A Developmental Examination of the Conceptual Structure of Animal, Artifact, and Human Social Categories

by

Marjorie Rhodes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2009

Doctoral Committee:

Professor Susan A. Gelman, Chair
Professor Shinobu Kitayama
Professor Henry Wellman
Associate Professor Barbra Meek
Research Scientist E. Margaret Evans

Acknowledgements

I would very much like to thank Susan Gelman, for collaborating with me on this research, and for being a fantastic adviser throughout graduate school. I would also like to thank the other members of my committee, Henry Wellman, E. Margaret Evans, Shinobu Kitayama, and Barbra Meek, as well as the members of the Cognitive Development and Language Lab, for their input into this work.

I am also very grateful to the research assistants who enthusiastically helped me collect, enter, and code these data: Holly Keilch, Erica Ranade, Korie Zink, Alissa Koloff, Christine Shenouda, Kimberly Canter, Dillon Prefer, Sam Kolkey, and Amanda Karp. I would also like to extend many thanks to the teachers and administrators who allowed me to visit their schools, and who provided so much support for participant recruitment. For help with schools and recruitment, I am especially grateful to Principal Patrick O'Neill. This type of research would also not be possible without the participation of parents and children, and so I am deeply grateful to the many families who generously donated their time to participate in these projects.

Finally, I would like to thank Daniel Brickman, for his invaluable insight into all of my work, and my family, for their constant support.

Table of Contents

List of Tables

List of Appendices

Abstract

This dissertation examines how the ontological status that people attribute to categories varies by domain, age, and cultural context. The first four studies examined whether categories of animals, artifacts, and people are represented as objectively-defined natural kinds, or alternately, as subjective conventionalized groupings. Participants included children (ages 5, 7, 10, and 17) drawn from two cultural contexts. Results demonstrated that abstract domain-specific ontological beliefs emerge early in development; in each age-group, children represented animal categories as natural kinds, but artifact categories as more conventionalized.

For human social categories, beliefs about naturalness and conventionality were predicted by interactions between the type of category (i.e., gender vs. race), cultural context, and age. Younger children, in both communities, viewed gender as a natural kind, but race categories as more conventionalized. The concepts of older children were found to vary by cultural context. Older children from an ethnically-homogeneous and socially-conservative environment viewed both types of social categories as natural kinds, whereas children from a more diverse environment viewed social categorization as flexible and subjective. Control conditions confirmed that domain effects in beliefs about category naturalness could not be attributed to perceptual features of the stimuli, category knowledge, category salience, or linguistic properties of the task.

The fifth study extended these findings to younger children. This work documented that three- and four- year-olds also conceptualize animal species and gender categories as natural kinds, but artifact and race categories as more subjective and flexible. The sixth study documented that the effects of age and cultural context on representations of social categories also extend to children's beliefs about how social categories influence individual behavior.

Overall, results from this dissertation suggest that abstract beliefs about ontology are incorporated into children's categories from at least the age of three, as well as that the development of social categorization involves complex interactions between intuitive biases and cultural input. Implications for the origins of social categories and theories of conceptual development are discussed.

Chapter 1

Introduction

An important developmental achievement is that of organizing experience into categories (e.g., tigers, chairs, girls). Categories enable us to store information efficiently and to generalize and extend knowledge in new ways. To develop functional adult-like concepts, children must learn not only how objects are commonly classified, but also what kind of meaning to attribute to categories. The goal of the present set of studies was to examine the fundamental meaning that individuals affix to categories in various domains across development. Specifically, I examine developmental and cultural influences on whether particular categories are understood as reflecting the objective, natural structure of the world (i.e., as natural kinds) or as subjective groupings that are dependent on convention (i.e., as conventionalized categories).

The question of whether categories reflect objectively-determined, natural structure or subjective, flexible conventions has been the subject of rich philosophical discussion about the relations among representation, language, and nature (Putnam, 1975; Quine, 1977; Schwartz, 1979; Wilson, 1999). With regard to cognitive psychology, this distinction provides an organizing framework for examining people's *beliefs* about category structure and how concepts vary across domains of knowledge. Thus, the present research examines the meaning that people attribute to categories and category

boundaries, not the metaphysical status of the categories themselves. Previous work indicates that adult concepts include at least two distinct interpretations of the meaning of category boundaries, which are often applied in domain-specific patterns (Diesendruck & Gelman, 1999; Kalish, 2002; Malt, 1990). Specifically, some categories are understood as marking objectively correct, real distinctions that exist in the world (e.g., between dogs and cats), whereas others are understood as marking subjective and flexible distinctions that are decided upon by humans, either according to social convention (e.g., bowls vs. plates) or in response to pragmatic concerns (e.g., goal-derived categories).

In the present research, I examined three questions that arise from considering the differentiated nature of adult conceptual structure. First, I examined whether young children, like adults, view some everyday categories as natural kinds and others as conventionalized groups. Second, I examined whether children share adults' intuitions that category meaning varies by domain, by comparing concepts of animal species, artifacts, and human social categories. Third, I examined the extent to which concepts of animals, artifacts, and social categories vary across cultural contexts, by including participants from two cultural communities. Thus, this work informs theoretical questions regarding the domain-specificity of human concepts, whether young children's concepts differ fundamentally from those held by adults, as well as the extent to which conceptual development depends on cultural input.

The proposal that adults draw principled distinctions between natural and conventional categories is supported by a number of studies revealing that adults have domain-specific intuitions about category structure and the meaning of category boundaries (Gelman & Coley, 1991). For example, adults generally believe that animal,

but not artifact, categories have absolute boundaries. Thus, when judging whether an exemplar is a member of an animal category, adults respond that individuals either are or are not category members, and that partial category membership is not possible (e.g., an individual animal either is or is not a bird—it cannot be "sort of" a bird, even if it is highly atypical of the category); in contrast, partial category membership is considered possible for artifacts (e.g., a tray may be considered "sort of" a table; Diesendruck & Gelman, 1999; Estes, 2003; 2004; Kalish, 2002). Adults also appear to view animal categories as more tightly structured and homogeneous than artifact categories; thus, knowledge of category identity promotes a broader range of inferences about individual animals than artifacts (Gelman, 1988; Gelman & O'Reilly, 1988). Most relevant to the present work, adults view the boundaries that determine categories as an objective matter of reality for animals, but as subjective and dependent on human decision for artifacts. For example, Malt (1990) found that adults believed that questions about whether a particular animal is a bird should be referred to an expert (e.g., a biologist)—implying that there is a *correct* category identity for animals, even if it is unknown—but that questions about the category identity of artifacts (e.g., whether an object is a piece of furniture) are subjective and resolvable by appeal to human intuition.

That adults hold domain-specific beliefs about the meaning of category boundaries is consistent with the proposal that concepts are embedded in coherent intuitive theories that include information about ontology and causality (Murphy & Medin, 1985). For example, adults' understanding of artifacts as human inventions, created to fulfill human intentions (Bloom, 1996), may support the belief that artifact classification is similarly dependent on human action and thus, is subjective, culturally-

situated, and flexible. In contrast, adults' understanding of animals and plants as originating from natural processes, independent of human intent and action, may be incorporated into their understanding of animal and plant classification as similarly natural and objective. Although adults' domain-specific theories may include inter-connected beliefs about object origins and the ontological status of category boundaries, it is important to note that these are distinct components of conceptual structure (Atran, 1990). As noted by Kalish (1998), some naturally-occurring items may be classified in a manner that is clearly conventionalized. For example, both the categories GOLD and GEMSTONE consist of naturally occurring substances, but GOLD is a represented as a natural kind, whereas GEMSTONE is understood as conventionalized and flexible (Kalish, 1998).

There are a number of different possibilities regarding how representations of category meaning might develop. According to traditional cognitive development theories, young children's categories are atheoretical, based on superficial similarities or scripts, and are not stable across contexts (Piaget, 1929). A similar perspective has more recently been advanced by researchers documenting the important influence of perceptual features in a variety of cognitive processes in early childhood (e.g., encoding, Fisher & Sloutsky, 2005; induction, Sloutsky, Kloos, & Fisher, 2007). If indeed young children's categories are superficial and atheoretical, then children may be likely to treat all categorization as subjective and flexible (e.g., they may be more likely than adults to view unusual groupings of both animals and artifacts as acceptable means of categorization). Alternately, from this perspective, if young children were found to reliably judge some categories as natural and others as conventional, then this pattern

should be predicted by the degree of similarity across item sets, not domain-specific ontological commitments. In either case, both traditional and more contemporary similarity-based approaches to cognitive development predict that systematic domain-specific beliefs about category naturalness should be relatively late developments.

Alternately, theory-based approaches to early conceptual development suggest that domain-specific naïve theories drive early categorization (Gelman & Koenig, 2003; Wellman & Gelman, 1992). From this perspective, very young children's categories include principled beliefs about what it means to be an animal or artifact, and the distinct causal mechanisms that operate on each, thus leading to a variety of domain-dependent influences on early categorization and category-based reasoning. This perspective is supported by research showing that infants reliably distinguish between animals and artifacts in the first year of life (Mandler & McDonough, 1993), and that by preschool, children demonstrate a great deal of domain-specific theoretical knowledge about animal and artifact categories. For example, preschoolers acknowledge the role of human action in the creation of artifacts; they view animals as created by nature and artifacts as created by people (Gelman & Kremer, 1991), and they believe there is an important role for the intent of a human creator in assigning artifact identity (Bloom, 1996; Diesendruck, Markson, & Bloom, 2003; Gelman & Bloom, 2000) and function (Defeyter & German, 2003; German & Johnson, 2002).

Preschoolers also make domain-specific judgments about category immutability. They report that the category identity of an artifact can change over time, either because of external transformations (Gelman & Wellman, 1991; Keil, 1989), or when groups of people change the way they use an artifact (Siegel & Callanan, 2007), whereas they

expect animal identity to be consistent across external transformations (Gelman & Wellman, 1991; Keil, 1989) and environments (Gelman & Wellman, 1991; Waxman, Medin, & Ross, 2007). There is also some evidence that young children view category membership as more fundamental to animal than artifact identity. For example, when presented with novel animals and artifacts, preschoolers are more likely to ask questions about category identity for animals (e.g., "What is it?"), but to ask about function for artifacts (e.g., "What is it for?"; Greif, Kemler-Nelson, Keil, & Guitierrez, 2006). Finally, preschoolers incorporate causal knowledge about animals and artifacts into their predictions about a range of object properties. For example, preschoolers infer that animals and artifacts have different kinds of insides (Gelman & Wellman, 1991; Simons & Keil, 1995), and that animals, but not artifacts, are self-moving (R. Gelman, 1990; Massey & R. Gelman, 1988), grow larger over time (Rosengren, Gelman, Kalish, & McCormick, 1991), and are governed by a range of internal causal mechanisms (e.g., tiredness, hunger; Hatano & Inagaki, 1994).

Although the distinction between animals and artifacts appears well-entrenched in preschoolers' reasoning, the issue of whether young children also make distinctions between animal and artifact categories in terms of representations of category structure and the ontological status of categories boundaries (e.g., as naturally- vs. conventionally-defined, or as discovered vs. created by people) is much less clear. The little prior work that has directly examined children's beliefs about category structure has suggested that preschoolers' judgments do not follow adult-like, domain-specific patterns. Instead, young children appear to treat both animal and artifact categories as objectively-defined, discovered (as opposed to created) by humans, and homogeneous. For example, Kalish

(1998) found that kindergarteners reliably rejected categories that violated their own beliefs about how to categorize on questions about both animals and artifacts, suggesting that they viewed the criteria for classifying both animals and artifacts as objectively-determined. Similarly, Gelman (1988) and Gelman and O'Reilly (1988) found that children viewed basic-level animal and artifact categories as equivalently homogeneous and inductively rich, whereas adults viewed animal categories as more homogeneous and inductively rich than artifact categories.

Thus, although there is consistent evidence that young children distinguish between animal and artifact categories on a number of dimensions, it is not clear that they hold differentiated beliefs about which of these categories are objective reflections of natural reality and which are more flexible and determined by convention. Kalish (1998) suggested that despite young children's recognition of the distinction between animals and artifacts, they view artifact categories as reflecting real, objective distinctions in the world in part because basic-level artifact categories (e.g., chairs) are relatively homogeneous from a structural perspective. In other words, young children may believe that both animal and artifact categories reflect true structure because they correspond to groups that share richly correlated properties (e.g., that appear to form "natural bundles" in the environment; Mervis & Crisafi, 1982; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Though this is a plausible account, only a very small number of studies have addressed young children's beliefs about the meaning of category boundaries. Therefore, in the present work I aimed to examine more thoroughly children's beliefs about the naturalness vs. conventionality of categories in various domains.

Another primary aim of this research was to examine social categories, particularly gender and race, across development and cultural contexts. Previous work has proposed that individuals conceptualize categories based on gender and race much the same way as they represent categories of animals—as natural kinds (Allport, 1954; Atran, 1990; Hirschfeld, 1996; Rothbart & Taylor, 1992). This proposal has powerful theoretical appeal, in part because categories that are represented as natural kinds often take on special importance in a variety of cognitive processes (e.g., inductive reasoning, information processing; Gelman & Coley, 1991). Thus, the proposal that social categories are construed as natural kinds provides a cognitive framework for understanding the important role that social categories play in social cognition (e.g., for stereotyping and person encoding; Macrae & Bodenhausen, 2000). However, only a small amount of empirical work has directly examined the extent to which adults construe human social categories as natural or conventional (see Haslam, Rothschild, & Ernst, 2000), and there is no prior work directly on this topic with children.

Examining the development of social categories with respect to naturalness and conventionality also provides an important opportunity to evaluate the role of cultural context in conceptual development. For example, whereas the input that children receive about animals and artifacts may be relatively stable across contexts, the input they receive about social categories is likely to be more variable (Diesendruck, 2003; Diesendruck & Haber, 2009). Thus, social categories provide an important test case for examining whether beliefs about category naturalness in particular domains are variable across contexts, as well as how local cultural input influences concepts across age.

There have been several different theoretical proposals regarding the developmental origins of social categories, and the extent to which they vary across cultures. Hirschfeld (1996) suggested that there are specific evolved cognitive biases to view social categories as natural kinds, which emerge early in development and are relatively stable across diverse ages and cultural contexts (see also Atran, 1998; Gil-White, 2001). In contrast, others have proposed that children's initial social categories are superficial, and only become imbued with meaning through socialization (for a review, see Aboud, 1988), such that construing social categories as natural should be a later development and more variable across contexts. A third possibility suggests that the role of cultural context depends on the type of social category. For example, Cosmides, Tooby, and Kurzban (2003) suggest that concepts of gender may be constrained by intuitive biases, due to the particular evolutionary significance of gender, but that other categories, such as race, depend on social experience. From this perspective, younger children may be more likely to view gender than race as a natural kind. To test these competing hypotheses, I included children of various ages from two cultural contexts, and systematically compared concepts of gender and of race, as a preliminary examination of the interaction between development and cultural input in children's social categories (see Astuti, Solomon, & Carey, 2004).

To begin to distinguish these various proposals regarding the role of cultural context in conceptual development, I sampled children from two communities which I expected to vary in adult beliefs about social categories. In particular, I selected two communities within Michigan that differed in a number of ways, including urban vs. rural status, racial and ethnic composition, and in political conservatism (see Study 2).

Because these two communities differed in many important ways, this research was not designed to determine which specific factors contribute to group-level differences in children's concepts. Rather, this research was designed to be informative regarding *whether* and *when* cultural context influences conceptual development in the social domain.

I selected these particular communities, however, because I expected community-level differences in political conservatism to be a good indicator of variability in adult beliefs about social categories. In survey work with adult populations, political conservatism has been found to correlate with increased biological explanations for social categories (e.g., conservatism relates to more endorsement of genetic theories to explain differences between African Americans and European Americans; Jayaratne et al., 2006). Thus, I hypothesized that political conservatism would relate to the belief that social categories (e.g., based on race) represent natural, objectively-correct ways to classify people. Also, political conservatism has been proposed to involve viewing categories as having strict, absolute boundaries (Jost, Glaser, Kruglanski, & Sulloway, 2003). Thus, political conservatism may also relate more generally to the belief that everyday categories are objectively-defined kinds. For these reasons, I expected older participants from the more conservative community to view social categories as natural kinds to a greater extent than older participants from the less conservative community.

If so, then comparing the concepts of younger children from these communities will be informative regarding the role of cultural input in conceptual development. Particularly, if early concepts are constrained by intuitive biases to think of social categories as natural kinds, I should find that younger children in both communities

construe social categories as natural kinds, and that cross-cultural differences emerge only in older childhood (see discussion in Astuti et al., 2004, and Mead, 1930). Alternately, if concepts of social categories are the product of cultural learning, then young children's concepts should mirror those of adults in their own communities, such that cross-cultural differences should emerge early and be stable across childhood.

Overview

Study 1 is a preliminary study with 5-year-olds and adults from a single community, and was designed to develop a method for assessing animal, artifact, and social categories. Studies 2 and 3 provide the main data for evaluating the central research questions, and included children ages 5, 7, 10, and 17 from two cultural contexts. Study 4 presents a series of control conditions designed to rule out some alternate explanations for the domain effects found in Studies 1-3. Study 5 extends these findings to younger preschoolers (3- and 4-year-olds). Finally, Study 6 examines whether the effects of age and cultural context on representations of social categories found in Studies 2 and 3 extend to other dimensions of children's social concepts.

Chapter 2

Study 1

Study 1 is a preliminary study with 5-year-olds and adults from a single community, and was designed to develop a method for assessing animal, artifact, and social categories. In this study, I also examined how adults' self-ratings of political conservatism relate to their beliefs about category naturalness. My main research questions involving the role of cultural context in conceptual development are addressed in Studies 2 and 3. The method for Study 1 was based on Kalish's (1998) research on kindergarteners' beliefs about animal and artifact categories. In that work, five-year-olds responded to hypothetical vignettes involving visitors that "came from some place very far away where many things are done differently." Participants were then shown a series of categorization decisions that they were told reflected the categories held by members of the visitors' home culture and were asked to evaluate the acceptability of these decisions. Many of the visitors' categories violated expected criteria (e.g., the visitors grouped a deer with a horse, instead of with another deer). If young children believe that their own everyday categories represent objectively correct, natural ways of organizing the world, they should respond that the visitors' categories are "wrong". If, however, they believe that their own categories are subjectively determined and that there are multiple acceptable ways to categorize the given items, then they should be willing to

accept the visitors' categories. By varying whether questions referred to animals or artifacts, Kalish (1998) examined whether five-year-olds' judgments about category objectivity varied systematically by domain. Results indicated that kindergarteners tended to reject the visitors' categories for questions about both animals and artifacts at the basic level. For example, they were equally likely to reject categories that grouped a deer with a horse, separately from another deer, as they were to reject categories that grouped a hammer with a screwdriver, separately from another hammer.

Interestingly, Kalish (1998) did find that five-year-olds demonstrated a more flexible approach to artifact than animal categories at the superordinate level. For example, they were more likely to reject a category that grouped a lion and wolf, separately from a tiger, than they were to reject a category that grouped a table with a car, separately from a bookshelf. As discussed by Kalish (1998), however, because superordinate artifact categories may be less coherent than animal categories from a structural perspective, as well as less familiar to young children generally, it is unclear whether this pattern of responses is revealing of domain-specific theoretical beliefs.

