

THE ESSENTIAL CHILD



ORIGINS OF ESSENTIALISM
IN EVERYDAY THOUGHT

SUSAN A. GELMAN

OXFORD SERIES IN COGNITIVE DEVELOPMENT

The Essential Child

Oxford Series in Cognitive Development

SERIES EDITORS

Paul Bloom and Susan A. Gelman

The Essential Child

Susan A. Gelman

The Essential Child

Origins of Essentialism in Everyday Thought

Susan A. Gelman

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi

Kuala Lumpur Madrid Melbourne Mexico City Nairobi

New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece

Guatemala Hungary Italy Japan Poland Portugal Singapore

South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2003 by Oxford University Press, Inc.

First published in 2003 by Oxford University Press, Inc.

198 Madison Avenue, New York, New York 10016

www.oup.com

First issued as an Oxford University Press paperback, 2005

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Gelman, Susan A.

The essential child : origins of essentialism in everyday thought/
Susan A. Gelman.

p. cm.—(Oxford series in cognitive development)

Includes bibliographical references and index.

ISBN-13 978-0-19-515406-1; 978-0-19-518198-2 (pbk.)

ISBN 0-19-515406-1; 0-19-518198-0 (pbk.)

1. Categorization (Psychology) in children. 2. Cognition in children.

3. Children—Language. I. Title. II. Series.

BF723.C27 G45 2003

155.4'1323—dc21 20022008635

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

For my parents, Jane and Robert Gelman

This page intentionally left blank

Acknowledgments

I have been working on the problem of essentialism, in one form or another, for the past twenty years. Over this time I have had the rare good fortune to collaborate with wonderful colleagues and students whose ideas have shaped my thinking in profound ways. Foremost among these is Ellen Markman, my Ph.D. adviser, who first sparked my interest in essentialism in a graduate seminar she taught while I was in my second year of graduate school. In the seminar we read widely about concepts and natural kinds, from Bruner to Kripke, from Macnamara to Quine. My term paper for this seminar was a research proposal to study children's natural kind concepts. Out of that grew our collaborative work on category-based induction, and ultimately this book. Ellen has been the best mentor and collaborator imaginable.

While at the University of Michigan, I have been especially grateful to Larry Hirshfeld, Marilyn Shatz, Twila Tardif, and Henry Wellman for their wisdom and insights. I am also indebted to all my collaborators on the research reported here, especially Woo-kyoung Ahn, Paul Bloom, Eve Clark, John Coley, Gil Diesendruck, Karen Ebeling, Jonathan Flukes, Gail Gottfried, Grant Gutheil, Erin Hartman, Gail Heyman, Anne Hickling, Michelle Hollander, Jen Jipson, Chuck Kalish, Melissa Koenig, Kathy Kremer, David Lizotte, Ellen Markman, Doug Medin, Melanie Nyhof, John Opfer, Tina Pappas, Lakshmi Raman, Karl Rosengren, Jon Star, and Marianne Taylor. Throughout the years I have been guided and encouraged by members of the University of Michigan Language Lab, who have given me outstanding critical advice on this research. Tom Rodriguez gave expert help in setting up the CHILDES database for the generics study reported in chapter 8.

For reading and commenting on an earlier draft of the ms., my thanks go to Paul Bloom, Gil Diesendruck, Andrew Gelman, Alison Gopnik, Gail Heyman, Larry Hirschfeld, Bruce Mannheim, Ellen Markman, Marjorie Taylor, and Henry Wellman. I am particularly grateful to Paul Bloom, whose enormously detailed and

trenchant reviews made me rethink major portions of the book. All problems, gaps, and inconsistencies that remain are my own. Rita Astuti, Gillian Feeley-Harnik, Alan Gelman, Andrew Gelman, Toby Jayaratne, Chuck Kalish, Greg Murphy, Lou Moses, and Harriet Whitehead provided useful help in answering questions. I thank Paul Bloom for suggesting the main title of the book, and Nancy Gelman, Larry Hirschfeld, Laura Novick, Barbara Sarnecka, and Marjorie Taylor for helpful feedback on cover ideas.

