingham, D., & Brougère, G. (Eds.). (2004). *Toys, games, and media*. Mahwah, N.J: L. Erlbaum Associates.

Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental trajectories of sex-typed behavior in boys and girls: A longitudinal general population study of children aged 2.5–8 years. *Child Development*, *79*(5), 1583–1593. doi: 10.1111/j.1467-8624.2008.01207.x

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*(1), 3–32. doi: 10.1037/0033-295X.111.1.3

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, *10*(3), 281–287. doi: 10.1111/j.1467-7687.2007.00584.x

Green, V. A., Bigler, R., & Catherwood, D. (2004). The variability and flexibility of gender-typed toy play: a close look at children's behavioral responses to counterstereotypic models. *Sex Roles*, *51*(7–8), 371–386.

Gregory, R. J., & Gregory, J. E. (2002). Conflict resolution on Tanna, Vanuatu. *South Pacific Journal of Psychology*, *14*, 60–68.

Guastello, S. J., Koopmans, M., & Pincus, D. (eds). (2009). *Chaos and complexity in psychology: Theory of nonlinear dynamical systems*. New York: Cambridge University Press.

Haken, H., & Jumarie, G. (2006). *A macroscopic approach to complex system*. Springer.

Hassett, J. M., Siebert, E. R., & Wallen, K. (2008). Sex differences in rhesus monkey toy preferences parallel those of children. *Hormones and Behavior*, *54*(3), 359–364. doi: 10.1016/j.yhbeh.2008.03.008

Hellendoorn, J., & Harinck, F. J. H. (1997). War toy play and aggression in Dutch kindergarten children. *Social Development*, *6*(3), 341–354.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*(2–3), 111–135.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). Most people are not WEIRD. *Nature*, *466*(7302), 29–29.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010c). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. doi: 10.1017/S0140525X0999152X

Herzog, J. D. (1984). The socialization of juveniles in primate and foraging societies: implications for contemporary education. *Anthropology & Education Quarterly*, *15*(1), 70–79.

Hewlett, B. S., Fouts, H. N., Boyette, A. H., & Hewlett, B. L. (2011). Social learning among Congo Basin hunter-gatherers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1168–1178.

Hines, M. (2005). *Brain gender*. Oxford University Press.

Hines, M. (2006). Prenatal testosterone and gender-related behaviour. *European Journal of Endocrinology*, *155*(suppl 1), S115–S121. doi: 10.1530/eje.1.02236

Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences*, *14*(10), 448–456. doi: 10.1016/j.tics.2010.07.005

Hines, M. (2013). Gonadal hormone influences on human neurobehavioral development: Outcomes and mechanisms. In D. W. Pfaff & Y. Christen (Eds.), *Multiple origins of sex differences in brain* (pp. 59–69). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-33721-5_5

Hines, M., & Alexander, G. M. (2008). Monkeys, girls, boys and toys: A confirmation: Letter regarding “Sex differences in toy preferences: Striking parallels between monkeys and humans." *Hormones and Behavior*, *54*(3), 478–479. doi: 10.1016/j.yhbeh.2008.05.012

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

Hines, M., Golombok, S., Rust, J., Johnston, K. J., Golding, J., Parents, … others. (2002). Testosterone during pregnancy and gender role behavior of preschool children: a longitudinal, population study. *Child Development*, *73*(6), 1678–1687.

Hines, M., Spencer, D., Kung, K. T., Browne, W. V., Constantinescu, M., & Noorderhaven, R. M. (2016). The early postnatal period, mini-puberty, provides a window on the role of testosterone in human neurobehavioural development. *Current Opinion in Neurobiology*, *38*, 69–73. doi: 10.1016/j.conb.2016.02.008

Humphreys, A. P., & Smith, P. K. (1987). Rough and tumble, friendship, and dominance in schoolchildren: Evidence for continuity and change with age. *Child Development*, *58*(1), 201–212. doi: 10.2307/1130302

Hurlbert, A., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, *17*(16), R621–R623.

Hurlbert, A., & Owen, A. (2015). Biological, cultural, and developmental influences on color preferences. In A. J. Elliot, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of color psychology* (pp. 454–478). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107337930.023

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, *64*(6), 384–404.

Huston, A. C. (1985). The development of sex typing: Themes from recent research. *Developmental Review*, *5*(1), 1–17. doi: 10.1016/0273-2297(85)90028-0

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w^2 and d. *American Psychologist*, *36*(8), 892.

Idle, T., Wood, E., & Desmarais, S. (1993). Gender role socialization in toy play situations: Mothers and fathers with their sons and daughters. *Sex Roles*, *28*(11–12), 679–691.

Jacklin, C. N., Maccoby, E. E., & Dick, A. E. (1973). Barrier behavior and toy preference: sex differences (and their absence) in the year-old child. *Child Development*, *44*(1), 196–200. doi: 10.2307/1127703

Jackson, C. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software, 38*(8), 1-28. doi: 10.18637/jss.v038.i08

Jadva, V., Hines, M., & Golombok, S. (2010). Infants’ preferences for toys, colors, and shapes: Sex differences and similarities. *Archives of Sexual Behavior, 39*(6), 1261–1273. doi: 10.1007/s10508-010-9618-z

Joel, D., & Fausto-Sterling, A. (2016). Beyond sex differences: new approaches for thinking about variation in brain structure and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1688), 20150451. doi: 10.1098/rstb.2015.0451

Jones, A., & Glenn, S. M. (1990). Gender differences in pretend play in a primary school group. *Early Child Development and Care*, *72*, 61–67. doi: 10.1080/0300443910720105

Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, *52*(2), 93–100.