For this reason, as well as because basic-level categories are the first to emerge ontogenetically (Rosch et al., 1976; but see Mandler, Bauer, & McDoonough, 1991), are universally perceived as the privileged level for induction for biological kinds (Medin & Atran, 2004), and are fundamental to determining object identity (see Rosch et al., 1976; Waxman, 1999), the present studies all examine children's basic-level categories for animals and artifacts. Particularly, I test whether children are willing to accept that items from different basic-level animal and artifact categories may be considered "the same kind." Similarly, I focused on social categories based on race and gender because race

and gender have been suggested to be the fundamental dimensions upon which adults categorize others (Hewstone, Hantzi, & Johnston, 1991; Taylor, Fiske, Etcoff, & Rudermann, 1978), and also appear to be salient to young children (Hirschfeld, 1996; Maccoby, 1988; Martin, 2000; Taylor, 1996). Study 1 is a preliminary study, which closely followed the method employed by Kalish (1998) in a sample of kindergarteners and adults from a midsize city in the midwestern United States. The main modification to his procedure in this initial study was to extend this method to examine concepts of gender and race.

Methods

Participants

Participants of two ages (N = 58; 27 kindergarteners, 9 male, 18 female; M age = 5,3; range = 4,0 – 6,4; 31 adults, 10 male, 21 female; M age = 18,7, range 18,1 – 19,7) were randomly assigned to either a baseline control condition or the focal experimental condition. Approximately 75% of participants were White, with the rest from diverse racial-ethnic backgrounds. Adult participants were recruited from introductory psychology subject pools at a large university in the Midwestern United States. Adults received partial course credit for participating. Children were recruited from a single public elementary school, located in the same city as the university from which adults were recruited, via letters sent home to parents in participating classrooms. Only children who returned signed consent forms were asked to participate. Trained experimenters interviewed children individually in a quiet area of the school. Sessions lasted approximately 15 minutes.

Procedures for Children

Experimental Condition

Introduction. Procedures closely followed Kalish (1998). Children were introduced to a colorful human-like puppet. They were told, "This is Feppy. Feppy is a visitor from a place far far away where they do lots of things differently than I do. Some of the things they do are wrong, but some of the things are just different. Your job is to help me figure out when Feppy is saying something *wrong* and when Feppy is saying something that is *maybe right*, but just different. Let's practice. I'll ask Feppy a question, and you help me decide whether his answer is *wrong* or whether it is *maybe right*."

Warm-up. Children then completed two warm-up questions. First, the experimenter asked Feppy what the participating child's name was, and Feppy responded that the child's name was "Sammy." The experimenter then asked the child, "Is it maybe right for Feppy to say your name is Sammy, or is it wrong?" Children should respond that this is "wrong", and were corrected if necessary. Second, the experimenter asked Feppy whether his favorite snack was carrots or cookies, and Feppy responded that his favorite snack is carrots. The experimenter then asked the child, "Is it maybe right for Feppy to say he likes carrots, or is it wrong?" Children should respond that this is "maybe right," and were corrected if necessary. The experimenter then reminded the child, "So, let's remember… First, Feppy said your name was Sammy, was that maybe right, or was that wrong? Then, Feppy said he liked carrots, was that maybe right, or was that wrong?" To begin the experiment, the experimenter said, "Great! Now, we are going to do some more questions, and for each one, you help me decide whether the things that Feppy says are maybe right or wrong."

Experimental Task. Children completed questions about four domains of items, including animals, artifacts, gender, and race. Questions were blocked by domain, the order of presentation of the blocks was counterbalanced across participants following a Latin square design, and items within each block were presented in a separate random order for each participant. Each block consisted of five questions in which the puppet was asked to determine whether a target item was the same kind of thing as one of two items (see Table 1). For example, in an animal set, participants saw two animals from different basic-level categories (e.g., a collie dog and a gray-and-white cat); in an artifact set, participants saw two artifacts from different basic-level categories (e.g., a table and a bookshelf); in a gender set, participants saw two exemplars of the same race but different gender; for race items, participants saw two exemplars of the same gender but different race (see Table 1). All pictures were colorful 3 X 5 photographs (with blank backgrounds) collected from online sources. Each example was placed in a separate tray, and the experimenter said, "Here are two [animals, things, people]". Placement of each exemplar on the participant's right or left was counter-balanced for each question across participants.

Then, participants were shown a target item that matched one of the examples at the basic-level for animals and artifacts (e.g., a black Labrador dog), and for gender and race for social categories, and the experimenter said, "Here is another [animal, thing, person]". The experimenter then said, directing the question to the puppet, "'Feppy, which one of these [animals, things, people] (*pointing to the exemplars in each tray*) is the same kind as this one (*pointing to the target*)? Which one does it go with?' Let's see what Feppy says… Feppy says, 'It is the same kind as this one!'" The target was then

placed in the tray that contained the exemplar that the puppet had indicated. Children were asked, "Is it wrong for Feppy to say that, or is he maybe right?" Order of presentation of the answer choices was counterbalanced across participants (i.e., half of participants were asked "is it maybe right for Feppy to say that, or is it wrong?").

On four out of the five questions in each domain, the puppet selected the exemplar that did not match the target on the expected criteria (basic-level status for animals and artifacts, gender or race for people; these choices are referred to throughout as *unexpected categories*), whereas on one question in each set the puppet chose the exemplar that matched the target based on these criteria (referred to throughout as *expected categories*). These expected categories are control questions, designed to assess whether children developed response biases to reject all of the presented categories.

Following Kalish (1998), after children responded to the initial question, they were asked a follow-up question based on their initial response. If the child indicated the category was "wrong", they were asked, "Everyone where Feppy lives would also say that it [*the target*] is the same kind as this one [*the unexpected exemplar that had been selected by the puppet*]. So, is it wrong for Feppy to say that, or is he maybe right?" If the child indicated that the presented category was "maybe right" on the initial question, then the target was then taken out of the tray that Feppy had selected first and placed in the other tray. Children were then asked, "Would it also be maybe right for Feppy to say that it [*the target*] is the same kind as this one [*the exemplar that had not been chosen for the first question*]?"

After all of the experimental questions, children were asked two additional control questions. First, children were shown a picture of a circle and a square and were told that

Feppy identified the circle as a square. They were asked whether this choice was "maybe right" or "wrong". Because children were expected to believe that what makes something a square is objective, they should respond that the answer is "wrong". Secondly, they were shown pictures of two ice cream cones and told that Feppy identified chocolate as his favorite ice-cream. Because children of this age understand the subjective nature of preferences (Wellman & Liu, 2004), they should respond that the answer given by the puppet is acceptable in this case. These questions test whether children responded reasonably and thoughtfully to the study questions.

Scoring. For both the initial questions and the follow-up questions, responses in which children indicated that the categories were "wrong" were scored as a '1,' responses in which children responded that the categories were "maybe right" were scored as a '0.' The same scores were given to children's responses to the follow-up question regardless of which follow-up question they received, following the procedures used by Kalish (1998). This scoring procedure was used because an answer of "wrong" to either of the follow-up questions indicates an inflexible approach to categorization, whereas an answer of "maybe right" to both follow-up questions indicates that the participant accepts multiple correct ways to categorize the items. Therefore, regardless of which follow-up question participants receive, lower scores indicate more flexibility towards how items may be categorized.

Baseline Control Condition

The purpose of the baseline control condition was to assess whether participants would independently choose to categorize the items at the basic level for artifacts and animals and based on gender and race for people, as well as whether they were equally

likely to use these criteria across domains (e.g., for gender and race items). This control was critical because if participants' categories were not based on these dimensions, then the visitor's responses would not seem unexpected to them.

Children were presented with all of the same materials as in the experimental conditions. Children were also introduced to the puppet: "This is Feppy. Feppy is a visitor from a place far far away where they do lots of things differently than we do. So, Feppy doesn't know a lot of the things that you know. Your job is to help Feppy learn about things that he doesn't know about." As in the experimental condition, participants completed two warm-up questions. They were told, "First, Feppy asks, 'what is your name?' Can you tell Feppy what your name is? Now, Feppy asks 'which snack do you like better? Do you like cookies or carrots?'"

Children then completed four sets of items, including all of the items from the experimental condition grouped in the same manner. For each question, participants were presented with the examples and told, "Here are two (animals, things, or people)." Each exemplar was set in a separate tray and the target was placed equidistant to both trays. Children were then asked, "Which one of these (animals, things, or children, *point to exemplars*) is the same kind as this one (*point to target).* Which one does it go with?" No follow-up questions were asked in this condition.

Scoring. Responses on which participants selected the exemplar that matched at the basic level for animal or artifacts, or on gender or race for people, were scored as a '1,' the alternate response was scored as '0.'

Procedure for Adults

Adults were also randomly assigned to the baseline control condition or the focal experimental condition. They completed the study independently via a computer program. The tasks were very similar, with small modifications to make procedures more suitable for older participants. For example, no puppet was used; the character was described simply as a "visitor from some place far away" and was referred to as "the visitor" throughout. Also, the control question at the end of the task involved a simple math question that the visitor responded to incorrectly, instead of the mislabeling of shapes. At the end of the task, participants were asked to self-report on their gender, race, birth-date, and political views (on a scale: 4 = *very liberal*; 3 = *liberal*; 2 = *conservative*; 1 = *very conservative*).

Results

In all studies, because participants' responses consisted of a series of dichotomous choices, data were analyzed through a combination of techniques, including analyses based on the binomial distribution and analyses involving non-parametric methods, except where otherwise noted. Specifically, I used binomial regression models to compare proportions of category rejections to the proportions expected by chance (.50; see Tables 2, 4, and 5), Mann-Whitney tests to compare means from independent samples, and the Generalized Estimating Equations (GEE) procedure in SPSS 16 for assessing main effects and interactions. The GEE procedure is appropriate because this procedure can account for the underlying binary structure of the data (the dependant variable is analyzed as the number of category rejections out of the total possible —4— in each domain), and can assess both within- and between- subjects effects (Harden &

Hilbe, 2003). These analyses yield Wald χ^2 values as indicators of main effects and interactions. All post-hoc contrasts that are discussed throughout the paper were significant at $p < .05$, following sequential Bonferroni corrections.

These methods are appropriate given the structure of the data; however, all analyses were also conducted using standard parametric statistics (e.g., one-sample t-tests, independent samples t-tests, and repeated measures analysis of variance), and in all cases, the same patterns of significant results were obtained. Also, for all studies, preliminary analyses revealed that there were no main or interactive effects of participant gender, so this variable was excluded. Effects of participant gender are examined in more detail in the supplementary analyses following Study 3.

Baseline Control Condition

The purpose of the baseline control condition was to assess whether children and adults would choose to categorize the item sets based on the expected criteria in each domain. Table 2 presents these data and comparisons to chance. Both children and adults reliably chose to categorize based on the expected criteria, and did so significantly more often than expected by chance in each domain (ps < .001). A series of Mann-Whitney tests revealed that children's and adults' responses differed only for animals, $Z = 2.45$, $p < .05$, such that adults used the expected criteria more often than children; however, both children and adults matched at the basic level for animals more often than expected by chance (see Table 2).

Experimental Condition

Control questions

There were two types of control questions built into the experimental condition. First, on one item in each domain, the visitor identified an expected category (referred to as embedded control questions). As expected, children rarely rejected these categories, doing so on only 17% of questions (0 for animals, 14% for artifacts, 36% for gender, 21% for race), and adults rejected these categories on only 5% of questions (0 for animals, 0 for artifacts, 7% for gender, 13% for race). These results confirm that neither children nor adults had strong biases to respond that the presented categories were always wrong. These control questions were not analyzed further; the primary analyses below include only the four items from each domain that presented unexpected categories.

Secondly, at the end of the experimental task, participants were asked two control questions, one in which the visitor gave an unexpected answer to a question about shapes (for children) or math (for adults) and one in which the visitor had a preference for a particular ice-cream flavor. As expected, all of the children and adults responded that the visitor's answer was wrong on the control question about shapes/math, whereas only 8% of children and no adult responded that the visitor's answer was "wrong" on the question about ice-cream. Thus, analyses of all of the control questions suggest that children and adults were able to follow the task, responded thoughtfully, and did not have strong response biases that persisted across questions.

Main Analyses

Table 2 presents the mean proportion of questions on which participants rejected the unexpected categories for both the initial and follow-up questions, as well as comparisons to chance. These data were examined through a 4 (domain: animal, artifact,

gender, race) X 2 (age: five-year-olds, adults) GEE, with domain as a within-subjects variable. These analyses were conducted for responses to the initial question, and then repeated for responses to the follow-up question.

For responses to the initial question, 5-year-olds reliably rejected the unexpected categories on questions about animals, artifacts, and gender, but not race, whereas adults reliably rejected these categories on questions about animals and artifacts only (see Table 2). The GEE analysis yielded significant effects of Domain, $\chi^2(3) = 52.83$, $p < .001$, and an Age X Domain interaction, $\chi^2(3) = 19.18$, $p < .001$. Children rejected categories for race significantly less often than for gender or artifacts, and marginally less often than for animals ($p = .07$). Children's judgments for animals, artifacts, and gender were equivalent. Adults rejected categories for both gender and race significantly less often than for animals or artifacts. Adults' judgments about animals and artifacts were equivalent, as were their judgments about race and gender. For age effects, children rejected categories more often than adults only on items about gender ($p < .05$).

Inspection of Table 2 reveals that on the follow-up question, children again reliably rejected animal, artifact, and gender categories, but not race categories. In contrast, adults reliably rejected only the visitors' animal categories. Analyses of the follow-up question revealed significant effects of Age, $\chi^2(1) = 7.40$, $p < .01$, Domain, $\chi^2(3) = 16.63$, $p = .001$, and an Age X Domain interaction, $\chi^2(3) = 10.99$, $p < .05$. Children rejected categories less often for race than for artifacts or gender, and their responses for animals, artifacts, and gender were again equivalent. In contrast, adults rejected categories for artifacts, race, and gender less often than for animals. Adults'

responses to artifacts, gender, and race were equivalent. For artifacts and gender, children rejected categories more often than adults.

Analyses of the follow-up question showed somewhat different patterns than the initial question, in that adults appeared more likely to accept the artifact categories on the follow-up question than on the initial question. One possibility was that adults were more likely than 5-year-olds to change their answers across the two questions asked for each item. Indeed, follow-up analyses on the proportion of questions on which participants changed their response from a rejection of the unexpected category on the initial question to acceptance on the follow-up question revealed that adults were more likely to do so than children (*M* children = .14, *SE* = .03; *M* adults = .30, *SE* = .05), Mann-Whitney, $Z = 2.35$, $p < .05$. Examining responses separately by domain revealed that children were significantly less likely than adults to change their responses to questions about animals, $Z = -2.32$, $p < .05$, and artifacts, $Z = -2.85$, $p < .01$. In fact, children never changed their initial responses for animals or artifacts, whereas adults did so 17% of the time for animals and 27% of the time for artifacts. This suggests that the two-question format used in Study 1 may have been too demanding for young children; accordingly, this procedure was modified in Study 2.

Variability in Adults' Responses

For exploratory purposes, some possible sources of variability in adults' responses were examined, including participant gender, race, and political attitudes. The only significant findings involved political attitudes, which were significantly associated with participants' responses for race. Specifically, individuals who rated themselves as either "very conservative" or "conservative" rejected categories that conflicted with

expected racial categories more often than individuals who rated themselves as "liberal" or "very liberal" (Liberal: M = .29, SE = .11, Conservative: M = 1, SE = 0; Z = 2.49, p < .05). Thus, whether adults viewed race as a natural kind appeared to be associated with broad political beliefs. The number of individuals rating themselves as conservative in the present sample was relatively small (n = 5; out of 14 in the analysis). Nevertheless, these results are suggestive and will be explored more thoroughly in Study 2.

Discussion

In Study 1, adults viewed the criteria for forming animal categories as objective; they were unwilling to accept unexpected animal categories on either initial or follow-up questions. This finding is consistent with prior work suggesting that adults view animal categories as reflecting real distinctions found in nature, and consequently, expect animal categorization to be universal (Malt, 1990). They were considerably more flexible about whether unexpected criteria could be used to form categories of artifacts, or people (for both gender and race). For artifacts, however, they were only willing to accept these unexpected categories on the follow-up questions, after they were reminded that the presented categories were shared by all members of the visitor's home culture. Overall, these findings suggest that adults view animal categories as natural, but artifact and social categories as more conventional.

Consistent with Kalish (1998), 5-year-olds viewed both animal and artifact categories as objectively determined. Also, 5-year-olds, unlike adults, viewed gender as a natural kind, consistent with a number of studies documenting relatively strict gender concepts among young children (Berndt & Heller, 1986; Biernat, 1991; Rhodes & Gelman, 2008; Taylor, 1996). Children were, however, willing to accept that the criteria

for some categories could be flexible. Particularly, they accepted more of the visitors' categories for race than for any other domain. Thus, Study 1 suggests that 5-year-olds view animal, artifact, and gender categories as natural, but race categories as more conventional. Importantly, 5-year-olds reliably formed categories based on animal species, artifact kind, gender, and race in the baseline control condition. Therefore, their flexibility regarding racial categories cannot be accounted for by a lack of awareness of or attention to physical cues to race. Taken together, data from the control and experimental conditions suggest that although 5-year-olds are aware of categories based on the expected criteria in each domain, they view categories based on race as more flexible than those based on gender, as well as more flexible than animal or artifact categories. Overall, data from Study 1 suggest that there may be developmental changes in gender and artifact categories across childhood development.

Chapter 3

Study 2

The goals of Study 2 were to provide a more sensitive measure of children's concepts and to examine the influence of cultural context on conceptual development.

Methodological Changes

Based on the results of Study 1, I considered that two features of the design, and the design used by Kalish (1998), might have discouraged young children from a flexible approach toward categorization. First, in order to accept the presented categories, children had to accept two unusual conditions: 1) that it was acceptable to consider two items from different common categories to be the same kind (e.g., a dog and a cat) and 2) that it was acceptable to *not* view two items of the *same* expected category as the same kind (e.g., two dogs). Thus, the presence of a "more correct" item among the stimulus set could have discouraged children from viewing the visitors' categories as plausible (e.g., Nguyen & Murphy, 2003).

In Study 2, questions were designed to tap only the first of these features, whether children will accept the possibility that two items from different commonly-used categories are the same kind. The "more correct" match (based on expected criteria) was removed from consideration, such that questions asked only whether it was "maybe right" to consider two animals or artifacts from different basic-level categories, or two

children who differed in terms of either gender or race, as the same kind. Participants saw only two items at a time (e.g., a dog and a cat) and were told that they reflected a category formed by the visitors (e.g., the puppet and his friends). In the absence of an exact expected match, one possibility was that children and adults would accept all of these categories. For example, they could accept the categories because the two exemplars share membership in a superordinate category (e.g., both animals are mammals, both artifacts are tools; similarly, the girl and boy are both White, the African American and White child are both girls). Therefore, with this procedure, I should find that participants reliably reject the categories only when they believe that their expected categories are objectively the only, or best, groupings. In domains in which participants believe that there are multiple ways to organize kinds, they should accept the unexpected categories.

Secondly, the two-question format for each item used in Study 1 might have limited young children's ability to understand fully that the categories presented by the visitor reflected categories held by all members of his home culture (not simply his own idiosyncratic beliefs; Kalish, 1998). In Study 1, this information was presented only in the follow-up question; participants were first told that the visitor made a decision, and then they were told that members of his home culture also shared the visitor's belief. Children may have based their initial decision on the evidence that only the *visitor* had formed an unusual category, without recognition that this category represented shared beliefs, and then simply not adjusted their responses on the follow-up question because this procedure was too demanding for them.

To address this concern, in Study 2, the information that the formed category (e.g., of a dog and a cat) was held by members of the visitor's community was presented through a single question, as follows: "Feppy and his friends ALL say that these are the same kind of animal. Are they maybe right?" Participants also were shown a picture depicting the visitor's family and friends, to emphasize that a community shared belief in the formed categories. In this way, participants were asked only a single question for each item. By simplifying the presentation of the questions, these items allowed for more sensitive measurement of young children's concepts.