My research has been sponsored by the National Institute of Child Health and Human Development, the National Science Foundation, the Spencer Foundation, and the Guggenheim Foundation. I gratefully acknowledge their support. In writing this book, I have drawn from other published work, including Ahn, Kalish, Gelman, Medin, Luhmann, Atran, Coley, and Shafto (2001); S. Gelman (1996); S. Gelman (2000); S. Gelman (in press); S. Gelman and Coley (1991); S. Gelman, Coley, and Gottfried (1994); S. Gelman, Coley, Rosengren, Hartman, and Pappas (1998); S. Gelman and Diesendruck (1999); S. Gelman and Heyman (1999); S. Gelman and Hirschfeld (1999); S. Gelman, Hollander, Star, and Heyman (2000); S. Gelman and Kalish (1993); S. Gelman and Koenig (in press); S. Gelman and Opfer (2002); S. Gelman and Tardif (1998); S. Gelman and Taylor (2000); and Morris, Taplin, and Gelman (2000).

This book could not have been written without the generous assistance of the children, parents, and teachers who have participated in my research studies over the years. I am particularly indebted to the University of Michigan Children's Center and the University of Michigan Center for Working Families for their enthusiastic participation.

I have received much encouragement, help, and editorial advice from Catharine Carlin and John Rauschenberg of Oxford University Press. They have been both graceful and patient in working with me on this project.

And most of all, I am indebted to Bruce and our own essential children: Stephanie, Adam, and Ben. Adam (now age nine) asked me to add a final anecdote to this book, which as far as he can tell is about funny things that children say. Here it is: when Adam was about two and a half years old and taking a bath, I said, "I'm going to get the shampoo" (reaching for the bottle of shampoo, which had a cap in the form of Winnie-the-Pooh's head). He replied, without missing a beat, "I want sham-Piglet" (pointing at the bottle of bath bubbles, which had a cap in the form of Piglet's head). This example says nothing about essentialism, but I include it nonetheless as a reminder that young children are ingenious and charming as they go about the serious business of figuring out their world.

Contents

Chapter 1	
Introduction	3

PART I. THE PHENOMENA

Introduction to Part I	
Notes on Research Methods	21
Chapter 2	
The Inductive Potential of Categories	26
Chapter 3	
Hidden, Nonobvious Properties	60
Chapter 4	
Children's Conceptions of Nature and Nurture	89
Chapter 5	
Causal Explanations, Causal Determinism	108
Chapter 6	
Conclusions to Part I	137

PART II. MECHANISMS OF ACQUISITION

Chapter 7	
What Parents Say—and Do Not Say—about Essences	155
Chapter 8	
Essentialism in Language	179
Chapter 9	
Theory Theories and DAM Theories	239

PART III. IMPLICATIONS AND SPECULATIONS

Chapter 10

Unfinished Business 277

Chapter 11

Why Do We Essentialize? 296

Notes 327

References 335

Author Index 369

Subject Index 377

The Essential Child

This page intentionally left blank

Chapter 1

Introduction

[Essence is] the very being of anything, whereby it is what it is. And thus the real internal, but generally . . . unknown constitution of things, whereon their discoverable qualities depend, may be called their essence.

John Locke, *An Essay Concerning Human Understanding*

I begin this book with a confession: I was a child essentialist.

When I was no more than four or five years old, I asked my mother how boys and girls were different. I already knew the obvious answers—different bodies, different clothes, different hairstyles, different roles (this was, after all, the prefeminist, early 1960s). What I was searching for was some explanation of all of these apparent differences. I am not sure exactly what answer I expected to hear, but the answer I did get from my very practical mother (“boys have penises, girls don’t”) struck me as entirely unsatisfying. “Is that all?” I remember asking with disbelief. It seemed to me that there had to be a more profound basis for sex differences than that.