Knickmeyer, R. C., Wheelwright, S., Taylor, K., Raggatt, P., Hackett, G., & Baron-Cohen, S. (2005). Gender-typed play and amniotic testosterone. *Developmental Psychology*, *41*(3), 517–528. doi: 10.1037/0012-1649.41.3.517

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82-173). Stanford, CA: Stanford University Press.

Kohlberg, L. (1976). Psychoanalytic and Cognitive-Developmental Approaches to Moral Education. Presented at the *AERA Symposium on Psychoanalysis and Education*.

Koller, V. (2008). `Not just a colour’: Pink as a gender and sexuality marker in visual communication. *Visual Communication*, *7*(4), 395–423. doi: 10.1177/1470357208096209

Laland, K. N., & Brown, G. R. (2006). Niche construction, human behavior, and the adaptive-lag hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews, 15*(3), 95–104. doi: 10.1002/evan.20093

Lamb, M. E., Sternberg, K. J., Hwang, C.-P., & Broberg, A. G. (2014). *Child care in context: Cross-cultural perspectives*. Psychology Press.

Lamminmäki, A., Hines, M., Kuiri-Hänninen, T., Kilpeläinen, L., Dunkel, L., & Sankilampi, U. (2012). Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. *Hormones and Behavior*, *61*(4), 611–616. doi: 10.1016/j.yhbeh.2012.02.013

Langlois, J. H., & Downs, A. C. (1980). Mothers, fathers, and peers as socialization agents of sex-typed play behaviors in young children. *Child Development*, *51*(4), 1237–1247. doi: 10.2307/1129566

Leaper, C. (2015). Chapter 19: Gender and social-cognitive development. In *Handbook of child psychology and developmental science* (7th ed., Vol. 2). Wiley.

LeVay, S. (2011). From mice to men: Biological factors in the development of sexuality. *Frontiers in Neuroendocrinology*, *32*(2), 110–113. doi: 10.1016/j.yfrne.2011.02.002

LeVine, R. A., & Lloyd, B. B. (1966). *Nyansongo, a Gusii community in Kenya*. Huntington, N.Y: Krieger.

LeVine, R. A. (2007). Ethnographic studies of childhood: A historical overview. *American Anthropologist*, *109*(2), 247-260.

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*(11), 841–845.

Liben, L. S., & Bigler, R. (2002). The developmental course of gender differentiation. *Monographs of the Society for Research in Child Development*, *67*(2), Serial No. 269, 1–147.

Lindstrom, L. (1983). Say what? Language and political boundaries on Tanna (Vanuatu). *Anthropological Linguistics*, 387–403.

Lobel, T. E., & Menashri, J. (1993). Relations of conceptions of gender-role transgressions and gender constancy to gender-typed toy preferences. *Developmental Psychology*, *29*(1), 150–155. doi: 10.1037/0012-1649.29.1.150

LoBue, V., & DeLoache, J. S. (2011). Pretty in pink: The early development of gender-stereotyped colour preferences: Children's colour preferences. *British Journal of Developmental Psychology*, *29*(3), 656–667. doi: 10.1111/j.2044-835X.2011.02027.x

Maccoby, E. E. (1966). *The development of sex differences*. Stanford University Press.

Malinowski, B. (2013). *Sex and repression in savage society: [1927]*. Routledge.

Maretzki, T. W., & Maretzki, H. (1966). *Taira: an Okinawan village*. New York: Wiley.

Marshall, G., & Jones, R. H. (1995). Multistate models and diabetic retinopathy. *Statistics in Medicine*, *14*, 1975-1983.

Martin, C. L., & Dinella, L. M. (2012). Congruence between gender stereotypes and activity preference in self-identified tomboys and non-tomboys. *Archives of Sexual Behavior*, *41*(3), 599–610. doi: 10.1007/s10508-011-9786-5

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, *52*(4), 1119–1134. doi: 10.2307/1129498

Martin, C. L., Kornienko, O., Schaefer, D. R., Hanish, L. D., Fabes, R. A., & Goble, P. (2013). The role of sex of peers and gender-typed activities in young children's peer affiliative networks: A longitudinal analysis of selection and influence. *Child Development*, *84*(3), 921–937. doi: 10.1111/cdev.12032

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, *13*(2), 67–70.

Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology*, *61*(1), 353–381. doi: 10.1146/annurev.psych.093008.100511

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*(6), 903–933. doi: 10.1037/0033-2909.128.6.903

McCarthy, M. M. (2010). How it's made: Organisational effects of hormones on the developing brain. *Journal of Neuroendocrinology*, *22*(7), 736–742. doi: 10.1111/j.1365-2826.2010.02021.x

Merler, M., Cao, L., & Smith, J. R. (2015). You are what you tweet #x2026;pic! gender prediction based on semantic analysis of social media images. In *2015 IEEE international conference on multimedia and expo (ICME)* (pp. 1–6). doi: 10.1109/ICME.2015.7177499

Miller, C. L. (1987). Qualitative differences among gender-stereotyped toys: Implications for cognitive and social development in girls and boys. *Sex Roles*, *16*(9–10), 473–487. doi: 10.1007/BF00292482

Miller, J. G. (1994). Cultural diversity in the morality of caring: Individually oriented versus duty-based interpersonal moral codes. *Cross-Cultural Research*, *28*(1), 3–39. doi: 10.1177/106939719402800101

Minturn, L., & Hitchcock, J. T. (1966). *The Rājpūts of Khalapur, India.* New York: Wiley.

Mischel, W. (1966). A social learning view of sex differences in behavior. In E. Maccoby (Ed.), *The development of sex differences* (pp. 57-81). Stanford, C.A.: Stanford University Press.