Cultural Context

In Study 2, I also began to examine the influence of cultural context on conceptual development. The findings from Study 1, which indicated that self-rated political conservatism correlated with adults' concepts of race, indicated that there is variability in adult concepts of social categories, as well as that political conservatism provides a good marker of this variability. Thus, to examine the developmental trajectory of cross-cultural differences in concepts of social categories, I compared the development of children growing up in two communities that differed in overall political conservatism. This will allow us to test when cross-cultural differences emerge across age, and will be informative regarding the relative role of intuitive biases and cultural input in the development of social categories.

Children from 5-18 years were recruited from two communities in Michigan. One community (which also participated in Study 1) was located in a midsized city (approximate population = 115,000; all demographic information was obtained from the United States Census Bureau website for the 2000 census), with an above-average

median household income ($46,299) and level of adult education (48% of adults over age 25 have a bachelor's degree or higher). This city is the location of a large public research university, and is located about 35 miles from a major urban center. The population of this city is approximately 75% White, 12% Asian, 9% African American, and 3% Hispanic. The population of the participating school was approximately 68% White, 17% Asian, 4% African American, and 4% Hispanic, and the participating samples approximately reflected this distribution. As a social and political context, this city is relatively liberal (see below).

I compared the development of children growing up in this community (referred to throughout as the midsize city) with children growing up in a community located about 75 miles to the west, in a more rural, and socially and politically conservative area of the state (referred to throughout as the rural community). For the second sample, children were recruited from two elementary schools in a rural community, both located in small villages within the same county. One of these villages had a population of 2,337, of which 96% were White, 2% were African American, and 1.5% were Hispanic. This village has a median household income of $35,223. The other village had a population of 514, of which 99% were White, and a median household income of $38,523. In the county where the two villages were located, 12% of adults over age 25 had obtained a bachelor's degree or higher. Although the midsize city and rural community differed in many ways, the percentages of elementary school children from each district meeting state averages on standardized test scores was approximately equivalent for both reading and math.

To establish that the midsize city and the rural community differed in terms of overall political conservatism, I examined their political behavior by analyzing publicly available voting records for the precincts where the schools that I sampled from were located (voting records, by precinct, were obtained from the Secretary of State's website). In the precinct surrounding the school in the midsize city, in 2004, 71% of voters voted for the liberal candidate for president, and in 2006, 77% voted for the liberal candidate for Michigan governor. In contrast, the majority of voters in the two precincts surrounding the schools in the rural community voted for the conservative candidates in these two elections (2004 presidential race: 73% in one precinct, 65% in the other; 2006 governor race: 62% in one precinct, 57% in the other). The same patterns were seen for recent conservative ballot initiatives to amend the state constitution (in favor of a conservative definition of marriage: 30% in the precinct in the midsize city, in the rural community, 73% in one precinct, 69% in the other; in favor of banning affirmative action: 35% in the precinct in the midsize city, in the rural community, 76% in one precinct, 71% in the other).

Because both communities were in Michigan, and were in fact only 75 miles apart, they shared many more features than the types of diverse contexts that are often included in cross-cultural comparisons (e.g., Astuti et al., 2004). For example, they shared many state institutions and public policies, and had access to similar media. Despite these important similarities, however, I expected that the two communities would provide sufficiently distinct input about the meaning of social categories. For example, ethnographic studies have revealed that distinct communities within the United States provide unique developmental contexts, which result in important local variation in

children's language and concept use (e.g., Heath, 1983). Similarly, prior developmental research has documented interactions between intuitive biases and cultural input, as marked by community-level variables, in conceptual development relating to biological phenomena (e.g., Evans, 2001; Diesendruck & Haber, 2009). Thus, although the two communities were relatively proximal, I expected that they provided unique developmental contexts that could lead to differences in concepts of social categories.

As the two communities appeared to present children with different social and political contexts, I examined whether, and at what age, children in these samples differ in their social categories. If social categories are driven by cultural input, then I would expect to find cultural variability from an early age, with children in the rural area more likely than children in the midsize city to construe social categories as natural. Alternately, if social categories are initially guided by intuitive biases, then there should be more similarities among young children's concepts across the two communities, and more cultural differences with age (see Astuti et al., 2004; Mead, 1930).

Method

Participants

Recruitment

Children in kindergarten, 2nd grade, and 5th grades from public elementary schools in the midsize city and rural community were recruited via letters sent home to the parents of students in participating classrooms. Only children who returned signed consent forms were included. During this consent process, parents in the rural community were also asked to provide contact information if they were willing to complete a brief demographic form. Parents who provided this contact information were contacted via

mail or email and asked to provide demographic information, including political attitudes, via return mail or through a secure online survey. Children completed individual sessions with trained experimenters in a quiet area of their elementary school.

Participants were also recruited from 11th and 12th grade classrooms in both communities. For these older students, letters were sent home describing the study; parents were asked to indicate if they did not want their child to participate. All students whose parents did not refuse consent were asked to participate. These older students completed a paper-and-pencil version of the study during their regular class periods, using procedures similar to those used for adults in Study 1.

Ages

From the midsize city there were 16 kindergarteners (9 male, 7 female; *M* age = 5,0; range = 4,7– 5,6), 15 2nd graders (8 male, 7 female; *M* age = 7,9; range 7,3– 8,7), 13 5th graders (5 male, 8 female; *M* age = 10,11; range = 9,11– 11,8), and 16 12th graders (9 male, 7 female; *M* age = 17,11; range = 16,0– 18,8). From the rural community there were 19 kindergarteners (8 male, 11 female, *M* age = 6,3; range = 5,7– 6,11), 18 2nd graders (8 male, 10 female; *M* age = 8,3; range = 7,10– 9,2), 19 5th graders (9 male, 10 female, *M* age =11,1; range = 10,8– 12,2), and 16 12th graders (7 male, 9 female; *M* age = 18,2; range = 17,7– 19,2).

Procedures

Younger Children

Children (5-, 7-, and 10-year-olds) were introduced to the puppet using the same introduction and warm-up questions as in Study 1, with the exception that they were also shown a picture of "Feppy's friends and family," which remained on the table throughout

testing. As in Study 1, all children were asked five questions in each of four domains (animals, artifacts, gender, and race, see Table 3). Before each domain children were told, "I'll show you the [animals; things; people] that Feppy and all his friends say are the same kind of [animal; thing; person] and you tell me whether they are maybe right." For each question, the experimenter said, "Feppy and his friends all say that these are the same kind of [animal, thing, person]. Are they maybe right?" Children simply responded with "yes" or "no."

Responses of "no" indicate an inflexible approach to categorization and were scored with a '1', whereas responses of "yes" were scored with a '0.' As in Study 1, within each block, children were shown four items in which unexpected categories were presented, and one expected category (see Table 3). All items used the same pictures as were used in Study 1. The blocks were given in one of four orders following a Latin square design. The items within each block were given in set random orders for each form.

Older Children

Experimental Procedure. 17-year-olds completed a paper and pencil version of this task. Following the same introduction used with adults in Study 1, they were shown pictures of the categories, with the phrase, "The visitor says he and everyone from his home community say these are the same kind of [animal, thing, person]. Are they maybe right?" Participants responded by circling either "yes" or "no."

Measures of Essentialist Beliefs. 17-year-olds also completed questionnaire measures of essentialist beliefs about gender and ethnicity groups, in which they were asked how much they agreed with a series of statements describing beliefs about the

nature of these human categories. These measures were adapted from Haslam et al. (2000), and are presented in Appendix A. To increase sample size, data from these measures were collected from 17-year-olds in both Studies 2 and 3, and are analyzed in Study 3.

Parental Demographic Questionnaire. Parents from the rural area who agreed to complete a brief demographic form via return mail or a secure online web-site were asked the following questions: 1) What is your relationship to the child who participated in our study? 2) What is your occupation? 3) If there are other adults living in your home, please list their relation to the child who participated in this project, and their occupation. 4) Please list the gender and age of all other children living in your home. 5) What is your religious background? How regularly do you attend religious services? (1 = *less than once per year*, 2 = *a few times per year*, 3 = *once a month*, 4 = *every week*). 6) How would you describe your political attitudes? (1 = *very liberal*, 2 = *liberal*, 3 = *moderate*, 4 = *conservative*, 5 = *very conservative*).

Results

Control questions

On the embedded control items in each domain, in which expected categories were presented, children rarely rejected these categories. They did so on only 14% of animal questions, 10% of artifact questions, 27% of gender questions, and 15% of race questions, all less than expected by chance, $ps < .001$. A series of logistic regression analyses were conducted to determine whether responses to these control questions were influenced by age, cultural context, or an interaction between age X cultural context. These analyses revealed significant effects only for items about gender. For gender, there

was a significant effect of cultural context, Wald $\chi^2(1) = 6.43$, $p < .05$. Fewer children in the midsize city (14%) than in the rural community (36%) responded that the visitors were "wrong" on this control item; though in both communities children rejected these categories less often than expected by chance (*p*s < .05).

On the control items at the end of the procedure, 98% of children responded that the visitors were "wrong" on the shape/math question, and only 10% did so for the question about ice cream preference, *p*s < .001. Thus, analyses of both the embedded control items and the control questions at the end of the task indicated that children did not develop pervasive response biases on these questions, and appeared to respond thoughtfully. Responses to the control items were not analyzed further; analyses below include the four items in each domain in which the visitors presented unexpected categories.

Experimental Questions

The mean proportions of questions on which children rejected the unexpected categories, and comparisons to chance, are presented in Table 4. These data were examined through a 4 (domain: animal, artifact, gender, race) X 2 (cultural context: midsize city, rural community) X 4 (age: 5-, 7-, 10-, 17-year-olds) GEE, with domain as a within-subjects factor. This analysis revealed significant main effects of Age, $\chi^2(3) = 11.72$, $p < .01$, Cultural Context, $\chi^2(1) = 6.38$, $p < .05$, and Domain, $\chi^2(3) = 168.29$, $p < .001$, and interactions between Age and Cultural Context, $\chi^2(3) = 19.09$, $p < .001$, and Cultural Context and Domain, $\chi^2(3) = 13.66$, $p < .01$. To examine the nature of these effects, a series of follow-up analyses were conducted separately for each domain, with

age and cultural context as the independent variables, and the proportion of rejected categories in each domain as the dependent variable.

The Effects of Age and Cultural Context in Each Domain

Animals. As shown in Table 4, children of each age rejected the visitors' animal categories more often than expected by chance. For animals, there were no effects of age or cultural context on participants' responses.

Artifacts. As shown in Table 4, no age-group rejected the visitors' artifact categories more often than expected by chance, and 10- and 17-year-olds in the midsize city reliably accepted these categories. For artifacts, there was a significant effect of age, $\chi^2(3) = 14.08$, $p < .01$, such that 5- and 7-year-olds rejected these categories more often than 10- or 17-year-olds. There were no significant effects involving cultural context.

Gender. As shown in Table 4, younger children in both communities, as well as older children in the rural community, rejected the visitors' categories for gender more than half of the time. In contrast, older children in the midsize city did not. For gender, there were significant effects of Age, $\chi^2(3) = 24.38$, $p < .001$, and Cultural Context, $\chi^2(1) = 7.12$, $p < .01$, and an interaction between Age and Cultural Context, $\chi^2(3) = 27.55$, $p < .001$. Examining this interaction revealed that in the midsize city, concepts of gender became more flexible with age: 5- and 7-year-olds rejected the visitors' categories more often than did 10- or 17-year-olds. In contrast, there was no effect of age on the concepts of children in the rural community. Also, the effect of cultural context was specific to the older children: 10- and 17-year-olds in the rural community rejected the visitors' categories more often than children of these ages in the midsize city, whereas the concepts of younger children did not differ by cultural context.

Race. As shown in Table 4, 5- and 7-year-olds' responses to questions about race did not differ from the proportion expected by chance. In contrast, 10- and 17-year-olds in the rural community rejected these categories more often then expected by chance, and children of these ages in the midsize community reliably accepted these categories. For race, there was a significant effect of Cultural Context, $\chi^2(1) = 39.03$, and an interaction between Age and Cultural Context, $\chi^2(3) = 33.90$, $p < .001$. Examining this interaction revealed that again, the effect of cultural context was specific to the older children; 10-year-olds and 17-year-olds in the rural community rejected the visitors' categories more often than did children of these ages in the midsize city, whereas there was no effect of cultural context on the concepts of younger children.

The Effects of Cultural Context and Domain for each Age

In order to compare children's responses directly across domains, I conducted another series of follow-up analyses, testing for effects of domain and cultural context, separately for each age-group.

5- and 7-year-olds. Among both 5- and 7-year-olds, these analyses revealed only an effect of Domain (5-year-olds: $\chi^2(3) = 46.65$, $p < .001$; 7-year-olds, $\chi^2(3) = 41.37$, $p < .001$). 5- and 7-year-olds rejected animal and gender categories more often than race or artifact categories. Their responses to animal and gender were equivalent, as were their responses to race and artifacts. There were no effects of cultural context.

10-year-olds. For 10-year-olds, there were effects of cultural context, $\chi^2(1) = 10.84$, $p < .01$, Domain, $\chi^2(3) = 75.84$, $p < .001$, and a marginal interaction between Cultural Context and Domain, $\chi^2(3) = 7.35$ $p = .06$. In the midsize city, 10-year-olds rejected the visitors' categories more often for animals than for artifacts or race, whereas

in the rural community, 10-year-olds rejected the visitors' categories for animals more often than artifacts only. In the rural community, 10-year-olds' responses for animals, gender, and race were equivalent. Also, the effects of cultural context were specific to the social categories; correcting for post-hoc comparisons, 10-year-olds in the rural community rejected the gender categories marginally more often than did children in the midsize city, $p = .08$, and they rejected the race categories significantly more often, $p < .01$.

17-year-olds. For 17-year-olds, there were significant effects of Domain, $\chi^2(3) = 32.92$, $p < .001$, Cultural Context, $\chi^2(1) = 11.55$, $p < .001$, and Domain X Cultural Context, $\chi^2(3) = 8.20$, $p < .05$. In the midsize city, 17-year-olds rejected animal categories more often than they rejected artifact, gender, or race categories, and their responses to artifacts, gender, and race were equivalent. In the rural community, 17-year-olds' responses to animals, gender, and race were equivalent, and they rejected each more often than they rejected artifacts. The effects of cultural context were specific to the social categories; 17-year-olds in both communities responded similarly to animals and artifacts, but 17-year-olds in the midsize city were less likely to reject the visitors' categories for gender or race than were 17-year-olds in the rural community.

Discussion

In Study 2, individuals of all ages from both communities rejected animal categories that did not follow expected criteria at the basic level. This finding supports prior work indicating that individuals represent animal species as natural kinds, believing that there are objectively correct ways to categorize animals (e.g., Atran, 1990). In contrast, participants of every age group demonstrated a more flexible approach to

artifact categories, suggesting that even young children understand something of the conventional nature of artifact groupings (Siegel & Callanan, 2007), and are willing to accept that their own expected categories represent only one of multiple acceptable ways to organize artifacts.

This study therefore revealed a level of differentiated reasoning about animal and artifact categories that was not present in Study 1 or in Kalish (1998). This greater flexibility may have reflected procedural modifications that provided a more sensitive measure of children's concepts. It is important to note that children did not reliably *accept* the unexpected artifact categories; rather they accepted them about half of the time (and more often than they accepted animals). This is likely because there were multiple factors that could have led children to reject these categories. For example, the artifact categories contained items with different labels (see Waxman, 1999), and were relatively incoherent from a structural perspective (Kalish, 1998). Thus, beliefs about the conventionality of artifact categories, which should lead children to accept the unexpected categories, conflicted with several other factors that could have led them to reject these categories. I interpret these data as indicating that children construe artifact categories as *relatively* more conventionalized than animal or gender categories, but not that they view them as completely subjective and flexible (see discussions of children's understanding of conventionality in Kalish & Sabbagh, 2007). Notably, children were also increasingly more likely to accept the artifacts with age, suggesting increased commitment to the conventional nature of artifact groupings across childhood.

Cultural context and age interacted to influence representations of human categories. For social categories, younger children's concepts were similar across

communities, but differed by type of social category. In both communities, 5- and 7-year-olds reliably rejected unexpected gender categories, but did so for race only about half of the time. Thus, younger children appeared to construe gender, but not race, as a natural kind. Among 10- and 17-year-olds, there was an important influence of cultural context. Older children in the rural community treated both gender and race as natural kinds, whereas older children in the midsize city treated both social categories as more conventionalized. For gender, the pattern across ages and communities suggests that natural kind concepts emerge early, and become more flexible with age only in particular cultural contexts. For race, the pattern across ages and communities is inconsistent with the proposal that early representations are guided by predispositions to view racial categories as natural kinds (Hirschfeld, 1996); rather, this pattern suggests that natural kind concepts of race emerges only later in childhood and are dependent on cultural input. Thus, these findings revealed different developmental trajectories for race and gender, consistent with the possibility that concepts of gender are constrained by intuitive biases, whereas concepts of race are more dependent on cultural input (e.g., Cosmides et al., 2003).

It is important to note that the questions for social categories differed somewhat from those involving animals and artifacts, in that all pictures necessarily included information about multiple category memberships. For example, for gender questions, items included a boy and a girl who were both White (and were also both young children). For race questions, items presented children of different physical appearances indicative of race, but who were both either male or female. Thus, for gender items, children could have accepted the categories because of shared race or age, and should

have rejected the categories only if they view gender as fundamental. Similarly, for race, they could have accepted the categories because of shared gender or age, and should have rejected the categories only if they view race as fundamental. Also, for social categories (perhaps more so than for animal or artifact categories), children could have considered a range of other features to determine whether two people should be considered the same kind (e.g., personality characteristics, interests, and so on). Thus, when children reliably rejected gender (as did younger children in both communities, and older children in the rural area) or race (as did only older children in the rural area), I interpret these findings as indicating that they view these categories as particularly fundamental to identity, and as necessary criteria to follow when forming social kinds. In Studies 4a and 4b, I address alternate explanations for these findings, related to the baseline salience of gender and racial categories to young children, as well as to children's abilities to recognize alternate means of classification (e.g., their ability to recognize shared gender on the race items and shared physical appearance on the gender items) across domains.

Chapter 4

Study 3

The results from Study 2 are the first that I am aware of to demonstrate effects of local cultural contexts for social categories with respect to naturalness and conventionality. Therefore, it seemed prudent to replicate these findings—particularly the developmental patterns obtained in the rural community. Also, because children in Study 2 were asked questions about multiple domains, it was desirable to obtain a measure of participants' beliefs about human categories that could not be influenced by their responses in other domains. For example, it seemed possible that responding to a set of questions regarding gender could bias children's responses to a set of questions regarding race (or vice versa), and so it was desirable to present questions about gender and race to different groups of participants. In Study 3, a new sample of children ages 5-18 from the same rural community was recruited, using the same recruitment procedures as in Study 2, and randomly assigned to either a Gender or a Race condition. Participants were always asked questions about the human category first (either gender or race depending on condition), followed by questions about animals and artifacts, in counterbalanced order. In order to replicate the interaction between cultural context and age, additional groups of 7-year-olds and 17-year-olds were also recruited from the midsize city. As in Study 2, demographic information was collected from parents and

high school students, and high school students completed additional measures of essentialist beliefs (see Appendix A).

Method

Participants

In the midsize city, participants included 22 2nd graders (6 male, 16 female; *M* age = 8,11; range = 8,5–9,5) and 30 12th graders (8 male, 22 female, *M* age = 18,0, range = 17,3–18,10). In the rural community, participants included 21 kindergarteners (14 male, 7 female, *M* age = 5,10; range = 4,8–7,8), 33 2nd graders (12 male, 21 female, *M* age = 8,0; range = 7,8–9,1), 17 5th graders (12 male, 5 female, *M* age = 11,1; range = 10,3–12,2) and 24 12th graders (13 male, 11 female, *M* age = 17,6; range = 16,6–18,8).