My early intuitions about intelligence were similar. I was in elementary school, maybe third grade. My classmates and I were waiting to take a standardized IQ test. As we sat, with our number 2 pencils sharpened and ready, I was awed by the power of this test. As I understood it, the IQ score would reveal my intellectual capacity—immutable, fixed, and unchanging. It was hidden and nonobvious, accessible to our teachers, but not something I would ever be allowed to know. The number would not change—it would stay in our permanent records, but more important, it would tell us who we were, and who we could become.

Both of these examples illustrate essentialist assumptions about categories (gender and intelligence, respectively). The present book concerns essentialism in everyday thought. Roughly, essentialism is the view that categories have an underlying reality or true nature that one cannot observe directly but that gives an object its identity. In other words, according to essentialism, categories (such as “boy,” “girl,” or “intelligence”) are real, in several senses: they are discovered (rather than invented), they are natural (rather than artificial), they predict other properties, and they point to natural discontinuities in the world. Essentialist accounts have been around, in one form or another, for thousands of years, extending back at least to Plato’s cave allegory in *The Republic*. Numerous fields—biology,

philosophy, linguistics, literary criticism, and psychology—stake claims about essentialism.

The question of whether children are essentialists is not to be taken lightly, because it runs directly against a powerful portrait of children's concepts as perceptually driven, concrete, and atheoretical. I will be making an argument that children's concepts are not merely perceptually based, concrete, or built up from specifics, but rather reflect folk theories and a powerful capacity to look beyond the obvious. To many readers, this viewpoint will sound familiar, as the zeitgeist in cognitive development for the past twenty-five years or so has been to acknowledge that children exhibit greater competence and conceptual sophistication than has been attributed to them in the past. As any student of developmental psychology knows, Jean Piaget's vastly influential stage theory posited qualitative shifts in cognitive capacity over the course of development (Piaget, 1970). On his view, children in the preoperational stage (from two years of age to about age six or seven years) are characterized more by what they lack intellectually than by what they have achieved. In contrast, critics of Piaget have been arguing, at least since the 1970s, that children have greater cognitive competence and potential than Piaget allowed (R. Gelman, 1978). This "early competence" view has received wide support since the 1970s and is amply supported by research topics as varied as numerical reasoning, theory of mind, language learning, and physical reasoning (among others; see Wellman and Gelman, 1998, for review). Even the popular press has taken note of the amazing capacities of babies and young children.

Nonetheless, the issue is far from resolved, for two reasons. First, even the most ardent early-competence theorist cannot deny that children's categories look truly aberrant on many tasks. Two-year-olds overextend familiar words to unrelated objects (e.g., calling the moon "a ball"; Clark, 1973). On experimental tasks, three- to five-year-olds extend novel words to items matching in shape rather than taxonomic kind (e.g., extending "zav" from a birthday cake to a top hat, rather than to a pie; Imai, Gentner, and Uchida, 1994; see Figure 1.1). Children three years of age fail to incorporate function when learning words for which adults find function crucial (L. Smith, Jones, and Landau, 1996). Preschool and even early-elementary-school children define ordinary words, such as "island," in terms of characteristic features (a place that is sunny, with palm trees) rather than defining features (a land mass surrounded by water; Keil, 1989). In brief, children seem captivated by surface appearances during their first few years of life.

Children's well-documented focus on object appearances, even on categorization tasks, means that children provide a strong test of psychological essentialism. Children below age six or seven are most often characterized as not looking beyond the surface of things. To the extent that children's categories can be characterized as essentialist, this is newsworthy. At the same time, such a finding would demand an explanation. I cannot simply conclude that children are more sophisticated than previously thought; I will also need to reconcile children's apparent sophistication with their many errorful ways.

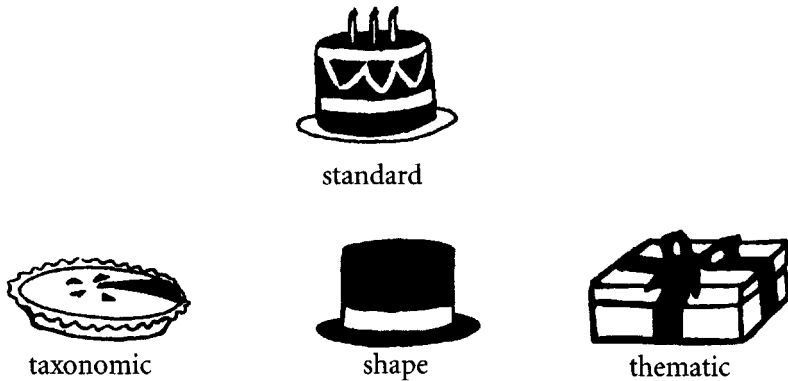


Figure 1.1. Sample item set. *Source:* Imai et al. (1994). Reprinted with the permission of Elsevier Science.