Moller, L. C., & Serbin, L. A. (1996). Antecedents of toddler gender segregation: Cognitive consonance, gender-typed toy preferences and behavioral compatibility. *Sex Roles*, *35*(7), 445–460.

Money, J. (1973). Gender role, gender identity, core gender identity: Usage and definition of terms. *Journal of American Academy of Psychoanalysis*, *1*(4), 397–402.

Mori, S., Nakamoto, H., Mizuochi, H., Ikudome, S., & Gabbard, C. (2013). Influence of affordances in the home environment on motor development of young children in Japan. *Child Development Research*, *2013*, e898406. doi: 10.1155/2013/898406

Murdock, G. P. (1967). Ethnographic atlas: A summary. *Ethnology*, *6*(2), 109–236.

Nelson, A. (2005). Children's toy collections in Sweden--A less gender-typed country? *Sex Roles*, *52*(1–2), 93–102. doi: 10.1007/s11199-005-1196-5

Nielsen, M., Tomaselli, K., Mushin, I., & Whiten, A. (2014). Exploring tool innovation: A comparison of Western and Bushman children. *Journal of Experimental Child Psychology*, *126*, 384–394.

Nordenström, A., Servin, A., Bohlin, G., Larsson, A., & Wedell, A. (2002). Sex-typed toy play behavior correlates with the degree of prenatal androgen exposure assessed by CYP21 genotype in girls with Congenital Adrenal Hyperplasia. *The Journal of Clinical Endocrinology & Metabolism*, *87*(11), 5119–5124. doi: 10.1210/jc.2001-011531

Nydegger, W. F., & Nydegger, C. (1966). *Tarong: An Ilocos barrio in the Philippines*. New York: John Wiley and Sons, Inc.

O'Brien, M., & Huston, A. C. (1985). Development of sex-typed play behavior in toddlers. *Developmental Psychology*, *21*(5), 866–871. doi: 10.1037/0012-1649.21.5.866

O'Brien, M., Huston, A. C., & Risley, T. R. (1983). Sex-typed play of toddlers in a day care center. *Journal of Applied Developmental Psychology*, *4*(1), 1–9. doi: 10.1016/0193-3973(83)90054-0

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., … Slaughter, V. (2016). Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. *Current Biology*, *26*(10), 1334–1338. doi: 10.1016/j.cub.2016.03.047

Ou, L.-C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part II: Colour emotions for two-colour combinations. *Color Research & Application*, *29*(4), 292–298. doi: 10.1002/col.20024

Overton, W. F. (2007). A coherent metatheory for dynamic systems: Relational organicism-contextualism. *Human Development*, *50*(2–3), 154–159.

Overton, W. F., & Lerner, R. M. (2014). Fundamental concepts and methods in developmental science: A relational perspective. *Research in Human Development*, *11*(1), 63–73.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877–8882. doi: 10.1073/pnas.0906172107

Pasterski, V., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2011). Prenatal hormones and childhood sex segregation: Playmate and play style preferences in girls with congenital adrenal hyperplasia. *Hormones and Behavior*, *59*(4), 549–555. doi: 10.1016/j.yhbeh.2011.02.007

Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development*, *76*(1), 264–278. doi: 10.1111/j.1467-8624.2005.00843.x

Patterson, T. A., Basson, M., Bravington, M. V., & Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, *78*(6), 1113–1123.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution*, *23*(2), 87–94.

Person, E. S., & Ovesey, L. (1983). Psychoanalytic theories of gender identity. *Journal of the American Academy of Psychoanalysis*, *11*(2), 203.

Phoenix, C. H., Goy, R. W., Gerall, A. A., & Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology*, *65*(3), 369–382.

Pike, J. J., & Jennings, N. A. (2005). The effects of commercials on children's perceptions of gender appropriate toy use. *Sex Roles*, *52*(1–2), 83–91. doi: 10.1007/s11199-005-1195-6

Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles*, *22*(5–6), 359–367. doi: 10.1007/BF00288339

Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior*, *31*(3), 157–175. doi: 10.1016/j.evolhumbehav.2010.02.005

Rekers, D. G. A., & Yates, C. E. (1976). Sex-typed play in feminoid boys versus normal boys and girls. *Journal of Abnormal Child Psychology*, *4*(1), 1–8. doi: 10.1007/BF00917600

Richardson, J. G., & Simpson, C. H. (1982). Children, gender, and social structure: An analysis of the contents of letters to Santa Claus. *Child Development*, *53*(2), 429–436. doi: 10.2307/1128986

Richert, R. A., Boyatzis, C. J., & King, P. E. (2017). Introduction to the British Journal of Developmental Psychology special issue on religion, culture, and development. *British Journal of Developmental Psychology*, *35*(1), 1–3. doi: 10.1111/bjdp.12179

Rogers, L. O., Scott, M. A., & Way, N. (2015). Racial and gender identity among Black adolescent males: An intersectionality perspective. *Child Development*, *86*(2), 407–424.

Romney, A. K., & Romney, R. (1966). *The Mixtecans of Juxtlahuaca, Mexico*. New York: Wiley.

Roopnarine, J. L., & Johnson, J. E. (1994). The need to look at play in diverse cultural settings. In J. L. Roopnarine, J. E. Johnson, & F. H. Hooper (Eds.), *Children's play in diverse cultures* (pp. 1–8). SUNY Press.

Roopnarine, J. L., Nathan, P., & Pellegrini, A. D. (2010). Cultural variations in beliefs about play, parent–child play, and children's play. In *The Oxford handbook of the development of play*. Oxford Handbooks Online.

Salk, R. H., & Hyde, J. S. (2012). Contemporary genetics for gender researchers: Not your grandma's genetics anymore. *Psychology of Women Quarterly*, *36*(4), 395–410.