Procedures

All procedures followed the method used in Study 2, with the exception that participants were randomly assigned to receive questions about only one kind of social category (either gender or race), and questions about human categories were always presented first. All other instructions, procedures, and scoring were identical to those in Study 2.

Results

Control Questions

On the embedded control items in each domain, in which the visitors presented expected categories, participants rarely rejected these categories. They did so on only 10% of animal questions, 7% of artifact questions, and 18% of people questions (23% gender, 14% race), all less often than expected by chance, *p*s < .001. Logistic regression

analyses found no effects of age, cultural context, or condition on responses to these questions.

On the control questions following the experimental task, participants responded that the visitors were "wrong" on 97% of shape/math questions, but only 10% of ice cream questions. Therefore, as in previous studies, children appeared to follow the task and respond thoughtfully to study questions.

Experimental Questions

The mean proportion of questions on which participants rejected the visitors' categories in each domain, and comparisons to chance, are presented in Table 5. Overall, the pattern was very similar to that obtained in Study 2. 5-year-olds (in the rural community) and 7-year-olds (from both communities) reliably rejected the visitors' categories for animals and gender, but not for artifacts or race. Among older children, 10- and 17-year-olds in the rural community reliably rejected the categories for animals, gender, and race. In contrast, 17-year-old in the midsize city reliably rejected categories only for animals, and they reliably accepted them for artifacts, gender, and race.

Because I only had data from 7- and 17-year-olds from the midsize city in Study 3, it was desirable to confirm that these data were similar to those obtained from younger and older children, respectively, from the midsize city in Study 2. Indeed, planned comparisons, conducted separately for each domain, confirmed that 7-year-olds from the midsize city in Study 3 performed similarly to both 5- and 7-year-olds from the midsize city in Study 2, $ps > .30$, and that 17-year-olds from the midsize city in Study 3 performed similarly to both 10- and 17-year-olds from this community in Study 2, $ps > .50$. These analyses confirm that the newly collected data from 7- and 17-year-olds in the

midsize city provide good estimates of the concepts of younger and older children, respectively.

To examine the effects of cultural context, data from 7- and 17-year-olds (as data from these ages were available from both communities) were analyzed through a 2 (age: 7-years, 17-years) X 2 (cultural context: midsize city; rural community) X 2 (type of social category: gender, race) X 3 (domain: social, animal, artifact) GEE, with domain as a within-subjects variable. This analysis revealed significant effects of Age, $\chi^2(1) = 9.44$, $p < .01$, Domain, $\chi^2(2) = 111.77$, $p < .001$, Cultural Context X Domain, $\chi^2(2) = 6.90$, $p < .05$, and Age X Cultural Context X Type of Social Category X Domain, $\chi^2(2) = 8.60$, $p < .05$. To consider this interaction, I ran follow-up analyses separately for each domain.

Animals

For animals, I ran a 2 (age) X 2 (cultural context) GEE. This analysis revealed a significant effect of Age, $\chi^2(1) = 6.18$, $p < .05$. Overall, 7-year-olds rejected the visitors' animal categories more often than 17-year-olds did. However, as shown in Table 5, both age-groups rejected these categories more often than expected by chance. There were no effects of cultural context, ps > .30.

Artifacts

For artifacts, I ran a 2 (age) X 2 (cultural context) GEE. This analysis revealed a significant effect of Age, $\chi^2(1) = 9.00$, $p < .01$. Overall, 7-year-olds rejected the visitors' artifact categories more often than did 17-year-olds. However, as shown in Table 5, neither group rejected these categories more often than expected by chance. There were no effects of cultural context, ps > .40.

Social Categories

For social categories, I ran a 2 (age) X 2 (cultural context) X 2 (type of social category: gender, race) GEE, with all three variables as between-subjects factors. This analysis revealed significant effects of Cultural Context, $\chi^2(1) = 9.43$, $p < .01$, and a Cultural Context X Age interaction, $\chi^2(1) = 4.22$, $p < .05$. The effect of age was specific to the midsize city; in this community, older children accepted more of the visitors' social categories than did younger children. In contrast, in the rural community, there was no evidence of developmental change. Also, the effect of cultural context was specific to the 17-year-olds; at this age, children in the midsize city accepted more of the visitors' social categories than did children in the rural community, whereas there were no effects of cultural context on the responses of younger children.

There were no effects involving the type of social category, ps > .15, suggesting similar patterns for gender and for race. However, although the effects of age and of cultural context were similar for these two social categories, inspection of the means (see Table 5) reveals that the two categories differed in the extent to which young children construed them as natural. Particularly, 7-year-olds in both communities (as well as 5-year-olds in the rural community) rejected the visitors' categories for gender significantly more than half the time, but they did not do so for race. Indeed, a planned comparison confirmed that 7-year-olds rejected the categories for gender significantly more often than for race, $Z = 2.07$, $p < .05$. Thus, young children appear to construe gender, but not race, as natural. In contrast, older children in the rural community reliably rejected the visitors' categories for both types of social categories, indicating that they construed both race and gender as natural, whereas older children in the midsize city appeared to view both categories as more conventional.

Supplementary Analyses

For our main experimental tasks, identical patterns were found across Studies 2 and 3. Therefore, data were combined from these two studies in order to increase the overall sample size to allow for supplementary analyses examining individual response patterns, item level analyses, participant gender and ethnicity, how parental factors relate to children's representations, and how individual differences in 17-year-olds' responses relate to other aspects of essentialist beliefs.

Item Level Analyses

I examined whether particular item sets appeared to elicit patterns of reasoning that were not captured by the mean of each domain, by examining the mean level of "no" responses for each item set by age and cultural context. This examination did not reveal any consistent item-level differences. The mean proportions of times that children rejected the visitors' categories for each item set are presented in Table 6.

Individual Response Patterns

In the combined data set, children were classified according to their overall response patterns in each domain. Specifically, children were coded as 'rejecters' if they rejected the unexpected categories on at least 3 out of 4 items in a domain, 'accepters' if they accepted the unexpected categories on at least 3 out of 4 items, and 'inconsistent' if they accepted the unexpected categories on 2 questions and rejected them on 2 questions. Overall, examining the data in this manner revealed the same patterns as were evident in our analyses of group means. These classifications are available in Appendix B.

Participant Gender and Ethnicity

Gender. In the combined data set, there were 279 participants, 128 male and 151 female, with enough participants of each age to test for effects of participant gender (5-year-olds: 31 male, 25 female; 7-year-olds: 34 male, 54 female; 10-year-olds: 26 male, 23 female; 17-year-olds: 37 male, 49 female). Separately for each age-group, I conducted a series of Mann-Whitney tests comparing responses of male and female participants for each domain. These analyses revealed no significant effects of participant gender (*p*s > .1).

Ethnicity. There was not sufficient diversity in either community to test for effects of participant ethnicity. Also, reliable data on individual participants' ethnic backgrounds were available only for 17-year-olds, who self-reported on this variable. One concern, however, was that because the samples from the midsize city were more ethnically diverse than the samples obtained from the rural community, group differences among older children by community could relate only to the differences in levels of diversity within the samples, as opposed to more general cultural effects on concepts. For example, children from racial-ethnic minority backgrounds could have more flexible concepts of social categories, and if so, then the greater proportion of children from diverse backgrounds in the midsize city could account for the community-level differences. To address this possibility, I conducted a series of Mann-Whitney tests comparing the concepts of 17-year-olds from the two communities, but limiting the sample only to participants who self-identified their racial-ethnic background as White, Caucasian, or European American (in the midsize city, this included 76% of the sample, $n = 35$, whereas in the rural community, this included 93% of the sample, $n = 37$). If community-level differences are driven by differences in the levels of racial-ethnic

diversity across the two samples, then with these more homogeneous samples, I should no longer find effects of cultural context for social categories. However, these analyses replicated those from Studies 2 and 3. Among these samples of 17-year-olds, there were no effects of cultural context for animals or artifacts (*p*s > .1), but there were significant effects of cultural context for gender, $Z = 3.04$, $p < .01$, and race, $Z = 3.84$, $p < .001$, such that children in the midsize city were more likely to accept these categories than were children in the rural community (midsize city: gender, $M = .39$, $SE = .07$, race, $M = .23$, $SE = .08$; rural community: gender, $M = .72$, $SE = .08$, race, $M = .71$, $SE = .07$).

Parental Political Attitudes

Parents of the children in the rural community self-reported on their political beliefs (How would you describe your political attitudes? *1= very liberal, 2 = liberal, 3 = moderate, 4 = conservative, 5 = very conservative*; $M = 3.0$, $SD = 1.23$; parent data were not available from the midsize city). Within the rural sample, level of parental conservatism significantly correlated with their children's score on race questions in the main experimental task ($n = 23$; $r = .42$, $p < .05$), such that parents who rated themselves as more conservative had children who were more likely to reject the visitors' categories that violated expected racial categories. The number of families included in this analysis was relatively small for several reasons: many parents chose not to respond to requests for this information, I did not have data on race concepts for some children in Study 3 (i.e., those who were assigned to the Gender condition), and some parents failed to provide the information necessary to link their responses to their child's data. Due to the small sample size, this correlation should be interpreted cautiously. However, the relationship between parental political conservatism and children's representations

appeared to be somewhat robust. I further examined this relationship by dichotomizing levels of parental conservatism—comparing the children of parents who described themselves as "very liberal" or "liberal" (n = 8) to those of parents who described themselves as "moderate", "conservative", or "very conservative" (n = 14). There was a significant effect of this grouping variable on children's responses to questions about race, Mann-Whitney test, $Z = 2.03$, $p < .05$. Children of parents in the conservative group reliably rejected the visitors' categories on questions about race ($M = .86$, $SE = .05$), whereas children of parents in the liberal group did not ($M = .47$, $SE = .16$). Parental political conservatism did not relate to children's responses for any other domain (ps > .30), also, no other parental factor (e.g., religiosity, family composition, employment status) related to children's concepts in any domain, ps > .30).

17-year-olds' Essentialist Beliefs

Combining the data from 17-year-olds across Studies 2 and 3 (n = 86), youth in the midsize city endorsed marginally lower levels of essentialist beliefs on the essentialism questionnaires (see Appendix A) than those in the rural community, on both the gender essentialism scale (α = .71, midsize city, $M = 5.55$, $SD = 1.25$, rural community, $M = 6.07$, $SD = 1.22$, $t(83) = -1.91$, $p < .06$) and the ethnic essentialism scale (α = .73, midsize city, $M = 4.70$, $SD = 1.32$, rural community, $M = 5.24$, $SD = 1.51$, $t(82) = -1.76$, $p < .09$). Scores on measures of gender and ethnic essentialism were significantly correlated, $r = .47$, $p < .001$, as were the scores of proportion of 'no' responses on the main experimental task on questions about gender and race ($r = .44$, $p < .05$). Endorsement of an essentialist perspective towards ethnicity on the essentialism scale correlated with adolescents' performance on the race questions from the main

experimental task ($r = .38$, $p < .01$), such that participants who endorsed more essentialist beliefs on the ethnicity essentialism scale were more likely to reject unexpected racial categories. I did not find a significant correlation between the gender essentialism scale and the gender score on the main experimental task.

Discussion

The findings from Study 3 replicated several key findings from Study 2. First, young children approached the question of whether categories are natural kinds or conventionalized groups in a domain-specific manner; they were more likely to reject unexpected categories for animals and gender than for artifacts or race. In both Studies 2 and 3, concepts of social categories changed with age, depending on cultural context. Gender and racial representations had different developmental trajectories: five-year-olds treated gender, but not race, as a natural kind. Over time, gender representations became more flexible only in the midsize city, and race representations became less flexible only in the rural community. In the supplementary analyses, I found a relationship between parental political conservatism and children's concepts of race, as well as between 17-year-olds' concepts of race and other measures of their essentialist beliefs. These studies suggest that conceptual distinctions between natural and conventional categories emerge early in development, and that concepts of gender and of race follow distinct developmental trajectories.

Chapter 5

Studies 4a – 4d

Across studies, I have found that children as young as five years of age have domain-specific intuitions about the ontological status of categories. Although this domain-specificity in children's concepts is predicted by some models of early conceptual development (Gelman, 2003), these findings are counter to predictions made by models that suggest that early categorization is driven by perceptual, rather than theoretical, features (Sloutsky & Fisher, 2004). In light of these considerations, before concluding that the patterns that I have reported reflect true conceptual distinctions on the part of children, I conducted a series of control studies designed to address some alternate hypotheses that may account for the pattern of findings in Studies 1-3. Alternatives related to four factors were considered: 1) baseline differences in category salience, 2) accessibility of category knowledge and alternate means of classification, 3) perceptual similarity, and 4) linguistic cues. Each of these alternate hypotheses and relevant data will be discussed in turn.

Study 4a: Baseline Category Salience

Study 1 included a baseline categorization condition in which five-year-olds were asked to form their own categories using the study stimuli. This condition demonstrated that even the youngest participants readily used the same expected criteria as adults for grouping items based on animal species, artifact category, gender, and race. Because

Study 1 involved participants only from the midsize city, however, it was important to demonstrate that these factors were also salient to children in the rural community. This was particularly important for items about race, because these children attended very racially homogeneous schools. Thus, their relative flexibility for race could simply be related to the fact that they have had little contact with racial diversity, and consequently, did not attend to race during the experimental task.

Method

To examine this possibility, a baseline control condition was conducted with a new group of kindergarteners from both communities (N = 33, 16 from the midsize city, 17 from the rural area; 16 female, 17 male, M age = 6,2; range = 5,8– 7,0). In this condition, children were introduced to the puppet using the same introduction as in previous studies. In this context, however, children were told that their job was to "help Feppy learn about things he doesn't know about". Children were told that "Feppy went around and looked at all sorts of [animals, things, people] that we have here", and that "Feppy wanted to know if he found [animals, things, people] that are the same kind". Children were asked to "help Feppy by telling him whether the [people, animals, things] he found are the same kind of [person, animal, thing]."

By random assignment, half of the children were shown pictures of two items from the same expected categories (i.e., the expected matches used as distracter items in Study 1), such that viewing the expected criteria as good criteria for classification should lead children to say "yes," whereas the other half of children were shown the unexpected categories used in Studies 2-3, such that if they recognize and use the expected criteria, they should say "no." In other words, whereas on the experimental task, children were

asked to evaluate whether the visitors' categories could be considered acceptable, on this task they were asked to use their own criteria. Within each condition, all children were shown all item pairs, with items blocked by domain and presented in one of four orders following a Latin square design.

Results and Discussion

When children were shown expected categories, and asked if they were the same kind, they responded "yes" for 85% of animal questions, 90% for artifacts, 66% for gender, and 94% for race—all more often than expected by chance (*p*s < 05). When children were shown unexpected categories (i.e., the same items used in the experimental tasks in Studies 2-3), they responded "no", meaning that they were not members of the same kind, 91% for animal questions, 77% for artifacts, 89% for gender, and 72% for race, which again were all more often than expected by chance (*p*s < .01). There were no effects of cultural context on children's responses in these control conditions, *p*s > .15.

Thus, when asked to apply their own criteria, children reliably used expected criteria to evaluate categories of animals, artifacts, and people (both race and gender). It is important to note that on questions about race, kindergarteners had to bring their own ideas about race to this task. Because children were shown only two items at a time, and asked whether they were the same, they could not simply match items together based on perceptual similarity (as they may have done in Study 1). This control study demonstrates that even the youngest participants were aware of the expected categorization criteria in each of the studied domains (e.g., they know that people can be classified on the basis of physical cues to race, and that artifacts are often grouped at the basic level). Taking these results together with those from Studies 2 and 3, however,

suggests that although 5-year-olds are aware of these expected criteria in each domain, they have domain-specific intuitions about which of these criteria are *necessary* to follow, and which are more subjective and flexible.

Study 4b: Category Knowledge

I next considered the possibility that perhaps children were more willing to accept unexpected categories when they could easily generate property similarities between the two paired items, or could easily generate an alternate means of categorization by which the two items could be considered the same (e.g., a superordinate or thematic category). For example, dogs and cats are from different basic-level categories, but are both mammals. They also share a thematic category membership, as both are pets. For artifacts, for example, both a table and a bookshelf are from the superordinate category *furniture* and share a thematic category membership of belonging in a living room. Similarly, for gender items, the presented categories included children that differed in gender, but were of the same race and age. For race items, children differed in terms of physical features indicative of race, but were of the same gender and age. Young children may have considered these other thematic and superordinate categories when determining whether two items could be considered the *same kind*. Thus, one possibility was that children could recognize these alternate means of categorization for some domains (e.g., artifacts and race) more than others (e.g., animals and gender). Similarly, children could have considered the number of shared properties (e.g., both have fur, both go to school) when making these judgments. If children could more easily bring to mind shared category memberships or shared properties for artifacts or race than for animals or gender, then this could account for the experimental findings.

Method

To test this possibility, new groups of 2nd graders from both communities were recruited (*N* = 23; 12 from the rural community, 11 from the midsize city; 7 male, 16 female *M* age = 8,1, range = 7,6– 8,11). Second graders were selected because their pattern of responses did not differ from kindergarteners on the main experimental task, but they were expected to be able to cope more successfully with the verbal demands of a feature-listing task. As in the main task, children were introduced to a puppet. In this task, however, children were told that, "Feppy and his friends don't know about a lot of the things that we know about. So, Feppy is here to learn from you about a lot of different things. Your job is to teach Feppy about things he doesn't know about." In turn, children were told that Feppy went around and looked at the animals, things, and people that we have here and now wants to learn. Children were shown the same unexpected pairs of items used in the main task, and for each pair, the experimenter had the puppet ask the child, "How are these two the same?" If necessary, the experimenter then repeated, "Tell Feppy some ways that these are the same." Children were shown all of the experimental items from Studies 2 and 3. Items were blocked by domain, and blocks were presented in one four random orders following a Latin square design. Children's responses were audio-taped, transcribed, and coded by two independent raters (inter-rater agreement = .97, differences resolved by discussion). Each response was coded as a shared property (e.g., a physical feature: "they both have fur", shared location "they are both at school") or as naming a superordinate or thematic category (e.g., "they are both reptiles", "they are both kids").

Results and Discussion

Children readily generated shared properties on this task (animals, $M = 8.57$, $SD = 3.49$; artifacts, $M = 6.39$, $SD = 2.27$; gender, $M = 6.83$, $SD = 3.04$; race, $M = 8.0$, $SD = 4.72$). A series of t-tests revealed that the only domain effects were that children generated more properties for animals than for artifacts, $t(22) = 3.30$, $p < .01$, or gender, $t(22) = 2.60$, $p < .05$, and more properties for race than for artifacts, $t(22) = 2.58$, $p < .05$. These findings suggest that the ease with which children could generate shared properties does not predict the findings from Studies 2 and 3, as children most readily identified how animals of different basic-level kinds were similar to each other, yet in experimental studies, they were consistently unwilling to accept that these animals could be considered the *same kind*. Also, there were no domain differences in how often children generated superordinate or thematic categories when describing these items (Animals, $M = 1.65$, $SD = 1.67$; Artifacts, $M = 1.39$, $SD = 1.15$, Gender, $M = 1.22$, $SD = 1.31$, Race, $M = 1.74$, $SD = 1.63$). There were also no effects of cultural context on the number of properties or categories generated in any domain. Thus, the accessibility of shared properties or shared membership in superordinate or thematic categories does not predict the pattern of results obtained on the experimental tasks.

Study 4c: Perceptual Similarity

Similarity accounts of categorization suggest that young children group together items that are perceptually similar to each other; constructing theory-laden categories is considered to be a later developmental achievement. From this perspective, any obtained patterns of domain-specific responding should reflect perceptual differences in the experimental stimuli. For this to be a possible alternative, artifact and race items should

be more perceptually similar than animal or gender items, as young children consistently accepted artifact and race categories more often than animal or gender categories.