A second reason for reexamining an early-competence view is the recent resurgence of interest in explanatory accounts that attribute less cognitive sophistication to the child (e.g., Cohen and Oakes, 1993; Rakison, 2000; L. Smith, Jones, and Landau, 1996). Even when young children's behavior seems mature, their competence is argued to be illusory, as children are relying on simple strategies for processing information. In the words of one group of researchers, children make use of “dumb attentional mechanisms” to learn words (L. Smith, Jones, and Landau, 1996). It is no accident that this proposal is characterized by such a pejorative phrase; the aim is to convey unambiguously that children are truly *unsophisticated* word learners. Instead, it is suggested, children rely on automatic, associative learning that operates independently of reflective thought. In contrast, one aim of the present book is to show how and why “dumb” learning mechanisms mischaracterize young children. (See especially chapter 9.)

ESSENTIALIST CONTROVERSIES

In any discussion of essentialism, controversies rage. Is essentialism in the world or in the mind? Is essentialism innate in the infant (Atran, 1998) or “a late and sophisticated achievement” (Fodor, 1998, p. 159)? Is essentialism a universal “habit of the mind” (Atran, 1998, p. 551) or limited to certain points in history (Fodor, 1998)? Do we essentialize all concepts (Carey, 1996) or most readily just biological species (Atran, 1998)? Does essentialism reflect the logic of nouns (Benveniste, 1971; Carey, 1996; Mayr, 1991), or is it independent of language? Debates about essentialism sit astride debates about the very nature of human cognition.

It is startling, and more than a little daunting, to recognize that all these competing claims have been argued in the literature—and argued with passion. As I step into these minefields, I will lay claim to a position that views essentialism from

both a psychological perspective and a developmental (child) perspective. My colleagues and I, along with a growing number of other researchers in the fields of developmental and cognitive psychology, have examined essentialism as an empirical phenomenon, amenable to scientific study. Methods are diverse and converging: experimental studies of categorization, inductive inferences, and semantic interpretations; thought experiments with children and adults; and natural observation and analysis of ordinary language. The research participants are primarily middle-class, U.S., English-speaking children and adults, but we are beginning to see some much-needed cross-cultural and cross-linguistic investigations, too. This book is a progress report on the state of what we know about essentialism, in light of these results. It does not claim to be—and cannot be—the final word. For example, I cannot systematically address the metaphysical question of whether essentialism exists in the world, though I will briefly consider a few cases in order to see how essentialism misleads and assists us. But there are, I suggest, interesting and surprising findings to be discussed at this point.

I contend that essentialism is a pervasive, persistent reasoning bias that affects human categorization in profound ways. It is deeply ingrained in our conceptual systems, emerging at a very young age across highly varied cultural contexts. Our essentializing bias is not directly taught, nor does it simply reduce to a direct reading of cues that are “out there” in the world. Most decidedly, it is neither a late achievement nor a sophisticated one. The question of which categories we essentialize is a tricky one. In a nutshell, I argue that essentialism is the result of several converging psychological capacities, each of which is domain-general yet invoked differently in different domains. Collectively, when these capacities come together to form essentialism, they apply most powerfully to natural kinds¹ (including animal and plant species, and natural substances such as water or gold) and social kinds (including race and gender), but not artifacts made by people (such as tables and socks). This proposal rejects two alternative positions: the suggestion that we essentialize all concepts, and the suggestion that we essentialize only biological species. A final issue concerns how essentialized concepts are related to language. I suggest that essentialism does not require language, but language is one important cue children use when trying to figure out when and what to essentialize.