Schau, C. G., Kahn, L., Diepold, J. H., & Cherry, F. (1980). The Relationships of Parental Expectations and Preschool Children's Verbal Sex Typing to Their Sex-Typed Toy Play Behavior. *Child Development*, *51*(1), 266–270. doi: 10.2307/1129620

Schliehe-Diecks, S., Kappeler, P. M., & Langrock, R. (2012). On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus,* rsfs20110077.

Schmitt, D. P., & Pilcher, J. J. (2004). Evaluating evidence of psychological adaptation: How do we know one when we see one? *Psychological Science, 15*(10), 643–649.

Semin, G. R., & Palma, T. A. (2014). Why the bride wears white: Grounding gender with brightness. *Journal of Consumer Psychology, 24*(2), 217–225. doi: 10.1016/j.jcps.2013.09.003

Serbin, L. A., Connor, J. M., Burchardt, C. J., & Citron, C. C. (1979). Effects of peer presence on sex-typing of children's play behavior. *Journal of Experimental Child Psychology*, *27*(2), 303–309. doi: 10.1016/0022-0965(79)90050-X

Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, *25*(1), 7–15. https://doi.org/10.1080/01650250042000078

Servin, A., Nordenström, A., Larsson, A., & Bohlin, G. (2003). Prenatal androgens and gender-typed behavior: A study of girls with mild and severe forms of congenital adrenal hyperplasia. *Developmental Psychology*, *39*(3), 440–450. doi: 10.1037/0012-1649.39.3.440

Shields, S. A. (2008). Gender: An intersectionality perspective. *Sex Roles*, *59*(5–6), 301–311. doi: 10.1007/s11199-008-9501-8

Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, *46*(4), 849–856. doi: 10.2307/1128389

Sorokowski, P., Sorokowska, A., & Witzel, C. (2014). Sex differences in color preferences transcend extreme differences in culture and ecology. *Psychonomic Bulletin & Review*, *21*(5), 1195–1201. doi: 10.3758/s13423-014-0591-8

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*, *126*, S124–S128. doi: 10.1053/j.gastro.2003.10.016

Spiro, M. (1984). Psychoanalysis and cultural relativism: The Trobriand case. *Emotions & Behavior Monographs*.

Stangor, C., & Ruble, D. N. (1989). Differential influences of gender schemata and gender constancy on children's information processing and behavior. *Social Cognition*, *7*(4), 353–372.

Stern, M., & Karraker, K. H. (1989). Sex stereotyping of infants: A review of gender labeling studies. *Sex Roles*, *20*(9–10), 501–522. doi: 10.1007/BF00288198

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development*, *9*(4), 545–569.

Sweet, E. V. (2013). *Boy builders and pink princesses: Gender, toys, and inequality over the twentieth century* (Ph.D.). University of California, Davis, United States -- California.

Takeuchi, M. (1994). Children's play in Japan. In J. L. Roopnarine, J. E. Johnson, & F. H. Hooper (Eds.), *Children's play in diverse cultures* (pp. 51–72). SUNY Press.

Taylor, C., Clifford, A., & Franklin, A. (2013). Color preferences are not universal. *Journal of Experimental Psychology: General*, *142*(4), 1015–1027. doi: 10.1037/a0030273

Taylor, C., Schloss, K., Palmer, S. E., & Franklin, A. (2013). Color preferences in infants and adults are different. *Psychonomic Bulletin & Review*, *20*(5), 916–922. https://doi.org/10.3758/s13423-013-0411-6

Teller, D. Y., Civan, A., & Bronson-Castain, K. (2004). Infants' spontaneous color preferences are not due to adult-like brightness variations. *Visual Neuroscience*, *21*(03), 397–401. doi: 10.1017/S0952523804213360

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318. doi: 10.1016/j.tics.2006.05.009

Tinbergen, N. (1963). On aims and methods of ethology. *Ethology*, *20*(4), 410–433.

Trumble, B. C., Gaulin, S. J. C., Dunbar, M. D., Kaplan, H., & Gurven, M. (2016). No sex or age difference in dead-reckoning ability among Tsimane forager-horticulturalists. *Human Nature*, *27*(1), 51–67. doi: 10.1007/s12110-015-9246-3

Tudge, J. R. H., Brown, J., & Freitas, L. B. L. (2010). The cultural ecology of play: Methodological considerations for studying play in its everyday contexts. In P. Nathan & A. D. Pellegrini (Eds.), *The Oxford handbook of the development of play*. Oxford University Press.

Vaisman, C. L. (2016). Pretty in pink vs pretty in black: Blogs as gendered avatars. *Visual Communication*, *15*(3), 293–315. doi: 10.1177/1470357216643909

van de Beek, C., van Goozen, S. H., Buitelaar, J. K., & Cohen-Kettenis, P. T. (2009). Prenatal sex hormones (maternal and amniotic fluid) and gender-related play behavior in 13-month-old infants. *Archives of Sexual Behavior*, *38*(1), 6–15. doi: 10.1007/s10508-007-9291-z

van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Development Perspectives*, *5*(4), 273–278.

Vanuatu Women's Centre (2014). *Vanuatu Women's Centre program against violence against women*. Progress report 3, Phase 6. Port Vila, Vanuatu: Vanuatu Women's Centre.