Method

To examine the degree of perceptual similarity across item sets, 13 college students were presented with the stimuli used in Studies 2 and 3 and asked to "look carefully at each pair of pictures and rate how *visually similar* you find each pair." Participants rated every item on a seven-point scale ranging from 1 = *'not visually similar at all'* to 7 = *'very visually similar'*. Items were blocked by domain, and the blocks, and items within the blocks, were presented in a separate random order for each participant.

Results and Discussion

Items had moderate to low perceptual similarity (animals, $M = 3.38$, $SD = 1.1$; artifacts, $M = 2.52$, $SD = 1.01$; gender, $M = 3.33$, $SD = 1.25$; race, $M = 3.29$, $SD = 1.14$). Repeated measures analysis of variance, with domain as a within-subjects factor, revealed a significant effect of domain, $F(3, 36) = 4.58$, $p < .01$. Follow-up analyses revealed that artifacts were perceived as *less* similar to each other than items from all other domains. This finding is exactly opposite to what would be expected if perceptual features could explain the pattern of findings obtained in Studies 2 and 3, as the domain containing items that were the *least* perceptually similar were *most* likely to be accepted as possible categories by children of each age. Because these data were obtained from adults, it was also desirable to examine whether young children were likely to perceive similarity in a comparable manner. To do so, I re-examined the data collected from the feature listing task described in Study 4b, to determine whether the number of shared perceptual properties generated by children (e.g., they both have fur, they both

have handles, they both have brown eyes) varied by domain. Consistent with the adult ratings, perceptual features were generated *less* often for artifacts (M = .62 perceptual features per item) than for any other domain (Animals, M = 1.61; Gender, M = 1.68; Race, M = 1.65), ps < .001. This provides strong evidence that children did not view the artifacts as more perceptually similar than the items from other domains. Note that, also, in both the adult and child data, ratings for gender and race were equivalent. I next examined the correlations between the number of perceptual features listed by children and the visual similarity ratings provided by adults, across items, and found a strong positive relationship, r = .51, p < .05. This suggests that, in this context, young children and adults attended to similar dimensions when judging perceptual similarity, and that the degree of perceptual similarity within item sets does not predict the experimental findings.

Study 4d: Linguistic Cues

Finally, I considered that subtle linguistic cues within the experimental questions could have contributed to domain effects. On each item in the experimental task, children received some verbal input about the domain of focus. Specifically, children were told that the visitors considered two individuals to be the "same kind of animal," "same kind of thing", or "same kind of person." One possibility is that children were more flexible in their categorization of artifacts, as compared to animals, because the word "thing" is less informative than the word "animal." Thus, children may have understood the animal questions as more constrained with respect to appropriate categorization criteria than the artifact questions. For example, directing children to consider whether the items could be considered the "same kind of animal" could have led

them to focus on the basic level for animals (e.g., dog and cat), whereas the phrase "same kind of thing" may have led them to focus on the superordinate level for artifacts (e.g., both clothing).

Method

To address this possibility, I conducted two control conditions with new groups of kindergarteners (N = 29, 16 male, 13 female, M age = 5,5, range = 4,11– 6,2; all from the midsize city). In the "Same Kind" condition, children completed identical questions and procedures as were used in Study 2, with the exception that when each item was shown to children, regardless of domain, children were told, "Feppy and his friends all say these are the same kind." That is, the last part of the sentence (e.g., "of animal", "of thing", or "of person") was dropped. If the linguistic cues in the previous experiments contributed to our findings that children rejected categories of animals more than categories of artifacts, then providing them with identical linguistic input for all types of items should alter the pattern of findings.

In the "Same Kind of Thing" condition, children received only animal and artifact items, and heard the same wording in both domains: "Feppy and his friends all say these are the same kind of thing." If children respond flexibly to the phrase "same kind of thing", then when this phrase is used for both animals and artifacts, they should accept more unexpected animal categories than they did in the previous studies. Children received items blocked by domain, with order of presentation of domains counter-balanced across participants.

Results and Discussion

In the "same kind" condition, the results were identical to those from kindergarteners in Studies 2 and 3. They rejected the visitors' categories more often on questions about animals ($M = .81$, $SE = .04$) than on questions about artifacts ($M = .65$, $SE = .05$), Mann-Whitney, $Z = 2.13$, $p < .05$. They also rejected the visitors' categories for gender ($M = .81$, $SE = .11$) more often than for race ($M = .21$, $SE = .09$), Mann-Whitney, $Z = 2.72$, $p < .01$. In the "same kind of thing" condition, children again reliably rejected the visitors' categories for animals ($M = .79$, $SE = .08$, $p < .01$), and they did so more often than for artifacts ($M = .56$, $SE = .08$), Mann-Whitney, $Z = 2.28$, $p < .05$. Thus, children's performance on the main experimental tasks appears not to have been influenced by unintended linguistic cues.

Discussion of Studies 4a-4d

These four control studies documented that the domain effects found in Studies 2 and 3 do not relate to children's baseline categories, the accessibility of superordinate or thematic categories, perceptual features of the stimuli, or linguistic characteristics of the task. Thus, these findings support the interpretation that the differences in children's responses by domain reflect true conceptual distinctions on the part of young children. In particular, by age 5, children appear to view animal and gender categorization as objective processes, but artifact and race categorization as more subjective and flexible. As these control studies have ruled out a number of alternate explanations for these findings, the results from Studies 2 and 3 may be interpreted as indicating that children

incorporate abstract ontological commitments into their categories, from at least the age of five.

Chapter 6

Study 5

The aim of Study 5 was to extend the investigation to younger preschoolers. Pilot testing revealed that the procedures used in Studies 2 and 3 were too difficult for younger preschool children. Therefore, a new method for assessing beliefs about category meaning was developed, based on a method used by Gelman, Croft, Fu, Clausner, and Gottfried (1998).

In Study 5, preschoolers were shown a target item (e.g., a dog), a picture of an item from a different basic-level category (e.g., a cat), and a picture that was covered by construction paper and so hidden from view. After training activities (described below), preschoolers were asked to point to the picture that showed the "same kind" of [animal, thing, person] as the target item (e.g., the dog). Thus, if preschoolers think that the visible picture (e.g., the cat) could plausibly be considered the same kind of animal as the target (e.g., the dog), they should select the visible picture. Alternately, if they do not think it is possible for the visible picture to be considered the same kind as the target, they should point to the hidden picture. The rationale of this task, as in the previous studies, is to assess whether children believe it might be acceptable to identify items from different expected categories as being "the same kind", in order to assess whether expected categories are objectively or subjectively based. Based on the prior work, I

hypothesized that children would be more likely to select the hidden picture on questions about animals and gender than on questions about artifacts or race.

Method

Participants

Participants included 29 preschoolers (*M* age = 4,0, range = 3,0– 4,11; 13 three-year-olds, 16 four-year-olds; 15 male, 11 female), recruited from preschools in a small town neighboring the midsize city sampled from in Studies 1-4. Preschoolers were randomly assigned to one of two conditions. In both conditions they received questions about animals and artifacts, but in one condition they received questions about gender and in the other they received questions about race.

Procedures

Participants first completed a warm-up activity (modeled after Gelman et al., 1998) in which they were trained to select the hidden pictures when the visible picture did not match an experimenter's request. For example, the experimenter presented the child with a card that showed one visible picture (e.g., a television) and one hidden picture (e.g., a picture of flowers, covered by a paper screen). First, the experimenter said "point to the television", and children pointed to the visible picture. Second, the experimenter said "point to the flowers". Children were encouraged to look under the screen to find the flowers, and then to point to the hidden picture. Children completed a series of four warm-up questions, in which they were allowed to look under the screen, in order to communicate to them that when the visible picture did not match the experimenter's request, they should then select the hidden picture.

For the experimental questions, the experimenter presented a series of cards that were laminated, such that children could not look under the screen to see the hidden pictures. Each card contained one visible picture, and one screen covering a hidden picture. The experimenter held a target item equidistant from the visible and hidden pictures and said, "I'm looking for a(n) [animal, thing, person] that is the same kind as this one. Where is the [animal, thing, person] that is the same kind?" If necessary, the experimenter then said, "Is it here? [*pointing to the visible picture*] Or is it under here? [*pointing to the hidden picture*]" The order in which the experimenter offered the two answer choices was counterbalanced across participants. Children responded by pointing to either the visible or the hidden picture.

Each child was presented with three blocks of items, including animals, artifacts, and one of the human categories. Blocked domains were presented in one of three orders corresponding to a Latin square design, and the items within each block were presented in a separate random order for each participant. Each block included four items in which the target and the visible picture were the unexpected category pairs presented in previous studies (see Table 3), and two control questions. The control questions included one question in which the target matched the visible pictures on the expected criteria (similar to the control items used in Studies 2 and 3), such that children should select the visible picture. For the other control question, the visible picture was more distantly related to the target than in the experimental pairs (animals, target = fish, visible picture = bird; artifacts, target = clock, visible picture = doghouse; gender: target = young girl, visible picture = elderly man; race, target = young African American boy, visible picture = elderly White man), such that children should select the hidden picture. Trials in which

the children selected the hidden pictures were scored as '1', and selections of the visible picture were scored as '0'.

Results

Control Questions

Analysis of the control items indicated that children tended to select the visible pictures when they matched the target on expected criteria, but tended to select the hidden pictures when the visible item was drawn from a more distant class. When expected matches were visible, they were selected the vast majority of the time (96% for animals, 96% for artifacts, 85% for gender, and 100% for race, all greater than chance, *p*s < .05). When items from more distant classes were visible, children typically selected the *hidden* pictures (96% for animals, 89% for artifacts, 85% for gender, and 77% for race, all greater than chance, *p*s < .05). Responses to these control items demonstrate that children understood the experimental format and did not have strong response biases that persisted across questions. Responses to the control items are not analyzed further. Analyses below include only those questions in which children were shown the items from previous experiments, in which the target and visible pictures were drawn from different basic-level categories for animals and artifacts, and different gender or racial categories for human social categories.

Experimental Questions

For the experimental items, descriptive statistics are presented as the proportions of times that children selected the hidden pictures. Three-year-olds reliably selected the hidden pictures for animals ($M = .85$, $SE = .07$) and for gender ($M = .81$, $SE = .12$), *p*s < .001. In contrast, they reliably did not select the hidden picture for artifacts ($M = .25$, *SE*

= .11) and for race (M = 0, SE = 0), ps < .01. Four-year-olds also reliably selected the hidden pictures for animals (M = .84, SE = .04) and gender (M = .83, SE = .12), ps < .001. They reliably did not select the hidden picture for race (M = .14, SE = .09), p < .001, and they selected the hidden and visible pictures equally often for artifacts (M = .55, SE = .10). Responses did not differ by age for animals, gender, or race, but three-year-olds were more likely than four-year-olds to select the visible picture for artifacts, Mann-Whitney, Z = 1.96, p = .05. At each age, children selected the hidden pictures more often for animals than for artifacts (Wilcoxon Signed Rank Tests for related samples: 3-year-olds, Z = 2.72, p < .01; 4-year-olds, Z = 2.72, p < .01), and for social categories more often in the gender condition than the race condition (Mann-Whitney tests for independent samples: 3-year-olds, Z = 2.89, p < .01; 4-year-olds, Z = 2.97, p < .01).

Discussion

These findings suggest that 3- and 4-year-olds, like 5-year-olds in the previous studies, view the criteria for animal and gender categorization as objectively determined and inflexible, but view the criteria for artifact and racial categorization more flexibly. In particular, when preschoolers were told that that an experimenter was looking for two animals of the same kind, children reliably assumed that the experimenter was not looking for an animal that matched at the superordinate level, but was from a different basic-level category (e.g., when the target was a dog, they assumed the experimenter was not looking for a cat). Similarly, for gender, when the target was a boy, for example, and the experimenter was looking for two people of the same kind, children assumed that the experimenter was not looking for a girl. These findings indicate that preschoolers view

gender categories and animal species categories as indicating membership in fundamentally distinct kinds.

As in the previous experiments, it is important to note that there are a number of reasons why children could have selected the visible pictures for animals and gender. For the animal items, for example, the dog and cat are both mammals and pets. Similarly for the gender items, the boy and girl were both children of the same age, and also appeared to be of the same race. Thus, the consistent finding that preschoolers assumed that these were not acceptable categories (in the context of identifying two items that are the same kind) provides strong evidence that preschoolers view species categories as fundamental to animal identity, and gender as fundamental to human identity.

These findings are the first that I am aware of to document similarities in 3-year-olds' reasoning about animal species and human gender. These findings are consistent with other work from older preschoolers and kindergarteners. For example, Rhodes and Gelman (2008) found that 4- and 5-year-olds treated animal and gender categories as similarly predictive of novel individual properties. In particular, they found that these older preschoolers assumed that individual properties that varied by categories (e.g., a dog had one property and a cat had another, or a boy had one property and a girl had another) would be stable over time, whereas they expected properties that varied within categories (e.g., two dogs had different properties, or two boys had different properties) to be more variable. Thus, 4- and 5-year-olds treated human gender and animal species as equivalently informative for making predictions about individual properties.

Similarly, Taylor, Rhodes, & Gelman (2009) directly compared young children's reasoning about how being born a member of species category for animals, or a member

of a gender category for people, influenced later development. The youngest children in this work were 5-year-olds, and results indicated that children of this age judged the influence of species categories and gender categories to be very similar. For example, they expected that an animal who was born a cow but raised by pigs would still engage in "cow-like" behaviors (e.g., saying moo), much like they expected that a child who was born a girl but raised by boys would still engage in "girl-like" behaviors (e.g., liking ballet). On these items, children gave similar proportions of birth category-predictions for questions about animal species and human gender, and they also gave similar kinds of explanations for category-based behaviors across domains.

Thus, the present findings are consistent with this previous work documenting similarities in reasoning about human gender and animal species in early childhood, and extend this work in two key ways. First, I documented similar reasoning across these two domains in a younger sample of children (3-year-olds). Secondly, this study examined children's beliefs about categorization directly (e.g., whether categorization based on gender and species is necessary to consider two individuals the same kind), instead of beliefs about how categories predict properties (e.g., how membership in a gender or an animal category predicts what properties an individual will have; for further discussion of this point, see the General Discussion).

Documenting that children reason similarly about animal and gender categories by age three, representing both as natural kinds, has important theoretical implications for understanding conceptual development in the social domain. Atran (1990) proposed that the conceptual system for natural kinds emerges first for representing biological species kinds (e.g., animal species), and that, later in development, children transfer this

conceptual system to the social domain (e.g., to represent human gender or race, see also Gil-White, 1999). Alternately, Gelman (2003) and Hirschfeld (1996) propose that the conceptual system for natural kinds does not begin as linked to any particular domain, and that this more general conceptual system is recruited separately to represent species kinds and human social categories (see discussions in Gelman & Hirschfeld, 1999, and in Taylor et al., 2009).

These two proposals make different developmental predictions. In particular, according to the transfer model (Atran, 1990), early in development, children should represent only animal (and perhaps plant) categories as natural kinds. Gender should not be viewed as a natural kind until later in development, and concepts of animal species and of human social categories should converge with age. In contrast, if a more general conceptual system for natural kinds is recruited to represent both animal and human categories (Gelman & Hirschfeld, 1999), then we should find similarity in the concepts of younger children across domains, and perhaps more divergence with age, as children attend to a broader range of domain-dependent factors (e.g., for human categories: personality, skills, experiences).

Because the present study found evidence of early similarities in children's reasoning about animal species and human gender, these data are consistent with the possibility that the conceptual system for natural kinds is applied across multiple domains simultaneously. This pattern does not appear to be consistent with the proposal that this conceptual system is initially linked to the domain of animals and then transferred to represent social kinds. Although the present findings leave open the possibility that such a transfer occurs before age 3, the developmental pattern from Studies 1-3, which showed

increasing divergence in concepts of animals and gender across age in the midsize city, also appears inconsistent with the transfer hypothesis (see also Taylor et al., 2009). It is interesting to note that the proposal that children transfer the conceptual system for natural kinds to new domains across age (Atran, 1990) *is* consistent with the pattern obtained for race in studies 2 and 3 in the rural area. Thus, understanding how particular categories come to be represented as natural kinds across age appears to require more fine-grained considerations (of both specific categories within domains and of cultural context).

The present findings are also the first that I am aware of to document domain-specific beliefs about the ontological status of category boundaries among such young preschoolers. Like 5-year-olds in the previous studies, children as young as three appear to view artifact categories and racial categories as more subjective and flexible than those based on animal species or human gender. For example, when children were told that the experimenter was looking for two things of the same kind, they often identified the sought-after object as an artifact from the same superordinate class, but different basic-level category (e.g., when the target was a hammer, they chose the visible saw). Similarly, when the experimenter was looking for two people of the same kind, children often selected pictures that differed from the target in terms of physical cues to race.

Because these children were younger than those that participated in the previous studies, it will be important in future work to conduct control studies to document that these younger preschoolers, like 5-year-olds, were equally attentive to the relevant categorization criteria in each domain. For example, without follow-up control studies (similar to those conducted in Study 4), I cannot rule out the possibility that these

younger preschoolers failed to notice the physical cues to race in this experiment. Also, an important possibility is that superordinate categories are more familiar to young children for artifacts (e.g., tools, clothing) than for animals (e.g., mammals, reptiles). In future work, it will be necessary to pin down exactly why children had a flexible approach to artifact and racial categorization in this context. The present findings suggest, however, that children view the differences between basic-level animal categories, and between people of different genders, as very salient and fundamental to identity, but the distinctions between basic-level artifact categories, and between people of different races, as less so.

That children have these domain-specific ontological commitments by age 3 is consistent with the proposal that domain-specificity is a primitive component of human cognition (Hirschfeld & Gelman, 1994). Alternate perspectives suggest that domain-specificity is the product of learning histories and experience (Sloman, Lombrozo, & Malt, 2007). Whereas Study 5 cannot rule out the possibility that the present domain-dependent responses occurred because of learning that occurred prior to age 3, these findings present a challenge to domain-general learning accounts to identify learning mechanisms that could account for this pattern of abstract domain-dependent beliefs so early in childhood (see Sloutsky & Fisher, 2008).

The findings from Study 5 replicated the domain effects found in Studies 1-3 in a younger sample of children and using a different experimental task. Thus, this work provides evidence of the robustness of these domain effects in early childhood. Based on the findings of Studies 2-3, I would not expect there to be variation in the responses of

younger children by community (e.g., across the two communities included in the present dissertation), however, this remains an important area for future work.

Chapter 7

Study 6

The aim of Study 6 was to compare the development of another component of gender concepts in children drawn from the same cultural contexts sampled from in Studies 2 and 3. The previous studies in this dissertation documented an interaction between age and cultural context in the development of gender concepts. Younger children (ages 5 and 7) in both communities represented gender as a natural kind, as indicated by the consistent finding that children of these ages reliably identified categories that grouped together children of different genders as "wrong". With age, children in the midsize city came to construe gender categories as more conventional and flexible (they accepted social categories that did not follow gender about half of the time by ages 10 and 17). In contrast, there were no developmental changes for gender concepts in the rural community; older children construed gender as a natural kind to the same extent as younger children did. Overall, this pattern suggests that gender concepts are fairly rigid in early childhood, but may become more flexible, depending on children's cultural experiences.

These findings are somewhat consistent with previous work on gender concepts. For example, a number of other studies have documented fairly strict gender concepts among young children in the United States (Berndt & Heller, 1986; Biernat, 1991;

Rhodes & Brickman, 2008; Rhodes & Gelman, 2008; Taylor, 1996; Taylor et al., 2009). Consistent with the present findings from the midsize city, each of these previous studies has also reported that gender concepts become more flexible with age. For example, Berndt and Heller (1986) found that younger children relied on membership in gender categories to predict individual behavior (e.g., they predicted that a girl would prefer to bake brownies than go fishing, regardless of whether she previously demonstrated a dislike for baking). In contrast, adults attended more to individuating factors (e.g., they predicted that a girl who previously demonstrated that she disliked baking would prefer to go fishing instead of bake brownies). Also, Taylor (1996) found that younger children view gendered behaviors (e.g., liking ballet for girls and football for boys) as an inevitable part of being born a member of a particular gender group, whereas older children and adults viewed such behaviors as the product of providing children with particular environments.