These proposals may seem simple enough, yet they contradict several fundamental assumptions about concepts, children, and language: that concepts are all structured alike, that children are limited to considering superficial perceptual features of the world, and that words simply reflect preexisting concepts. My primary purpose in writing this book is to trace the developmental roots of essentialism. In the course of doing so, I will also explore the broader lessons that these results imply concerning human concepts, children’s thinking, and ways in which language influences thought.

In this chapter, I first go through some preliminaries: what is essentialism, and why is it important? I then lay preliminary groundwork for the three themes that weave through the book:

- As proposed earlier, essentialism is an early cognitive bias. It is not, as some have maintained, an historical accident. Essentialism has its source in the cognitive requirements of categorization in certain domains—particularly as they affect the young learner.
- Children’s concepts are embedded in folk theories. They are not learned by means of simple associative learning strategies. They do not proceed from perceptual to conceptual, but incorporate both levels at once. More generally, developmental dichotomies (e.g., concrete to abstract, simple to complex, surface to deep, etc.) mischaracterize the nature of cognitive development.
- Although essentialism is foremost a cognitive bias, it is also supported and shaped by language. In particular, I will argue that two forms of language (count nouns and generic noun phrases) promote essentialist reasoning.

At the end of the chapter, I sketch out the structure and organization of the book.

PRELIMINARIES REGARDING ESSENTIALISM

What Is Essentialism?

Defining “essence” or “essentialism” is not an easy task, as the terms are used broadly by different scholars to mean different things. Nonetheless, I start with a brief intuitive characterization of what I take essentialism to mean. To begin, I am concerned with people’s beliefs about the world (not metaphysical claims about the world *per se*). There are three components to psychological essentialism as an intuitive folk belief:

- First, people believe that certain categories are natural kinds: they are real (not fabricated by humans), discovered (not invented), and rooted in nature. (For this and each of the other components, the claims about what people “believe” refer to nonconscious, intuitive beliefs or assumptions, not metacognitive or explicit beliefs.)
- Second, people believe that there is some unobservable property (be it a part, substance, or ineffable quality)—the essence—that causes things to be the way they are. The essence gives rise to the observable similarities shared by members of a category.
- Third, people believe that everyday words reflect this real-world structure. Words such as *dog*, *tree*, *gold*, or *schizophrenic* are often believed to map directly onto the natural kinds of the world. Not all words do so, but at least words referring to basic-level categories of natural kinds, as well as many words for social categories.²

To summarize, roughly speaking, what I mean by an “essence” is an underlying reality or true nature, shared by members of a category, that one cannot observe directly but that gives an object its identity and is responsible for other similarities that category members share (James, 1890/1983; Locke, 1671/1959; Medin, 1989). This is not a metaphysical claim about the structure of the world but rather a psychological claim about people’s implicit assumptions. In the domain of biology, an essence would be whatever quality is thought to remain unchanging as an organism grows, reproduces, and undergoes morphological transformations (baby to adult human; caterpillar to butterfly). In the domain of chemistry, an essence would be whatever quality is thought to remain unchanging as a substance changes shape, size, or state (e.g., from solid to liquid to gas).

However, what I mean by essence is not what all scholars mean by essence. Even among those who would agree with my characterization, there are subtle distinctions that must be made explicit. Lawrence Hirschfeld and I proposed that three factors jointly serve to map out the various types of essentialism (S. Gelman and Hirschfeld, 1999).

1. Where is essence located? Is it in the world (*metaphysical essentialism*) or in human representations (*representational essentialism*)?
2. What is the ontological type of an essence? Is it *sortal* (serving to define categories), *causal* (having consequences for category structure), or *ideal* (having no real-world instantiation)?
3. What degree of specificity is entailed? Are essences *specific* (their particulars known and identified) or *placeholder* (their particulars unknown and perhaps unknowable)?

I discuss these factors below, then clarify which sense of “essence” is assumed in this book.