Vashro, L., & Cashdan, E. (2015). Spatial cognition, mobility, and reproductive success in northwestern Namibia. *Evolution and Human Behavior*, *36*(2), 123–129. doi: 10.1016/j.evolhumbehav.2014.09.009

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. doi: 10.1037/0033-2909.117.2.250

Wallen, K. (1996). Nature needs nurture: The interaction of hormonal and social influences on the development of behavioral sex differences in rhesus monkeys. *Hormones and Behavior*, *30*(4), 364–378. doi: 10.1006/hbeh.1996.0042

Wallen, K. (2009). The Organizational Hypothesis: Reflections on the 50th anniversary of the publication of Phoenix, Goy, Gerall, and Young (1959). *Hormones and Behavior*, *55*(5), 561–565. doi: 10.1016/j.yhbeh.2009.03.009

Weinraub, M., Clemens, L. P., Sockloff, A., Ethridge, T., Gracely, E., & Myers, B. (1984). The development of sex role stereotypes in the third year: relationships to gender labeling, gender identity, sex-types toy preference, and family characteristics. *Child Development*, *55*(4), 1493–1503. https://doi.org/10.2307/1130019

Weisgram, E. S., Fulcher, M., & Dinella, L. M. (2014). Pink gives girls permission: Exploring the roles of explicit gender labels and gender-typed colors on preschool children's toy preferences. *Journal of Applied Developmental Psychology*, *35*(5), 401–409.

Wells, D. L., McDonald, C. L., & Ringland, J. E. (2008). Color preferences in gorillas (*Gorilla gorilla gorilla*) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *122*(2), 213–219. doi: 10.1037/0735-7036.122.2.213

Whitam, F. L., & Mathy, R. M. (1991). Childhood cross-gender behavior of homosexual females in Brazil, Peru, the Philippines, and the United States. *Archives of Sexual Behavior*, *20*(2), 151–170.

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. doi: 10.1098/rstb.2009.0069

Whiting, J. W. M., Child, I. L., Lambert, W. W., & others. (1966). *Field guide for a study of socialization.* New York: Wiley.

Wilansky-Traynor, P., & Lobel, T. E. (2008). Differential effects of an adult observer's presence on sex-typed play behavior: A comparison between gender-schematic and gender-aschematic preschool children. *Archives of Sexual Behavior*, *37*(4), 548–557. doi: 10.1007/s10508-008-9342-0

Williams, C. L., & Pleil, K. E. (2008). Toy story: Why do monkey and human males prefer trucks? *Hormones and Behavior*, *54*(3), 355–358. doi: 10.1016/j.yhbeh.2008.05.003

Witkin, H. A. (1979). Socialization, culture and ecology in the development of group and Sex differences in cognitive style. *Human Development*, *22*(5), 358–372.

Wong, W. I., & Hines, M. (2014). Effects of gender color-coding on toddlers' gender-typical toy play. *Archives of Sexual Behavior*, *44*(5), 1–10. doi: 10.1007/s10508-014-0400-5

Wong, W. I., & Hines, M. (2015). Preferences for pink and blue: The development of color preferences as a distinct gender-typed behavior in toddlers. *Archives of Sexual Behavior*, 1–12.

Wong, W. I., Pasterski, V., Hindmarsh, P. C., Geffner, M. E., & Hines, M. (2013). Are there parental socialization effects on the sex-typed behavior of individuals with Congenital Adrenal Hyperplasia? *Archives of Sexual Behavior*, *42*(3), 381–391. doi: 10.1007/s10508-012-9997-4

Yokosawa, K., Schloss, K. B., Asano, M., & Palmer, S. E. (2016). Ecological effects in cross-cultural differences between U.S. and Japanese color preferences. *Cognitive Science*, *40*(7), 1590–1616. doi: 10.1111/cogs.12291

Yoshikawa, H., & Hsueh, J. (2001). Child development and public policy: Toward a dynamic systems perspective. *Child Development*, *72*(6), 1887–1903.

Zemach, I., Chang, S., & Teller, D. Y. (2007). Infant color vision: Prediction of infants' spontaneous color preferences. *Vision Research*, *47*(10), 1368–1381. doi: 10.1016/j.visres.2006.09.024

Zentner, M. R. (2001). Preferences for colours and colour--emotion combinations in early childhood. *Developmental Science*, *4*(4), 389–398. doi: 10.1111/1467-7687.00180

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology*, *45*(3), 688–701. doi: 10.1037/a0014053