In previous work, developmental changes in gender concepts have often been attributed to domain-general increases in cognitive flexibility (Bigler, 1995). The findings from Studies 2 and 3 in the rural community, however, suggest instead that developmental changes in gender concepts depend on specific cultural experiences. The goal of Study 6 was to link the present findings regarding gender to this previous body of work.

The method used in Studies 2 and 3 assessed a different component of gender concepts than has been measured in prior work. Particularly, I assessed whether children view gender as an objectively correct, natural way to classify people, whereas previous work has examined children's beliefs about the power of gender categories in

determining social behavior. Thus, one possibility is that there are cultural differences only in children's beliefs about the nature of social kinds (e.g., the extent to which kinds are objectively determined), and not in children's beliefs about gendered *behavior*. For example, Evans (2001) and Diesendruck and Haber (2009) have proposed that there is specific cultural variation in the extent to which individuals view social and biological categories as objectively determined, and that this variation relates to community-level differences in religious beliefs (e.g., the extent to which people view specific categories as created by God). Thus, one possibility is that cultural differences in social concepts relate only to beliefs about the objectivity of social kinds.

An alternate possibility, however, is that the cultural differences obtained in the earlier studies would also be found in other aspects of children's gender concepts. For example, beliefs about category objectivity may correspond to beliefs about inductive potential, in that categories that are understood to represent natural, objectively defined groupings are also understood to be those whose members share many properties (Gelman & Coley, 1991). Thus, viewing gender as a natural, objectively correct way to categorize may be related to the belief that gender has a powerful role in shaping human behavior. If so, then there may also be cultural differences in the extent to which beliefs about the inductive power of gender change across age, which perhaps have been overlooked in previous work.

The present study examines how beliefs about the inductive power of gender categories change across age among children from the two communities sampled from in Studies 2-4. This study presented children with a "switched-at-birth" reasoning task (Gelman & Wellman, 1991; Hirschfeld, 1996; Taylor, 1996). In this task, children were

told about a male infant who is raised from birth entirely by females, or about a female infant who is raised from birth entirely by males. Test questions probe what children think the target character will be like when he or she has grown up—will the character develop the physical or behavioral characteristics associated with the birth gender category or the gender category present in their environment? Thus, these tasks test whether children view the development of gender-linked properties as an inherent part of being born a particular gender, or as dependent on environmental influence. Prior research, conducted in communities similar to the midsize city sample from in Studies 1-4, has found that younger children (5-year-olds) generally predict that the characters in these stories will have both the physical and stereotypical behavioral properties (e.g., dressing up like a princess for girls vs. as a soldier for boys) associated with their birth gender category (Taylor, 1996; Taylor et al., 2009). In contrast, older children (10-year-olds) and adults generally predict that the story characters will have the physical properties associated with their birth category, but the behavioral properties associated with the gender category in their environment. Thus, younger children appear to believe that being born with a particular gender powerfully influences behavior—regardless of the environment—whereas older children and adults view stereotypic gender behaviors as more dependent on the environment. The goal of the present study was to test whether this developmental trajectory varies across cultural contexts.

Method

Participants

Participants include the same children who participated in Study 3, including the 5-year-olds, 7-year-olds, 10-year-olds, and 17-year-olds from the rural area, and 7- and 17- year-olds from the midsize city (for specific age ranges, see Study 3).

Procedure

After completion of Study 3, participants completed a switched-at-birth gender task, following the procedures used by Taylor (1996) and Taylor et al. (2009). Because Taylor et al. found differences in children's reasoning about male and female characters, participants were randomly assigned to be told about either a male infant who was raised entirely by women or a female infant who was raised entirely by men. The procedures for children are described here; 17-year-olds completed the task on their own in a paper test booklet.

For children, participants were told, "Now, I'm going to tell you about a baby [girl/boy] named Arza. Look, here's a picture of Arza when Arza was a baby. [*Show Arza*.] Right after Arza was born, when Arza was just a tiny baby, Arza went to live with [her uncle/ his aunt] on an island. See, here is Arza's [uncle/aunt]. [*Show picture*.] Can you point to Arza? Can you point to Arza's [uncle/aunt]? Now, I have to tell you something about this island. On this island there were only [boys and men/ girls and women]. Arza was the only [girl/boy]. Arza's uncle loved Arza and took care of Arza. Arza lived with her [uncle/aunt] and became part of their family. Arza grew up on the island with only [boys and men/ girls and women] and had a happy life, but Arza never got to see another [girl or woman/ boy or man]."

Then, participants were asked a series of control questions, "Is Arza a girl or a boy?", "Does Arza live with boys and men on the island?", "Does Arza live with girls

and women on the island?", and "Are there any other [girls or women/ boys or men] on the island?" If participants answered any of these questions incorrectly, the story was retold and the questions were asked again.

Next, participants were told, "Now, I have a picture of Arza now that Arza is a big kid. This picture is a surprise for later, but I want you to think about what Arza is like now that Arza is a big kid, and I'm going to ask you some questions."

Participants were then asked a series of six questions, three about physical properties, and three about stereotypical behavioral properties. Properties were blocked by type (physical, behavioral), and order of presentation of the blocks was counter-balanced across participants. Property questions all began with the stem, "When Arza is a big kid…" and included, "does Arza have a boy's blood inside or a girl's blood inside? Does Arza have a boy's body or a girl's body? Does Arza have a boy's brain or a girl's brain? Does Arza play with a tea set or does Arza play with a toy truck? Does Arza like to go fishing or put on make-up? Does Arza play dress-up or with baseball cards?" Responses in which participants indicated that Arza would have properties associated with his/her birth gender category were scored as '1', responses indicating that Arza would have the properties associated with the gender category in the environment were scored as '0'.

Results

Mean levels of birth-category predictions, and comparisons to chance, are presented in Table 7. The first set of analyses focused exclusively on the rural community, as all age groups were available from this community, and developmental changes in beliefs about gender in the rural community were the primary focus of this

study. Subsequent analyses compare children of equivalent age groups from the midsize city and rural area.

Data from the rural community were examined through a 4 (age) X 2 (gender of target: boy, girl) X 2 (property: behavioral, physical) GEE. These analyses revealed significant effects for age $\chi^2(3) = 10.59$, $p < .05$, target-gender, $\chi^2(1) = 9.32$, property, $\chi^2(1) = 31.56$, $p < .001$, and interactions between age and property, $\chi^2(3) = 20.22$, $p < .001$, target-gender X property, $\chi^2(1) = 3.81$, $p = .05$, and age X target-gender X property, $\chi^2(3) = 9.31$, $p < .05$. Follow-up analyses revealed that there were no effects of age or target-gender on predictions about physical properties; as shown in Table 7, children of each age reliably predicted that the character would have the physical properties associated with their birth category.

For behavioral properties, however, there were effects of age, $\chi^2(3) = 13.85$, $p < .001$, and target-gender, $\chi^2(1) = 13.85$, $p < .001$. As shown in Table 7, younger children reliably thought that characters would have the behavioral properties associated with their birth-category, whereas older children did not. For behavioral properties, birth-category predictions declined with age, such that 5- and 7- year-olds made more birth-category predictions for behavioral properties than did 10- or 17-year-olds. Also, overall, children predicted that male characters would have more of the behavioral properties associated with their birth category than would female characters (as in Taylor et al., 2009).

Next, I compared the data of 7- and 17-year-olds from the two communities through a 2 (cultural context) X 2 (age) X (target-gender) X 2 (property) GEE. This analysis revealed effects of target-gender, $\chi^2(1) = 16.23$, $p < .001$, cultural context, $\chi^2(1)$ 3.99, $p < .05$, property, $\chi^2(1) = 100.39$, $p < .001$, age X property, $\chi^2(1)$ 17.81, $p < .001$,

age X target-gender, $\chi^2(1) = 17.81$, property X target-gender, $\chi^2(1) = 5.58$, $p < .05$, and a four-way interaction between age, target-gender, location, and domain, $\chi^2(1) = 10.57$, $p < .001$.

To decompose this four-way interaction, I conducted separate analyses for predictions about physical and behavioral properties. For physical properties, there was an effect only of age, $\chi^2(1) = 3.79$, $p = .05$, such that 7-year-olds ($M = .81$) made fewer birth-category predictions than 17-year-olds ($M = .91$); however, both ages made more birth-category predictions than expected by chance. For behavioral properties, there were effects of age, $\chi^2(1) = 9.93$, $p < .01$, target-gender, $\chi^2(1) = 19.21$, $p < .001$, cultural context, $\chi^2(1) = 6.48$, $p < .05$, and a three-way interaction between age, target-gender, and cultural context, $\chi^2(1) = 6.79$, $p < .01$.

Among 2nd graders, children in the midsize city made more birth-category predictions about the behavior of boy targets than the behavior of girl targets. As shown in Table 7, for 7-year-olds in the midsize city, their responses to questions about the behaviors of boys did not differ from chance, whereas they reliably made *environment*-based predictions about the behavior of girls. In contrast, in the rural community, responses to male-targets and female-targets did not differ; for both male and female targets, 7-year-olds in the rural community made more birth-category predictions than expected by chance. Thus, at age 7, children in the rural community saw gender-linked behaviors as more strongly determined by birth categories than did children in the midsize city, particularly for female behavior.

Among 12th graders, participants from the midsize city reliably made environment-based predictions regarding the behavior of both male and female targets.

In contrast, 12th graders in the rural community did so only for female targets; they made birth-category predictions about half of the time for male targets. Thus, by age 17, children in both communities had similar attitudes about female behavior; in both communities, stereotypic behavior for females was considered to be environment-dependent. In contrast, the belief that male stereotypic behavior is environment-dependent had developed only in the midsize city.

Discussion

The findings from Study 6 suggest that developmental changes in gender concepts vary across local cultural contexts. Thus, this study indicates that the patterns found in Taylor (1996) and Taylor et al. (2009), in which gender concepts become more flexible across age, do not represent universal developmental trajectories. In the present study, the belief that stereotypic female behavior is dependent on the environment developed at an earlier age in the midsize city, as compared to in the rural area. In particular, 7-year-olds endorsed this view in the midsize city, but not in the rural area. In the rural area, children did not reliably predict that female behavior would be influenced by the environment until age 17. For beliefs about male behavior, there was also an important effect of cultural context. In the midsize city, neither age group saw male behavior as entirely determined by the birth-category and 17-year-olds reliably judged male behavior as dependent on the environment. In contrast, in the rural community, younger children reliably predicted that male behavior would be determined by the birth category, and at no age did participants reliably judge that male behavior would be influenced by the environment of upbringing. Thus, across communities and ages, there is evidence both that flexible beliefs about female behavior develop more readily than flexible beliefs

about male behavior, as well as that the development of such beliefs is importantly influenced by children's cultural contexts.

These findings suggest that the influence of cultural context on gender concepts is not restricted to beliefs about the objectivity of gender as a means of social classification. Instead, the effect of cultural context also extends to children's beliefs about the role of gender categories in determining human behavior. These findings are consistent with those from Studies 2 and 3, which indicated that in the rural area, children continued to construe gender as a natural kind across childhood, whereas in the midsize city, children developed the belief that gender is a flexile and subjective means of social classification by age 10. Across all of the studies in this dissertation, developmental changes in gender concepts depended on children's cultural context.

An important issue for future work will be to determine why gender concepts become more flexible with age in some communities, but not in others. In prior work documenting increased flexibility in gender concepts with age, researchers have proposed that such changes are caused by domain-general increases in cognitive flexibility. For example, Bigler (1995) suggested that domain-general improvements in multiple classification skills account for the decreasing salience of gender categories as children get older. The present findings suggest, however, that such changes are not the inevitable consequence of improved flexible thinking, but instead, depend on children's cultural experiences.

The theory-change model of conceptual development (Gopnik & Wellman, 1994; Gopnik & Meltzoff, 1997) may provide a more useful framework for understanding how gender concepts change across age. Within this framework, children are thought to rely

on naïve theories to explain and predict events in their environment. Often, children's initial theories are incomplete, however, such that they sometimes yield incorrect predictions. The accumulation of these unexpected experiences—where reality is inconsistent with predictions—eventually leads children to revise their theories to be able to more accurately explain and predict their environment. Thus, experiences that yield data that are inconsistent with children's initial theories prompt conceptual change.

Drawing on this framework, Taylor et al. (2009) proposed that children revise their gender theories across childhood. In that work, 5-year-olds appeared to hold gender theories which indicated that membership in gender categories directly causes the development of particular behaviors. Evidence that younger children hold such theories was found in their predictions about gendered behavior (e.g., younger children predicted that a baby who was born female but raised by all males would still prefer ballet to football), as well as in children's explanations for these predictions. For example, when 5-year-olds were asked to explain why this girl would like ballet, they very often responded with "because she is a girl". That is, they referenced category membership as a direct cause of behavior. Also, when children were given a choice of several causal mechanisms that could explain the behavior of interest, younger children tended to endorse internal mechanisms (e.g., she was born liking ballet) and to reject external mechanisms (e.g., she was taught to like ballet). Thus, 5-year-olds appeared to view gender as directly causing behavior, and to view internal mechanisms as responsible for these effects. Such a theory should lead children to predict that members of gender categories will consistently engage in stereotypical behaviors, regardless of individuating factors.

In contrast, in Taylor et al. (2009), 10-year-olds' theories of gender appeared to include the beliefs that stereotypical gendered behaviors are associated with, not causally linked to, membership in gender categories, as well as that the link between category membership and behavior depends on the environment of upbringing. For example, older children reliably predicted that a girl raised with boys would prefer football to ballet. These older children also tended to generate environment-based explanations for these behaviors (e.g., because her friends like football) and to endorse external causal mechanisms (e.g., teaching, environmental pressure) instead of internal mechanisms. A minority of the time, older children did predict that characters would engage in behaviors associated with their birth category (e.g., that the girl described above would prefer ballet). However, even when they made these predictions, older children still did not view birth categories as directly causing behavior. Instead, they viewed gendered behaviors as associated with, but not directly caused by, gender categories (see Prasada & Dillingham, 2006). For example, 10-year-olds explained that a girl (who was raised with boys) would like ballet by referencing associations (e.g., "because girls usually like ballet better."). Thus, 5- and 10-year-olds differed not only in the frequency with which they made birth-category based predictions (with 10-year-olds less likely to make such predictions than 5-year-olds), but also in their explanations for why such behaviors developed.

In the present study, children in the midsize city appeared to revise their theories of gender across childhood, much like the children in Taylor et al. (2009). However, children in the rural area revised their theories of female behavior much more slowly than did children in the midsize city, and they had not fully revised their theories of male behavior by age 17. Within the concepts-as-theories model described above, the younger

child's theory indicates that gendered behavior is causally linked to category membership, and predicts that such behaviors will develop regardless of individuating factors (e.g., one's personality, skills, or environment). Thus, observations of counter-stereotypic behaviors (e.g., a boy who likes ballet) provide evidence that is inconsistent with younger children's theory-driven predictions. Within this framework, how much children are exposed to counter-stereotypic behaviors should predict when (and whether) they revise these initial theories.

Thus, one possibility is that children in the rural area have less exposure to counter-stereotypical gender behavior. For example, if social pressures to conform to gender norms are felt more strongly in the rural area, then children may encounter fewer individuals who engage in behaviors that conflict with their expectations (e.g., they may encounter fewer boys who do ballet). In contrast, if social pressures to conform are felt less strongly in the midsize city, children may encounter more examples of individuals that disconfirm their expectations. If children in the midsize city encounter many counter-stereotypical examples, this may eventually lead them to revise their beliefs that gendered behaviors are a direct consequence of category membership. Alternately, children in the rural area may not have as many experiences, and thus not encounter the need to revise their theory. In future work, it will be important to document both whether the frequency of counter-stereotypic behaviors varies by context, as well as to experimentally test whether exposure to counter-stereotypic gender behaviors prompts children to revise their theories.

Another possible contributor to cultural variability in gender concepts could relate to linguistic input. In particular, generic noun phrases (e.g., "girls like baking")

communicate information about a category as an abstract whole (Carlson & Pelletier, 1995; Cimpian & Markman, 2008; Gelman & Raman, 2003). Such sentences differ from non-generic sentences (e.g., "this girl likes baking"; "many girls like baking") by implying that the referenced categories are relatively homogeneous, and that many properties will be shared across all members of the category. Previous research has demonstrated that generic noun phrases involving gender are common in parental speech to young children (Gelman, Taylor, & Nguyen, 2004). There is also evidence that exposure to generic noun phrases causally influences children's concepts. For example, children who hear information about novel categories stated generically view category memberships as more stable and more inductively rich than children who hear similar information expressed in non-generic forms (Gelman, Ware, & Kleinberg, 2009). Thus, if children in the rural area hear gender information expressed in generic form more often than children in the midsize city, this could account for why children in the rural area continue to view gender as a coherent and inductively rich social category. In future work, it will be important to examine whether differences in parental language do vary in this manner, as well as whether such differences in parental input are causally linked to children's concepts.

Overall, the studies in this dissertation indicate that the effects of cultural context on children's social categories span several aspects of conceptual structure. Whereas Studies 2 and 3 documented variability in the belief that gender is an objective way to classify people, Study 6 found similar cross-cultural differences in beliefs about the relations between gender categories and behaviors. These studies also point to an important need to closely examine the mechanisms that account for developmental

changes in gender concepts across childhood, as well the processes that contribute to cultural variability.

Chapter 8

General Discussion

The studies in this dissertation indicate that, by age three, children make systematic, domain-specific distinctions between natural and conventional categories. Particularly, young children viewed animal and gender categories as objectively correct, natural ways to categorize their environment. They appeared to view basic-level animal categories and gender categories as fundamental ways of defining what it means to be an animal, or a person, respectively. In contrast, they viewed artifact classification and social categorization based on race as relatively conventionalized, flexible, and subjective. Several control conditions documented that young children were aware of commonly used classification criteria in each domain. For example, when they were asked to form their own social categories, they used physical appearance differences indicative of race as often as they used gender. Children were also equally likely to use basic-level criteria to form their own categories of animals and artifacts. Thus, domain differences in children's responses do not appear to relate to differences in their awareness of common categories, or in the salience of these categories. Domain differences in beliefs about naturalness and conventionality were also not predicted by the accessibility of superordinate or thematic categories, perceptual features of the stimuli, or linguistic cues (see Study 4).

Taken together, all of these data suggest that, from a young age, children incorporate systematic beliefs about the ontological status of categories into their concepts of animals, artifacts, and people. These findings are consistent with a theory-driven approach to early cognitive development, which suggests that domain-specific naïve theories are powerful motivators of conceptual development (Gelman & Koenig, 2003; Wellman & Gelman, 1992). For example, the expectation that the domain of animals is composed of discrete kinds may propel knowledge acquisition about the biological world (Atran, 1990). Similarly, an expectation that artifact kinds are conventionalized could facilitate knowledge acquisition for artifacts by encouraging children to learn from others about the artifacts in their local environment (see Kalish & Sabbagh, 2007). The hallmark of theory-driven approaches to cognitive development is that these early emerging expectations, in this case about the ontological status of categories by domain, facilitate early cognitive development, as opposed to being the product of an extended period of learning and developmental change.