METAPHYSICAL VERSUS REPRESENTATIONAL (PSYCHOLOGICAL, NOMINAL, CULTURAL). I distinguish between essentialism as a philosophical position and essentialism as a folk belief. The former addresses the nature of objective reality and is concerned with whether or not essences are located in the world (a metaphysical question); the latter addresses the nature of people’s representations of the world and so largely sidesteps the metaphysical question. Representational essentialism could be manifested in ordinary belief systems (psychological essentialism), language (nominal essentialism), and/or cultural practices (cultural essentialism).

SORTAL, CAUSAL, AND IDEAL ESSENCES. The sortal essence is the set of defining characteristics that all and only members of a category share. This notion of essence is captured in Aristotle’s (1924) distinction between essential and accidental properties (see also Keil’s (1989) defining versus characteristic properties): the essential properties constitute the essence. For example, on this view the essence of

a grandmother would be the property of being the mother of a person's parent (rather than the accidental or characteristic properties of wearing glasses and having gray hair; see Landau, 1982). In effect, this characterization is a restatement of the classical view of concepts: meaning (or identity) is supplied by a set of necessary and sufficient features that determine whether an entity does or does not belong in a category (E. Smith and Medin, 1981). Specific essentialist accounts, then, provide arguments concerning which sorts of features are essential. The viability of this account has been called into question by more-recent models of concepts that stress the importance of probabilistic features, exemplars, and theories in concepts (e.g., Murphy and Medin, 1985; Nosofsky, Kruschke, and McKinley, 1992; Rosch and Mervis, 1975).

In contrast, the causal essence is the substance, power, quality, process, relationship, or entity that causes other category-typical properties to emerge and be sustained, and that confers identity. The quote from Locke that began the chapter depicts this view. The causal essence is used to explain the observable properties of category members. Whereas the sortal essence could apply to any entity (pencils, wastebaskets, and tigers are all categories for which certain properties may be "essential," i.e., crucial for determining category membership), the causal essence applied only to entities for which hidden inherent properties determine observable qualities. For example, the causal essence of water may be something like H_2O , which is responsible for various observable properties that water has (but see Malt, 1994). The cluster of properties "odorless, tasteless, and colorless" is not a causal essence of water, despite being true of all members of the category "water," because the properties lack causal force.

The ideal essence is assumed to have no actual instantiation in the world. For example, on this view the essence of "goodness" is some pure, abstract quality that is imperfectly realized in real-world instances of people performing good deeds. None of these good deeds perfectly embodies "the good," but each reflects some aspect of it. Plato's cave allegory, in which what we see of the world are mere shadows of what is real and true, exemplifies this view. The ideal essence contrasts with both the sortal and the causal essences, which concern qualities of real-world entities.

SPECIFIC VERSUS PLACEHOLDER NOTION. Specific essentialist construals can be found in concepts as divergent as "soul" and "DNA," though essentialism may also be sketchy and implicit—a belief that a category has a core, without knowing what that core is. (See also R. Gelman, 1990; Keil, 1989; Wellman, 1990; and Wellman and Gelman, 1988, for arguments concerning skeletal or framework concepts.) Medin proposes that people have an "essence placeholder" (Medin, 1989). For example, a child might believe, before ever learning about chromosomes or human physiology, that girls have some inner, nonobvious quality that distinguishes them from boys and that generates the many observable differences in appearance and behavior between boys and girls. Those who are scientifically informed may come to have quite detailed beliefs about an essence (e.g., for gold,

that it has the atomic number of 79), but such conceptions are rare in everyday thought. Instead, the placeholder claim is that people hold an intuitive belief that an essence exists, even if its details have not yet been revealed. One consequence of this point is that an essence typically could not be part of the semantic core of a word, nor could it determine word extensions. Nonetheless, it has implications for people's beliefs regarding the depth and stability of a concept (Rothbart and Taylor, 1990).

SO WHAT DO I MEAN BY ESSENTIALISM? Having made all these distinctions, what do I mean by essentialism? There are at least twelve different senses of "essence" (see Table 1.1), some of which have potentially either a specific or a placeholder version. (Placeholder notions make sense only for representational essentialism. It is not clear what a placeholder version of metaphysical essentialism could be.) I will be focusing on causal, representational, placeholder essentialism (marked in the table with asterisks). There are other distinctions one could make as well (e.g., is essentialism of a kind or of an individual? domain-specific or domain-general?), but I take these up in the course of the book.