Appendices

Appendix 1: Colour Preferences

Frequency Tables and Charts for Planned Comparisons

Table A1.1. Frequency of Boys' and Girls' Preferences for Pink vs Blue

Brisbane City

	Pink	Blue	Binomial
Girls	13	7	.383
Boys	10	20	.099
Total	23	27	.672

OR: 0.26 (0.06 – 1.03), Fisher's exact *p* = .039

Shipibo

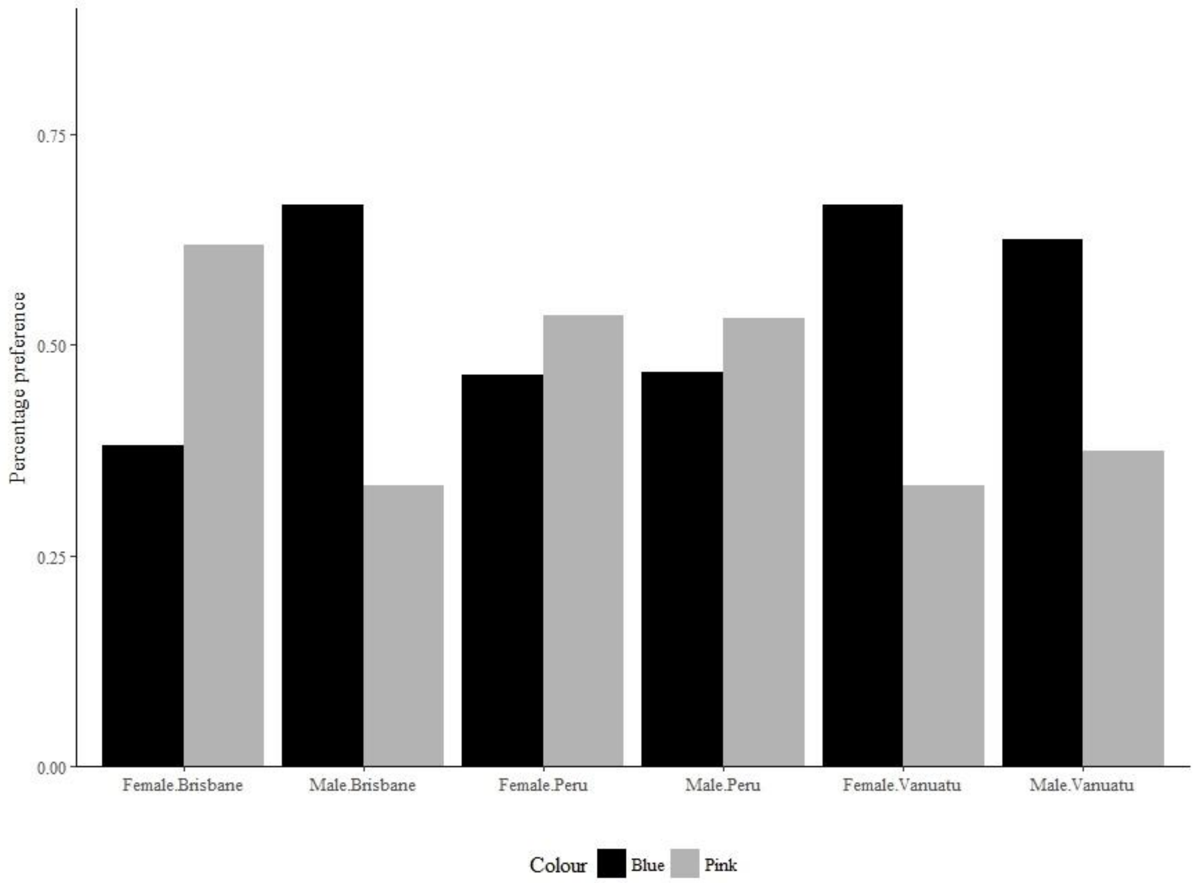
	Pink	Blue	Binomial
Girls	30	26	.689
Boys	25	22	.771
Total	55	48	.555

OR: 1.00 (0.42 – 2.39), Fisher's exact *p* > .999

Navhal

	Pink	Blue	Binomial
Girls	7	14	.189
Boys	12	20	.215
Total	19	24	.542

OR: 1.14 (0.30 – 4.48), Fisher's exact *p* > .999



Preferences for pink and blue in male and female children in Brisbane (Australia), Shipibo (Peru), and Navhal (Vanuatu) samples

Table A1.2. Frequency of Boys' and Girls' Preferences for Red vs Blue

Brisbane City

	Red	Blue	Binomial
Girls	11	10	.999
Boys	11	20	.201
Total	22	30	.332

OR: 0.48 (0.12 – 1.79), Fisher's exact p = .246

Shipibo

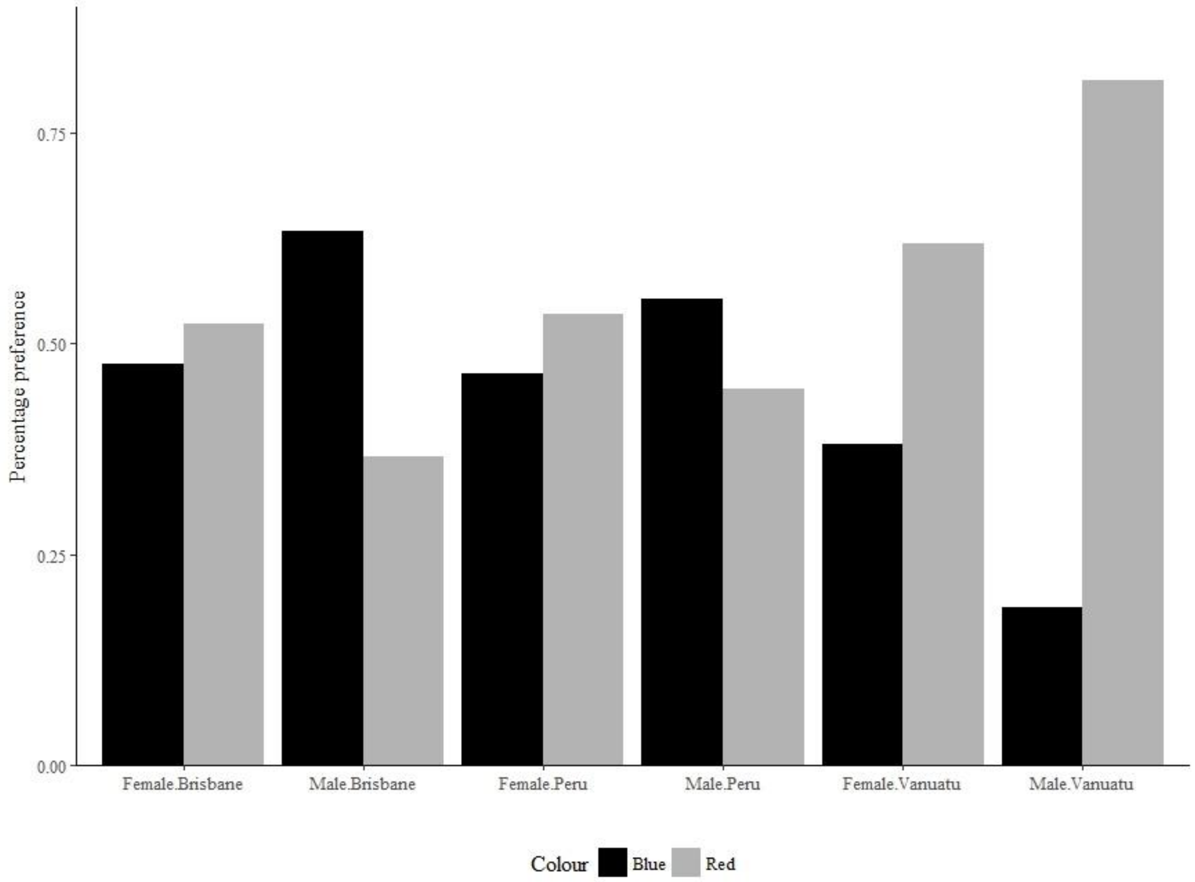
	Red	Blue	Binomial
Girls	30	26	.689
Boys	21	26	.560
Total	51	52	>.999