The present studies extend prior work on early theory-based concepts by explicitly examining children's beliefs about category structure. Whereas prior work has documented children's distinctions between animals and artifacts for a number of dimensions (see Gelman & Koenig, 2003), previous studies had not documented domain differences in children's beliefs about whether categories are discovered vs. created by people, or have objectively-determined vs. subjectively-set boundaries. As described by Atran (1999), beliefs about the causal mechanisms that operate within a domain and beliefs about category structure are two distinct components of conceptual structure. For example, in the domain of biology, Atran proposes that there are two intuitive

expectations that universally constrain concepts. These include expectations about causality, which are proposed to focus on essentialist beliefs about identity and development (e.g., that category identity is inherited from parents, stable across development, and causes the development of species-typical properties; Gelman, 2003), as well as an expectation about taxonomy, which Atran specifies as an expectation that the biological world is composed of discrete kinds. A number of developmental and cross-cultural studies have documented evidence for the first expectation, involving causality, for animal concepts. For example, young children from very diverse cultural contexts appear to engage in essentialist reasoning about animal identity and development (Astuti et al., 2004; Sousa, Atran, & Medin, 2002; Waxman et al., 2007). The findings from the present studies are the first, however, to present direct evidence for the early development of the second expectation, involving domain-specific beliefs about the objectivity and discreteness of category boundaries.

Concepts of Animals and Artifacts

For animals and artifacts, there was clear evidence of conceptual stability across age. Younger children differentiated these categories, viewing animal categories as objectively-defined natural kinds and artifact categories as more conventional, much like older children and adults did. There was also evidence, however, that as children got older, they were increasingly likely to view both animal and artifact categories as conventional, consistently accepting unexpected categories (both animals and artifacts) more often at older ages than at younger ages. Interestingly, whereas the cultural context X age interaction was never significant for animals or artifacts, more pronounced developmental change appeared to occur in the midsize city for both domains (see Tables

4 and 5). For example, in the midsize city there was a 20% reduction in the proportion of artifact category rejections from the younger to older participants in both Studies 2 and 3, whereas in the rural area, these reductions were much smaller.

In future work, it will be important to examine both why concepts of animals and artifacts become more flexible with age, as well as why this change is more pronounced in some communities. One possibility is that, with age, children's judgments are less influenced by relatively superficial features (e.g., that the two items have different names or shapes), and more influenced by their ontological commitments. Similarly, with age, children may accumulate more experiences that justify the belief that categorization varies across contexts. For example, they may have experiences with cultural variability in artifact use, and with superordinate animal categories. Also, improvement in domain-general skills for flexible thinking could contribute to these changes. The patterns for 5- and 7-year-olds suggest that such skills in cognitive flexibility are fairly early emerging; however, such abilities may also improve with age.

These explanations leave open, however, the question of why developmental changes for animals and artifacts varied by context. One possibility is that these differences reflect only differences in response patterns. For example, because 17-year-olds in the midsize city accepted more of the unexpected social categories than did 17-year-olds in the rural area, they may have carried over this tendency to accept unexpected categories to the other domains. Another possibility, however, is that these differences by context reflect stable individual differences in variables such as believing that category boundaries are important for maintaining order (Jost et al., 2003) or endorsing

cultural relativism. In future work, it will be interesting to examine directly these various possibilities.

Concepts of Gender

For social categories, I obtained different developmental trajectories for concepts of gender and race, as well as important effects of cultural context. Young children in both communities viewed gender as a natural kind category. This is consistent with a number of other developmental studies on gender concepts, which suggest that young children engage in essentialist thinking about gender. For example, preschoolers view gender categories as strongly predictive of behavior (Berndt & Heller, 1986; Biernat, 1991; Taylor, 1996; Taylor et al., 2009; Study 6 of this dissertation) and expect two people of the same gender category to share many physical and behavioral properties, even if they differ in appearances (Gelman, Collman, & Maccoby, 1986). Also, consistent with the present findings from the midsize city, there is previous evidence that young children have stricter gender concepts than older children and adults do. For example, younger children view gender categories as directly causing a range of behavioral properties, whereas adults view such behaviors as more dependent on individuating factors and the environment (see Taylor, 1996; Taylor et al., 2009; Study 6).

The present findings suggest, however, that developmental changes in gender concepts importantly vary by cultural context. In the midsize city, children viewed gender as relatively conventional by age 10, but in the rural community, both older and younger children construed gender as a natural kind. These findings highlight the importance of considering children's cultural context in studies of gender concepts.

Because much developmental research is likely to be carried out in the communities that surround research universities (i.e., that are more similar to the midsize city than the rural community), the present findings suggest that caution should be taken in generalizing findings from such communities. These studies also suggest a need to examine the mechanisms of conceptual change. In prior work, increased flexibility in gender concepts has often been proposed to result from increases in general cognitive flexibility (e.g., multiple classification skills; Bigler, 1995). The present findings suggest that domain-general factors are unlikely to fully account for developmental changes in gender concepts. Instead, there appears to be an important role of specific cultural experiences (see Study 6).

These studies suggest several interesting questions for future research on gender concepts. An important question is why the younger children's gender concepts were similar across the two communities, given that the concepts of older children varied by cultural context. If cultural learning were wholly responsible for conceptual development in the social domain, then young children's concepts should be expected to mirror those of adults in their communities. Thus, one possibility is that early concepts of gender are constrained by intuitive biases. For example, Cosmides et al. (2003) suggest that there are evolved cognitive mechanisms for representing gender, due to its evolutionary significance. From this perspective, children bring to the task of social categorization the expectations that gender categories are fundamentally distinct. The current data are consistent with this proposal, although it is important to note that early concepts of gender as a natural kind appear to be open to revision across childhood.

Although these data are consistent with the possibility that there is innate support for gender concepts, these studies fall short of fully testing this proposal. An alternate possibility is that the social context that young children experience is markedly different from that experienced by older children, as well as more uniform across communities. For example, parents and teachers could emphasize gender more when communicating with young children, and when selecting their toys, clothing, and activities, thus leading to a heightened salience of gender for young children across communities. On this account, some communities may focus less on gender norms as children age, whereas other communities may maintain this focus. Another possibility is that gender categories are an early emerging social category because gender is presented to children as *binary*, and children have a great deal of experience with both genders. Other social categories (e.g., based on race) do not have an easily learned binary structure. Thus, gender may emerge as a salient category to young children because it is easy to learn (see Landau & Gleitman, 1985; Waxman & Shumer, 2008).

In future work, it will be important to examine the extent to which rich concepts of gender emerge as a result of intuitive biases, learning, and/or the particular social environment presented to young children. On any of these three accounts, early gender concepts result from a constructive process on the part of children. That is, gender concepts do *not* seem to result from a simple assimilation of adults' concepts. Examining more directly the process by which gender emerges as a natural kind category in early childhood will require testing children from more distinctly different cultural contexts, as well as younger children and infants.

Concepts of Race

Concepts of race presented a different developmental trajectory. Although the control conditions confirmed that by kindergarten age, young children were aware of race, and in fact used physical cues to race to classify people when they were asked to use their own criteria, younger children in both communities viewed categorization based on race as conventionalized and subjective. Natural kind concepts of race emerged later, and only in the rural community.

These findings are consistent with the proposal that distinct developmental processes support the acquisition of race and gender categories (Cosmides et al., 2003), and are inconsistent with the influential proposal that there are intuitive cognitive biases to represent race as a natural kind (Gil-White, 2001; Hirschfeld, 1996). Hirschfeld (1996) proposed that children have an innate "human-kind making" competence, which leads them to interpret categories of people (particularly those based on race) as marking fundamental, discrete boundaries. The evidence in support of this proposal comes from "switched-at-birth" studies, in which children are told that a baby is born to parents who are White, but grows up with parents who are Black, for example. Children are then asked to predict what the baby will look like as an older child. Such studies have demonstrated that children as young as three understand skin color as an inherited and stable property, impressively demonstrating that preschoolers have a biological understanding of physical appearance. In my view, however, they fall short of establishing that children view race as an important means of social classification, in that they do not assess whether young children view physical cues to race as having *social* significance (e.g., in predicting a person's social identity or behavioral properties).

Indeed, the present studies suggest that although young children are aware of physical cues to race, they view use of such cues for social categorization as flexible and subjective. Similarly, several other recent studies have also found that children fail to place special significance on race in making social judgments. For example, preschoolers rely on social categories based on gender and language, but not race, to make inferences about people's friends, toys, and activity choices (Kinzler, Shutts, DeJesus, & Spelke, in press; Shutts, Banaji, & Spelke, in press). Thus, although young children are aware of physical appearance-based categories, and view physical appearances indicative of race as inherited, the present studies and other recent work suggest that they do not represent race as a natural kind; instead, they view racial categories as flexible and subjective. It is important to note that the majority of children in these studies were White, and likely received little direct input about race. Given these limited samples, I do not suggest that young children *cannot* develop natural kind concepts of race—they may indeed develop such concepts by preschool age in some populations. Rather, these studies provide evidence against universal cognitive biases to do so. In future work, it will be important to include children from more diverse racial-ethnic backgrounds and communities.

Another important issue for future work is the process by which race concepts develop. Hirschfeld (1993) suggested that changes in race concepts occur with age as children become better able to match perceptual differences in appearance with their conceptual beliefs about human kinds. Thus, he proposed that young children have the conceptual belief that racial groups divide people into natural kinds, but are poor at recognizing these kinds from perceptual cues. Within this framework, based on the data from only the younger children, and from only the experimental conditions of Studies 1-

3, one possibility would be that younger children have the *belief* that race is a natural kind, but that I did not find evidence of this belief because children were provided with only perceptual cues to race.

There are several reasons, however, why I doubt that this possibility can account for these results. First, the control conditions (see Study 1 and Study 4a) documented that 5-year-olds did not have difficulty identifying perceptual cues to race within these stimuli. In these controls, when children were asked to apply their own criteria for categorization, children categorized by race as often as by gender (and just as often as they used expected criteria for animals and artifacts), suggesting that they recognized the relevant racial categories (much like they recognized the relevant artifact categories) but saw use of such categories as flexible. The patterns for race and artifacts differed markedly from those for gender and animals, where children applied their own criteria during both experimental and control conditions. I interpret the pattern across conditions as indicating that children recognized race as a conceptual category, but saw it as a conventionally-determined, instead of natural, kind. Thus, these findings do not appear to indicate that children displayed a flexible approach to categorization on these tasks because they failed to access their conceptual knowledge about race. Nevertheless, it would be interesting in future work to examine whether children treat racial categories as more natural and objective when they are provided with relevant category labels (which Hirschfeld, 1993, suggests increases the accessibility of children's conceptual knowledge about race).

The pattern across ages also does not appear to match the framework sketched out above. Hirschfeld proposed that, with increasing age, children become better able to

match up perceptual cues of race to conceptual beliefs. If so, we might expect that this developmental process would occur faster in communities where children have exposure to diverse racial groups, which in this case, would be the midsize city. However, the opposite pattern was found; natural kind concepts emerged in older childhood only in the rural area, where children had very little exposure to racial or ethnic diversity. Thus, I interpret the present findings regarding race as indicating that developmental changes involve more than improved matching between perceptual cues and conceptual beliefs, and instead include deeper conceptual changes (from the belief that race is one of many criteria for social categorization to the belief that use of race is fundamental and objective) that depend on cultural experiences.

Another important area for future work involves examining the relation between political conservatism and beliefs about the naturalness of racial categories. In the present studies, parental levels of political conservatism correlated with children's concepts of race, as did adults' own levels of political conservatism in Study 1. The basis for this relationship, however, remains unclear. One important question involves the extent to which this association is specific to political conservatism, or whether such an association was found in the present study because political conservatism served as a proxy for some other indicator (e.g., exposure to racial and ethnic diversity, economic status, etc.). For example, Evans (2001) and Diesendruck and Haber (2009) have suggested that religiosity importantly influences children's beliefs about kinds, such that children from more religious communities view categories as more natural and inflexible than do children from less religious communities. In particular, these researchers have proposed that beliefs that categories exist because they were specified by God is related

to the belief that categorization is an objective process. Thus, one possibility is that the effects of cultural context documented in the present work stem from community-level differences in religiosity, instead of in political conservatism.

In the present studies, although information on religiosity was collected from all parents who provided data on political attitudes, religiosity was not found to correlate with children's concepts of race ($p > .50$). These studies, however, were not specifically designed to evaluate whether political conservatism or religiosity plays a more important role in guiding this aspect of conceptual development, and thus are somewhat limited in their ability to address this question. For example, the number of parents who provided data was fairly small (see the supplemental analyses following Study 3) and the measures were fairly insensitive. Thus, the finding that parental political conservatism correlated with children's beliefs about race, whereas religiosity did not, is informative, but does not resolve this issue. Instead, these findings suggest the need for future work examining more directly how parental attitudes and beliefs influence children's concepts. Such work should examine a broader range of parental factors in more detail, and should also begin to move toward tests of possible mechanisms of transmission between parental beliefs and children's concepts (e.g., language use, exposure to diversity, and so on).

The Development of Social Categorization

The findings for gender and race suggest that there are multiple conceptual systems that contribute to the development of social categorization. The conceptual system for natural kinds, which includes the expectation that categories reflect an underlying natural reality, appears to contribute to the early development of gender concepts. The studies in this dissertation documented that the expectation that gender

categories indicate objectively distinct kinds of people is very early emerging, even in communities where older children and adults have more flexible gender beliefs. These findings are consistent with other work suggesting that children are sensitive to gender from early infancy (Leinbach, & Fagot, 1993), that toddlers use gender labels to guide their own interactions with new objects and activities (Bradford, Martin, Endsley, & Halverson, 1986), and that preschoolers use gender to guide their preferences (Shutts et al., in press) and achievement behaviors (Rhodes & Brickman, 2008). That gender is such an important category in early childhood, and more important than other salient categories (such as race), is consistent with the possibility that there is a dedicated conceptual system for gender, which is active very early in development.

In contrast, concepts of race as a natural kind were later emerging developmentally, as well as more variable across contexts. Also, individual differences in children's family experiences (as indicating by parental conservatism) were related to differences in children's beliefs about race. This evidence points to the conclusion that the conceptual underpinnings of race are flexible and dependent on cultural input.

In future work, it will be important to examine the conceptual underpinnings of race categories in more detail. One model of social categorization that has received little direct examination but would appear to predict the present data is the coalitional psychology hypothesis proposed by Kurzban, Tooby, and Cosmides (2001; see also Cosmides et al., 2003; Kurban & Leary, 2001; Tooby & Cosmides, 1988). On this account, concepts of race are supported by a cognitive system for reasoning about social alliances, within-group cooperation, and between-group competition, such that race takes on meaning as a marker of important social differences when it is correlated with the

structure of social alliances. From this perspective, young children would not be expected to have natural kind concepts of race (although they may be aware of and recognize racial groupings). Instead, concepts of race are expected to develop over time, particularly in contexts where input leads children to view race as a marker of social alliances. The form of this input has not been specified, but could involve direct observation, implicit communication (e.g., in the form of generic language, Gelman et al., 2004), media portrayals, and so forth.

Within this framework, whether a particular category takes on significance to children depends on the extent to which it serves as a good marker of social alliances. Thus, this framework can explain why race takes on increased meaning in some communities, but not in others, and predicts that the conceptual significance of other types of social categories (e.g., religion, linguistic groups) will vary by context as well. In other words, within this framework, some of social categorization is context-dependent and developmentally malleable.

In future work, it will be critical to directly test whether a conceptual system for social alliances supports the acquisition of social categories, and to examine the process by which this occurs. Consistent with the proposal that coalitional reasoning is central to the development of social categorization, there is evidence that young children are attentive to social alliances, and that they use categories based on alliance structures to make predictions about how people should behave and interact (Rhodes, Brickman, & Gelman, 2009). Also, children use coalitional categories, over other salient categories (e.g. race), to make predictions about individuals' preferences, obligations, and social networks (Rhodes et al., 2009).

The proposal that concepts of race are supported by a conceptual system for representing social alliances provides an alternative to the proposal that concepts of race are supported by the conceptual system for natural kinds (Hirschfeld, 1996). In light of the current data indicating that natural kind concepts of race do not emerge until late in childhood, such an alternate proposal, and one that can account for an extended developmental process and cross-cultural variability, is necessary. It is important to note, however, that the coalitional hypothesis also has some important similarities with Hirschfeld's proposals. In particular, the coalitional hypothesis is consistent with Hirschfeld's description of social categorization as supported by a unique, domain-specific cognitive system. Although I propose that social categorization is somewhat malleable and dependent on input, I do not think that the development of social categorization can be accounted for by general factors, such as labeling or perceptual features. Instead, children appear to form rich categories, centered on certain abstract properties (in the case of coalitional reasoning, about cooperation and competition). Thus, the coalitional hypothesis is consistent with the proposal that children bring a prepared, domain-specific cognitive system to the task of social categorization (Hirschfeld, 1994), but suggests that the skeletal framework of this system, for race (and a number of other categories), involves coalitional information.

In sum, I propose that there are at least two conceptual systems that contribute to the development of social categorization: the conceptual system for natural kinds, which, at least initially, focuses mainly or exclusively on gender, and a conceptual system for social alliances, which supports the development of social categories based on the dimensions that are salient predictors of social alliances in one's environment (including

race, religion, linguistic groups, and so on). Another important issue for future work will involve possible transfer between these two systems. For example, it will be interesting to examine whether once children identify particular categories as markers of social alliances (e.g., race), they also begin to represent those categories as natural kinds (e.g., as categories with sharp boundaries, inductive potential, etc.).

Overall, these studies highlight the value of examining conceptual development as embedded in culture (Rogoff, 2003; Waxman & Medin, 2007). I found evidence of some conceptual stability across contexts (e.g., for animals and artifacts), and also documented important variability (e.g., for social categories). Further, although there are undoubtedly important domain-general mechanisms that contribute to cognitive development (e.g., for causal inference, Gopnik et al., 2004; for statistical learning, Xu & Tenenbaum, 2007), the present studies reinforce the importance of examining how children appeal to domain-specific theories to make sense of the world of animals, artifacts, and people.

Appendix A

Essentialism Scale[1]

Directions: Circle the number that shows how much you agree or disagree with each statement. (9 = *Strongly Agree*, 1 = *Strongly Disagree*)

1. Gender is a very important part of what makes people who they are.

2. People that are the same gender have many things in common.

3. Knowing someone's gender tells you a lot about a person.

4. Gender is an all-or-none category; people are either male OR female, there is nothing in between.

5. Gender is a natural category.

6. Gender categories are important in all cultures around the world.

7. Males share an underlying property that causes them to have many similarities.

8. Females share an underlying property that causes them to have many similarities.

[1]Statements are written as they were presented for the Gender Essentialism Scale. For the Ethnicity Essentialism Scale, the word "Ethnicity" was substituted for "Gender", and items 7 and 8 were combined into one item that read, "People of the same ethnicity share an underlying property…"

Appendix B

Individual Response Patterns[1]

	Animals			Artifacts			Gender			Race		
	Reject	Incons	Accept	Reject	Incons	Accept	Reject	Incons	Accept	Reject	Incons	Accept
City												
5-years	15	1	0	5	7	4	14	1	1	3	6	7
7-years	35	1	1	17	12	8	20	2	4	13	6	7
10-years	10	2	1	1	3	9	6	3	4	2	2	9
17-years	31	5	10	11	6	29	12	5	15	3	8	19
Rural												
5- years	35	2	3	16	12	12	23	2	4	15	6	9
7-years	44	3	4	19	16	16	26	4	5	15	7	12

10-years	31	5	0	13	8	15	24	2	2	21	2	4
17-years	33	1	6	11	7	22	18	5	4	18	8	3

[1]Number of children classified as 'rejecters', 'accepters', or 'inconsistent' for Studies 2 and 3 combined, by cultural context, age, and domain. There are fewer five-year-olds and 10-year-olds from the midsize city relative to the other groups because these groups were included in Study 2 only.