Metaphysical essentialism is beyond the scope of this book, primarily because psychological methods cannot shed light on these issues. The empirical studies presented in the following chapters focus on beliefs about turtles, for example, not on turtles themselves. Thus, the claim that children are essentialist is not a claim that essentialism is accurate. Biologists and philosophers of science have seriously questioned whether essentialism can characterize biological kinds (Dupré, 1993; Mayr, 1982, 1991; Sober, 1994; see chapter 11 for more detailed discussion). When essentialism extends beyond the realm of the sciences to attach to social categories such as race and gender, there is little doubt that it woefully misrepresents reality (Templeton, 1998).

I also decline to consider sortal essences, primarily because they seem implausible from both a psychological and a linguistic perspective. Given decades of research on categorization, it is extremely unlikely that people represent features that can identify all and only members of a category (Rosch and Mervis, 1975), regardless of how confident they are that such features exist (Malt, 1994; McNamara and Sternberg, 1983).

Table 1.1. Varieties of essentialism

	<i>Sortal</i>	<i>Causal</i>	<i>Ideal</i>
Metaphysical	•	•	•
Representational			
Psychological	•	*	•
Nominal	•	*	•
Cultural	•	*	•

Ideal essences have until recently been virtually ignored in studies of concept representation (but see Sperber, 1975). If anything, people's representations of most object concepts seem to be based on the structure and variability of what they encounter rather than nonrealized idealizations. When people are asked to rate the typicality of various instances of a category, for instance, their ratings usually reflect central tendencies (Rosch and Mervis, 1975). Interestingly, though, other kinds of categories do elicit ideals rather than central tendencies—for example, the prototype of a rich person is fabulously rich and not “average” rich (Barsalou, 1985), suggesting that it may be fruitful to examine ideal concepts in some content domains (see also Atran, 1999; Lynch, Coley, and Medin, 2000).

How does causal, representational essentialism manifest itself? For one, it is a doggedly realist view of the world, presupposing a reality beyond the phenomenal. A nonobvious essence is assumed to provide a “truer” representation of reality than can be observed, and the world is organized into densely complex and predictive clusters of correlated features. For example: when we classify an animal as a turtle, we are interested in much more than its outward appearance. We typically assume that this classification may have a nonobvious basis (e.g., though the presence of a shell or particular markings may be useful to classifying a turtle, these features can be overridden by other, more “biological” properties), that it fosters many novel inferences (e.g., regarding body temperature, number of offspring typically produced, and means of gathering food), and that it is open to revision. We presume there may be turtles that look like rocks (but are not), and rocks that look like turtles (but are not), or that one could discover new species of turtles that are unusually tiny or unusually large or that do not even have distinct shells. Table 1.2 summarizes some of these (overlapping) manifestations of essentialism.

How Essences Relate to Categories and Kinds

At this point, some readers may wonder if the claim of psychological essentialism is just a fancy way of saying that people form categories. The answer to this question is a resounding no. Causal essentialism is related to—but distinct from—categorization. A *category* is any grouping together of two or more discriminably different things (Bruner, Goodnow, and Austin, 1956). All organisms form categories: even mealworms have category-based preferences, and higher-order animals such as pigeons or octopi can display quite sophisticated categorical judgments (e.g., Herrnstein and de Villiers, 1980). But there are differences in the scope and variety of categories and category systems employed by different species (Markman and Callanan, 1983). I would certainly not wish to attribute essentialism to a mealworm, or even an octopus. More controversially, I would not attribute essentialism to a monkey or ape (see chapter 10).