OR: 0.70 (0.29 – 1.67), Fisher's exact p = .423

Navhal

	Red	Blue	Binomial
Girls	13	8	.383
Boys	26	6	<.001
Total	39	14	.001

OR = 2.93 (0.58 – 16.80), Fisher's exact p = .157



Preferences for red and blue in male and female children in Brisbane (Australia), Shipibo (Peru), and Navhal (Vanuatu) samples

Table A1.3. Frequency of Boys' and Girls' Preferences for Pink vs Light Blue

Brisbane City

	Pink	Light blue	Binomial
Girls	3	18	.001
Boys	5	25	<.001
Total	8	43	<.001

OR: 1.16 (0.19 – 8.52), Fisher's exact $p > .999$

Shipibo

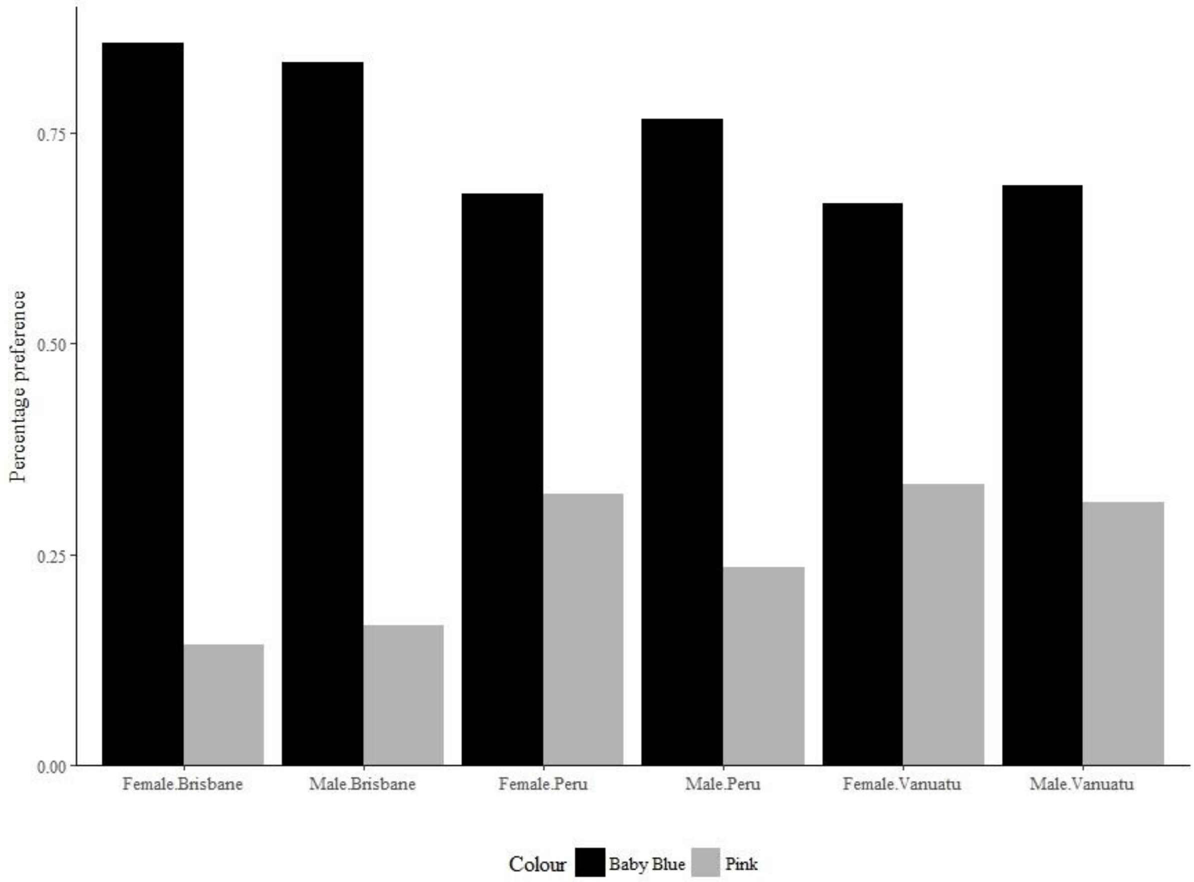
	Pink	Light blue	Binomial
Girls	18	38	.010
Boys	11	36	<.001
Total	29	74	<.001

OR: 0.65 (0.24 – 1.70), Fisher's exact $p = .380$

Navhal

	Pink	Light blue	Binomial
Girls	7	14	.189
Boys	10	22	.050
Total	17	36	.013

OR: 0.86 (0.22 – 3.44), Fisher's exact $p > .999$



Preferences for pink and light blue in male and female children in Brisbane (Australia), Shipibo (Peru), and Navhal (Vanuatu) samples