Table 1

Description of experimental stimuli, Study 1

Category Options		Target
Distracters	Visitors' Selection	
	Animals	
Collie dog	Gray-white cat	Black Labrador dog
Light fur lion	Wolf	Dark fur lion
Green frog	Box turtle	Orange frog
White goat	Pink pig	White and black goat
Black horse	*White-black cow	Brown cow
	Artifacts	
Silver train	Blue car	Yellow train
Brown square table	Brown bookshelf	White round table
Mallet hammer	Screwdriver	Hammer
Silver spoon	Silver fork	Green plastic spoon
Blue shorts	*Blue dress	Red dress
	Gender	
Blonde-hair boy	Blonde-hair girl	Brown-hair boy
Brown-hair boy	Brown-hair girl	Red-hair boy
Brown-hair girl	Brown-hair boy	Blonde-hair girl
Blonde-hair girl	Blonde-hair boy	Brown-hair girl

Brown-hair boy	*Brown-hair girl	Blonde-hair girl
	Race	
Black (boy)	White (boy)	Black (boy)
Asian (girl)	White (girl)	Asian (girl)
Black (girl)	Latino (girl)	Black (girl)
Asian (boy)	Black (boy)	Asian (boy)
White (boy)	*Latino (boy)	Latino (boy)

Note. Items marked with a * are control items, in which expected categories were presented.

Table 2

Mean proportions of expected categories (baseline condition), and mean proportions of rejected categories (experimental condition), Study 1.

	Baseline Condition[1]	Experimental Condition[2]	
	M (*SE*)	*M* (*SE*)	
		Initial Question	Follow-up Question
5-years			
Animals	.91* (.04)	.84* (.07)	.79* (.11)
Artifacts	.98* (.02)	.88* (.05)	.86* (.09)
Gender	.87* (.06)	.91* (.06)	.88* (.08)
Race	.85* (.06)	.55 (.10)	.52 (.11)
Adults			
Animals	1* (00)	.92* (.03)	.70* (.09)
Artifacts	1* (00)	.83* (.06)	.47 (.09)
Gender	.84* (.07)	.52 (.11)	.25* (.09)
Race	.95* (.04)	.43 (.11)	.35 (.11)

Note. *Binomial regression models comparing the proportion of expected category-matches (baseline condition) or rejected categories (experimental condition) to the proportion expected by chance (.5), *$p < .05$.

[1]Proportion of questions on which participants formed expected categories based on basic-level animal or artifact categories, gender, or race, by age and domain.

[2]Proportion of questions on which participants rejected the unexpected categories, by age and domain.

Table 3

Experimental stimuli, Studies 2 and 3

Categories Presented by the Visitors	
Animals	
Gray-white cat	Black Labrador dog
Wolf	Dark fur lion
Box turtle	Orange frog
Pink pig	White and black goat
*White-black cow	Brown cow
Artifacts	
Blue car	Yellow train
Brown bookshelf	White round table
Screwdriver	Hammer
Silver fork	Green plastic spoon
*Blue dress	Red dress
Gender	
Blonde-hair girl	Brown-hair boy
Brown-hair girl	Red-hair boy
Brown-hair boy	Blonde-hair girl
Blonde-hair boy	Brown-hair girl
*Brown-hair girl	Blonde-hair girl

Race

White (boy)	Black (boy)
White (girl)	Asian (girl)
Latino (girl)	Black (girl)
Black (boy)	Asian (boy)
*Latino (boy)	Latino (boy)

Note. Items marked with a * are control items, in which visitors presented categories that matched at the basic-level for animals and artifacts, and for race or gender for social categories.

Table 4

Mean proportions of rejected categories, Study 2.

	Animals	Artifacts	Gender	Race
	M (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)
City				
5-years	.92* (.04)	.56 (.08)	.86* (.07)	.39 (.08)
7-years	.90* (.04)	.53 (.09)	.88* (.06)	.57 (.10)
10-years	.85* (.07)	.25* (.07)	.56 (.12)	.27* (.09)
17-years	.73* (.09)	.34* (.09)	.42 (.09)	.16* (.06)
Rural				
5-years	.84* (.07)	.53 (.07)	.83* (.07)	.58 (.10)
7-years	.79* (.06)	.53 (.06)	.75* (.09)	.47 (.09)
10-years	.87* (.05)	.49 (.09)	.92* (.06)	.72* (.09)
17-years	.81* (.08)	.42 (.08)	.75* (.10)	.70* (.09)

*Binomial regression models comparing proportion of rejected categories to the proportion expected by chance (.5), $p < .05$.

Table 5

Mean proportions of rejected categories, Study 3.

	Animals	Artifacts	Gender	Race
	M (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)
City				
7-years	.89* (.04)	.58 (.05)	.68* (.13)	.59 (.13)
17-years	.69* (.07)	.34* (.06)	.38* (.11)	.32* (.10)
Rural				
5-years	.93* (.05)	.57 (.07)	.80* (.13)	.57 (.11)
7-years	.88* (.04)	.55 (.07)	.82* (.06)	.56 (.10)
10-years	.94* (.04)	.54 (.09)	.81* (.12)	.87* (.07)
17-years	.80* (.07)	.41 (.08)	.73* (.10)	.75* (.09)

*Binomial regression models comparing proportion of rejected categories to the proportion expected by chance (.5), $p < .05$.

Table 6

Proportions of category rejections for individual items, Studies 2 and 3 combined

Categories Presented by the Visitors		Proportion of times categories were rejected[1] *M* (*SE*)
Animals		
Gray-white cat	Black Labrador dog	.90 (.02)
Wolf	Dark fur lion	.84 (.02)
Box turtle	Orange frog	.77 (.03)
Pink pig	White and black goat	.86 (.02)
*White-black cow	Brown cow	.12 (.02)
Artifacts		
Blue car	Yellow train	.53 (.03)
Brown bookshelf	White round table	.68 (.03)
Screwdriver	Hammer	.39 (.03)
Silver fork	Green plastic spoon	.32 (.03)
*Blue dress	Red dress	.09 (.02)
Gender		
Blonde-hair girl	Brown-hair boy	.72 (.02)
Brown-hair girl	Red-hair boy	.73 (.03)
Brown-hair boy	Blonde-hair girl	.71 (.03)

Blonde-hair boy	Brown-hair girl	.76 (.03)
*Brown-hair girl	Blonde-hair girl	.26 (.03)
Race		
White (boy)	Black (boy)	.54 (.03)
White (girl)	Asian (girl)	.48 (.03)
Latino (girl)	Black (girl)	.55 (.03)
Black (boy)	Asian (boy)	.54 (.03)
*Latino (boy)	Latino (boy)	.15 (.02)

Note. Items marked with a * are control items, in which visitors presented categories that matched at the basic-level for animals and artifacts, and for race or gender for social categories.

[1]Mean proportion of questions on which children rejected each category across age and cultural context in Studies 2 and 3.

Table 7

Mean proportions of birth category predictions, Study 6.

	Male Physical	Male Behavioral	Female Physical	Female Behavioral
Midsize City				
7-years	.83* (.08)	.64 (.14)	.70* (.14)	.07* (.04)
17-years	.95* (.03)	.26* (.11)	.84* (.07)	.12* (.06)
Rural				
5-years	.93*(.05)	.93* (.05)	.90* (.10)	.77* (.11)
7-years	.92* (.05)	.79* (.09)	.75* (.08)	.73* (.11)
10-years	.89* (.11)	.63 (.16)	.75* (.08)	.38 (.21)
17-years	.89* (.06)	.56 (.13)	.92* (.04)	.06*(.06)

*Binomial regression models comparing proportion of birth category predictions to the proportion expected by chance (.5), $p < .05$.

References

Aboud, F. E. (1988). *Children and prejudice.* Cambridge: Basic Blackwell.

Allport, G. (1954). *The nature of prejudice.* Cambridge: Addison-Wesley.

Astuti, R., Solomon, G.E., & Carey, S. (2004). Constraints on conceptual development. *Monographs of the Society for Research in Child Development, 69,* vii – 135.

Atran, S. (1990). *Cognitive foundations of natural history.* New York: Cambridge University Press.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences, 31,* 547-609.

Atran, S. (1999). Itzaj Maya folkbiological taxonomy: Cognitive universals and cultural particulars. In. D.L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233-284). Cambridge, MA: The MIT Press.

Berndt, T.J., & Heller, K.A. (1986). Gender stereotypes and social inferences: A developmental study. *Journal of Personality and Social Psychology, 50,* 889-898.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology, 61,* 351-365.

Bigler, R.S. (1995) The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development, 66,* 1072-1087.

Bloom, P. (1996), Intention, history, and artifact concepts. *Cognition, 60,* 1-29.

Bradford, M., Martin, C., Endsley, R., & Halverson, C. (1986). Influence of sex stereotypes on children's exploration and memory: A competence versus performance distinction. *Developmental Psychology, 22*, 481-486.

Carlson, G.M., & Pelletier, F.J. (1995). *The generic book.* Chicago: University of Chicago Press.

Cimpian, A., & Markman, E.M. (2008). Preschool children's use of cues to generic meaning. *Cognition, 107,* 19-53.

Cosmides, L., Tooby, J., and Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7,* 173-179.

Defeyter, M.A., & German, T.P. (2003). Acquiring an understanding of design: Evidence from children's insight problem solving. *Cognition, 89,* 133-155.

Diesendruck, G. (2003). Categories for names or names for categories? The interplay between domain-specific conceptual structure and language. *Language and Cognitive Processes, 18,* 759-787.

Diesendruck, G., & Gelman, S.A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review, 6,* 338-346.

Diesendruck, G., & Haber, L. (2009). God's categories: The effect of religiosity on children's teleological and essentialist beliefs about categories. *Cognition, 110,* 100-114.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14,* 164-168.

Estes, Z. (2003). Domain differences in the structure of artifactual and natural categories. *Memory and Cognition, 31,* 199-214.

Estes, Z. (2004). Confidence and gradedness in semantic categorization: Definitely somewhat artifactual, maybe absolutely natural. *Psychonomic Bulletin and Review, 11,* 1041-1047.

Evans, E.M. (2001). Cognitive and contextual factors in the emergence of diverse belief systems: Creation versus evolution. *Cognitive Psychology, 42,* 217-266.

Fisher, A.V., & Sloutsky, V.M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development, 76,* 583-597.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science, 14,* 79-106.

Gelman, S.A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20,* 65-96.

Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday thought.* New York: Oxford University Press.

Gelman, S.A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition, 76,* 91-103.

Gelman, S.A., & Coley, J.D. (1991). Language and categorization: The acquisition of natural kind terms. In S.A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146-196).

Gelman, S.A., Collman, P., & Maccoby, E.E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development, 57,* 396-404.

Gelman, S.A., Croft, W., Fu, P., Clausner, T., & Gottfried, G. (1998). Why is a pomegranate an apple? The role of shape, taxonomic relatedness, and prior lexical knowledge in children's overextensions of apple and dog. *Journal of Child Language, 25,* 267-291.

Gelman, S.A., & Hirschfeld, L.A. (1999). How biological is essentialism? In D.L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 403-446). Cambridge, MA: MIT Press.

Gelman, S.A., & Koenig, M.A. (2003). Theory-based categorization in early childhood. In: D.H. Rakison & L.M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 330-359). New York: Oxford University Press.

Gelman, S.A., & Kremer, K.E. (1991). Understanding natural cause: Children's explanations of how objects and their properties originate. *Child Development, 62,* 396-414.

Gelman, S.A., & O'Reilly, A.W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development, 59,* 876-887.

Gelman, S.A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development, 74,* 308-325.

Gelman, S.A., Taylor, M.G., & Nguyen, S.P. (2004). Mother-child conversations about gender. *Monographs of the Society for Research in Child Development, 69,* vii – 127.

Gelman, S.A., Ware, E.A., & Kleinberg, F. (2009). Effects of generic language on category essentialism. Paper presented at the biennial meeting of the Society for Research on Child Development, Denver, CO.

Gelman, S.A., & Wellman, H.M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition, 38,* 213-244.

German, T.P., & Johnson, S.C. (2002). Function and the origins of the design stance. *Journal of Cognition and Development, 3,* 279-300.

Gil-White, F.J. (2001). Are ethnic groups biological "species" to the human brain? *Current Anthropology, 42,* 515-554.

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,* 3-32.

Gopnik, A., & Meltzoff, A.N. (1997). *Words, thoughts, and theories*. Cambridge, MA: The MIT Press.

Gopnik, A., & Wellman, H.M. (1994). The theory theory. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). New York: Cambridge University Press

Greif, M.L., Kemler-Nelson, D.G., Keil, F.C., & Guitierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science, 17,* 455-459.

Hardin, J.W., & Hilbe, J.M. (2003). *Generalized estimating equations.* Boca Raton, FL: CRC Press.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39,* 113-127.

Hatano, G., & Inagaki, K. (1994). Young children's naïve theory of biology. *Cognition, 50,* 171-188.

Heath, S.B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* New York: Cambridge University Press.

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology, 21,* 517-528.

Hirschfeld, L.A. (1993). Discovering social difference: The role of appearance in the development of racial awareness. *Cognitive Psychology, 25,* 317-350.

Hirschfeld, LA. (1994). Is the acquisition of social categories based on domain-specific competence or on knowledge transfer? In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 201-233). New York: Cambridge University Press.

Hirschfeld, L.A. (1996). *Race in the making.* Cambridge: MIT Press.

Hirschfeld, L.A., & Gelman, S.A. (1994). *Mapping the mind: Domain specificity in cognition and culture.* New York: Cambridge University Press.

Jayaratne, T.E., Ybarra, O., Sheldon, J.P., Brown, T.N., Feldbaum, M., Pfeffer, C.A., Petty, E.M. (2006). White Americans' genetic lay theories of race differences and sexual orientation: Their relationship with prejudice towards Blacks, gay men, and lesbians. *Group Processes and Intergroup Relations, 9,* 77-94.

Jost, J.T., Glaser, J., Kruglanski, A.W., & Sulloway, F.J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129,* 339-375.

Kalish, C.W. (1998). Natural and artificial kinds: Are children realists or relativists about categories? *Developmental Psychology, 34,* 376-391.

Kalish, C.W. (2002). Essentialist to some degree: Beliefs about the structure of natural kind categories. *Memory and Cognition, 30,* 340-352.

Kalish, C.W., & Sabbagh, M.A. (2007). Conventionality in cognitive development: How children acquire shared representations in language, thought, and action. *New Directions in Child and Adolescent Development, 115*, 1-97.

Keil, F.C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: Bradford Book/MIT Press.

Kinzler, K.D., Shutts, K., DeJesus, J., & Spelke, E.S. (in press). Accent trumps race in guiding children's social preferences. *Social Cognition.*

Kurzban, R., & Leary, M.R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127,* 187-208.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98,* 15387-15392.

Landau, B., & Gleitman, L. (1985) *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Leinbach, M.D., & Fagot, B.I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development, 16,* 317-332.

Maccoby, E. (1988). Gender as a social category. *Developmental Psychology, 24,* 755-765.

Macrae, C.N., & Bodenhausen, G.V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51,* 93-120.

Malt, B.C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language, 29,* 289-315.

Mandler, J.M., Bauer, P.J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology, 23,* 263-298.

Mandler, J.M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development, 8,* 291-318.

Martin, C.L. (2000). Cognitive theories of gender development. In T. Eckes & H.M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 91-121). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Massey, C.M., & Gelman, R. (1988). Preschoolers' ability to decide whether a photographed unfamiliar object can move itself. *Developmental Psychology, 24,* 307-317.

Mead, M. (1930). *Coming of age in Samoa*. New York: William Marrow and Company.

Medin, D.L., & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review, 111,* 960-983.

Mervis, C.B., & Crisafi, M.A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development, 53,* 258-266.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nguyen, S.P., & Murphy, G.L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development, 74,* 1783-1806.

Piaget, J. (1929). *The child's conception of the world.* London: Routledge and Kegan Paul.

Prasada, S., & Dillingham, E.M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99,* 73-112.

Putnam, H. (1975) The meaning of "meaning." In H. Putnam, *Mind, language, and reality* (pp.215- 271). Cambridge: Cambridge University Press.

Quine, P.C. (1977). Natural Kinds. In S.P. Schwartaz (Ed.), *Naming, necessity, and natural kinds* (pp. 155-176). Ithaca, NY: Cornell University Press.

Rhodes, M., & Brickman, D. (2008). Preschoolers' responses to social comparisons involving relative failure. *Psychological Science, 19,* 968- 972.

Rhodes, M., Brickman, D., & Gelman, S.A. (2009). *Children's concepts of race and social alliances.* Paper presented at the biennial meeting of the Society for Research on Child Development, Denver, CO.

Rhodes, M., & Gelman, S.A. (2008). Categories influence predictions about individual consistency*, Child Development, 79,* 1270-1287.

Rogoff, B. (2003). *The cultural nature of human development*. New York: Oxford University Press.

Rosch, E.., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Rosengren, K., Gelman, S.A., Kalish, C.W., & McCormick, M. (1991). As time goes by: Children's early understanding of biological growth. *Child Development, 62,* 1302-1320.

Rothbart, M, & Taylor, M. (1992). Category labels and social reality: Dow we view social categories as natural kind? In G. Semin and K. Fiedler (Eds.), *Language and social cognition* (pp. 11 -36). London: Sage.

Schwartz, S.P. (1979). Natural kind terms. *Cognition, 7,* 301-315.

Shutts, K., Banaji, M.R., & Spelke, E.S. (in press). Social categories guide young children's preferences for novel objects. *Developmental Science.*

Siegel, D.R., & Callanan, M.A. (2007). Artifacts as conventional objects. *Journal of Cognition and Development, 8,* 183-203.

Simons, D.J., & Keil, F.C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition, 56,* 129-163.

Sloman, S., Lombrozo, T., & Malt, B. (2007). Ontological commitments and domain specific categorization. In M.J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 105-129). Hove, UK: Psychology Press.

Sloutsky, V.M., & Fisher, A.V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133,* 166-188.

Sloutsky, V.M., & Fisher, A.V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development, 79,* 639-651.

Sloutsky, V.M., Kloos, H., & Fisher, A.V. (2007). When looks are everything: Similarity versus kind information in early induction. *Psychological Science, 18,* 179-185.

Sousa, P., Atran, S., & Medin, D. (2002). Essentialism and folkbiology: Evidence from Brazil. *Journal of Cognition and Culture, 2,* 195-223.

Taylor, M.G. (1996). The development of children's beliefs about social and biological aspects of gender differences, *Child Development, 67,* 1555-1571.

Taylor, M.G., Rhodes, M., & Gelman, S.A. (2009). Boys will be boys, cows will be cows: Children's essentialist reasoning about gender and animal development. *Child Development, 80,* 461-481.

Taylor, S.E., Fiske, S.T., Etcoff, N.L., & Rudermann, A.J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36,* 778-793.

Tooby, J., & Cosmides, L. (1988). *The evolution of war and its cognitive foundations.* Institute for Evolutionary Studies Technical Report, No. 88-1. Palo Alto, CA: Institute for Evolutionary Studies.

Waxman, S.R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In. D.L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233-284). Cambridge: The MIT Press.

Waxman, S.R. & Medin, D. (2007). Experience and cultural models matter: Placing firm limits on childhood anthropocentrism. *Human Development, 50,* 23-30.

Waxman, S.R., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology, 43,* 294-308.

Waxman, S.R., & Shumer, D. (2008). *Naming and inductive strength of racial and gender categories: Evidence from preschool-aged children*. Unpublished manuscript.

Wellman, H.M., & Gelman, S.A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43,* 337-375.

Wellman, H.M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523-541.

Wilson, R.A. (1999). *Species: New interdisciplinary essays.* Cambridge: MIT Press.

Xu, F., & Tenenbaum, J.B. (2007). Word learning as Bayesian inference. *Psychological Review, 114,* 245-272.