Appendix 2: Toy Preferences

Summary Statistics

Brisbane City, Australia

	Boys	Girls	*d*	*p*	
Boy-type toys	**43.18**	**20.99**	**0.88**	**<.001**	*
Racing car	8.73	0.85	0.80	.001	*
Action figure	53.38	7.99	1.60	<.001	*
Pirate costume	11.64	6.05	0.36	.099	
Sword	6.15	2.26	0.44	.052	
Girl-type toys	**22.49**	**39.59**	**-0.69**	**.006**	*
Pony carriage	1.91	6.18	-0.42	.078	
Barbie doll	0.49	21.55	-0.98	<.001	*
Princess costume	3.73	13.61	-0.57	.014	*
Baby doll	0.40	2.19	-0.40	.099	
Neutral toys	**34.33**	**39.42**	**-0.17**	**.482**	
Animals	3.40	4.47	-0.08	.724	
Pencils	8.99	30.17	-0.72	.003	*
Dinosaur costume	0.95	4.53	-0.55	.022	*
Stuffed dog	0.23	0.15	0.09	.682	

Table A2.1. Preferences for Sex-Typed Toys in Brisbane City, Australia. Numbers in Boys and Girls columns are mean proportion of play time. *d* is the standardised mean difference effect size for the sex difference. Positive sign indicates boys>girls, negative sign indicates girls>boys. *p* is the exact *p* value of the mean difference. Asterisk indicates statistically significant difference at $p<.05$.

Shipibo Villages, Peru

	Boys	Girls	*d*	*p*	
Boy-type toys	**76.49**	**14.72**	**2.87**	**<.001**	*****
Racing car	38.70	5.46	1.39	<.001	*
Action figure	33.56	6.18	1.13	<.001	*
Pirate costume	1.05	2.63	-0.43	.032	*
Sword	3.19	0.45	0.58	.014	*
Girl-type toys	**4.55**	**75.87**	**-3.50**	**<.001**	*****
Pony carriage	3.46	11.11	-0.58	.004	*
Barbie doll	0.50	27.65	-1.24	<.001	*
Princess costume	0.12	7.23	-1.07	<.001	*
Baby doll	0.47	29.87	-1.35	<.001	*
Neutral toys	**18.96**	**9.41**	**0.58**	**.008**	*****
Animals	17.75	6.54	0.69	.002	*
Pencils	0.15	0.91	-0.40	.039	*
Dinosaur costume	0.54	0.52	0.02	.933	
Stuffed dog	0.52	1.44	-0.43	.034	*

Table A2.2. Preferences for Sex-Typed Toys in Shipibo Villages, Peru. Numbers in Boys and Girls columns are mean proportion of play time. d is the standardised mean difference effect size for the sex difference. Positive sign indicates boys>girls, negative sign indicates girls>boys. p is the exact p value of the mean difference. Asterisk indicates statistically significant difference at $p<.05$.

Lenakel Town, Vanuatu

	Boys	Girls	*d*	*p*	
Boy-type toys	**55.87**	**29.19**	**1.04**	**<.001**	*****
Racing car	12.16	10.34	0.12	.674	*
Action figure	35.74	10.62	1.04	<.001	*
Pirate costume	2.28	5.11	-0.47	.094	
Sword	5.69	3.12	0.42	.130	
Girl-type toys	**14.98**	**56.67**	**-2.00**	**<.001**	*****
Pony carriage	6.28	8.31	-0.25	.372	
Barbie doll	4.64	32.84	-1.55	<.001	*
Princess costume	1.93	12.35	-1.06	<.001	*
Baby doll	2.13	3.17	-0.20	.467	
Neutral toys	**29.16**	**14.14**	**0.71**	**.013**	*****
Animals	27.22	10.88	0.77	.008	*
Pencils	1.23	0.93	0.11	.679	
Dinosaur costume	0.31	1.41	-0.52	.071	
Stuffed dog	0.40	0.92	-0.31	.279	

Table A2.3. Preferences for Sex-Typed Toys in Lenakel Town, Vanuatu. Numbers in Boys and Girls columns are mean proportion of play time. d is the standardised mean difference effect size for the sex difference. Positive sign indicates boys>girls, negative sign indicates girls>boys. p is the exact p value of the mean difference. Asterisk indicates statistically significant difference at p<.05.

Navhal *Kastom* Villages, Vanuatu

	Boys	**Girls**	***d***	***p***	
Boy-type toys	**43.18**	**20.99**	**0.88**	**<.001**	*****
Racing car	14.39	12.66	0.09	.698	
Action figure	23.02	7.25	0.70	.001	*
Pirate costume	1.71	0.56	0.30	.121	
Sword	4.06	0.52	0.59	.002	*
Girl-type toys	**22.49**	**39.59**	**-0.69**	**.006**	*****
Pony carriage	10.08	5.76	0.28	.167	
Barbie doll	3.06	12.38	-0.87	.003	*
Princess costume	3.10	10.53	-0.63	.030	*
Baby doll	6.26	10.93	-0.29	.255	
Neutral toys	**34.33**	**39.42**	**-0.17**	**.482**	
Animals	30.39	31.35	-0.03	.900	
Pencils	1.01	0.96	0.01	.952	
Dinosaur costume	0.68	0.33	0.19	.343	
Stuffed dog	2.24	6.78	-0.47	.087	

Table A2.4. Preferences for Sex-Typed Toys in Navhal Kastom Villages, Vanuatu. Numbers in Boys and Girls columns are mean proportion of play time. d is the standardised mean difference effect size for the sex difference. Positive sign indicates boys>girls, negative sign indicates girls>boys. p is the exact p value of the mean difference. Asterisk indicates statistically significant difference at p<.05.