Even considering only those categories used by humans, it is apparent that they are remarkably varied and diverse (Markman, 1989; Waxman, 1991). At one extreme, they include groupings that are encoded in language, that incorporate dense clusters of highly correlated features, and that display rich inductive potential

Table 1.2. Essentialist versus nonessentialist positions regarding the nature of categories

<i>Essentialist</i>	<i>Nonessentialist</i>
Discovered	Invented
Intrinsic to individual category members	Product of external forces
Unalterable	Easily changed
Stable over transformations	Transient
High inductive potential	Low inductive potential
Nonobvious basis	Superficial basis
Mutually exclusive traits	Overlapping traits
Absolute category membership	Graded category membership

(such as “cat”); at the other extreme, they include groupings that are arbitrarily constructed, with but a single featural basis, and with little inductive potential (such as “white things”; Markman, 1989; Mill, 1843). There are also many categories intermediate between the two, such as simple artifacts (e.g., “cup,” “chair”), which capture correlated feature clusters but have more-limited inductive potential than found in species of living things.

From these examples, it is clear that not all categories are essentialized. The richly structured types are often referred to as “natural kinds” (Lakoff, 1987; Schwartz, 1977, 1979), and these are the sorts of categories that I propose are most likely to be essentialized. Categories at the other extreme (such as “white things”) do not have essences except in the most trivial sortal sense. Markman (1989) refers to this latter type of category as “arbitrary categories”; Shipley (1993) refers to them as “classes.” A *natural kind* is a category that is treated by those who use it as being based in nature, discovered rather than invented, and capturing many deep regularities. In contrast, a category such as “white things” is treated as arbitrary, invented rather than discovered, and capturing little information beyond the basis of the original grouping. “Tigers” is a natural kind; the set of “striped things” (including tigers, striped shirts, and barbershop poles) is not, because it captures only a single, superficial property (stripedness); it does not capture nonobvious similarities, nor does it serve as a basis of induction. Similarly, ad hoc categories, such as “things to take on a camping trip,” do not form natural kinds (Barsalou, 1991). Beyond these obvious examples, which categories are essentialized kinds (e.g., artifact categories? social categories?) is an open empirical question that I consider at various points throughout the book. I will be arguing that essentialism in the sense I mean is found in people’s categories of natural kinds (both living and nonliving) and many social kinds (including races, ethnicities, and traits), but not artifact categories.³

There is a second sense, too, in which essentialism requires a notion of kind. Specifically, the essence of a category is attached not to an *individual object* but

rather to the *kind* in which it is classified. We cannot meaningfully interpret the question “What is the essence of *that*?” (with the speaker pointing to a dog) without knowing more about what “that” is. The question is ambiguous: does “that” refer to Fido? poodles? dogs? animals? white things? The essence of dogs presumably differs from the essence of animals or the essence of white things. Even when one contrasts two distinct kind categories (such as dogs and poodles), the hypothesized essence presumably varies. In order to determine the relevant essence, we need to know which kind is under consideration. It is helpful here to review briefly Macnamara’s (1986) argument regarding the logic of sortals. (Unfortunately the terminology here is confusing. “Sortals” here are not to be confused with “sortal essences,” which are entirely separate.) Sortals are simply those categories to which common nouns refer (e.g., dog, cat, chair, pencil). Macnamara notes that sortals are required for individuating entities. For example, the question “How many?” makes no sense without supplying the sortal—how many *what* (e.g., dogs? legs? molecules?). Likewise, sortals are required for making judgments of identity. “Are these two things the same?” makes no sense without supplying the sortal—the same *what* (see also Carey and Xu, 1999, for discussion). I suggest that sortals are likewise required for determining essence. One cannot answer the question “What is the essence of this?” without supplying the sortal—of this “*what*.”

Why Is Childhood Essentialism Important to Study?

Essentialism in children is important to study for several reasons. First, it is remarkably pervasive. It is pervasive over time (discussed at least over the past two thousand years), across radically different philosophical traditions (e.g., embraced by both Plato and Locke), and perhaps across cultures (e.g., Atran, 1998; Diesendruck, 2001). It is important to lay bare this set of persistent assumptions and to examine its origins and implications (good and bad) for human reasoning.

Second, the framework has revealed previously unsuspected abilities in young children, thereby contradicting a widely accepted view that children’s concepts are limited to concrete, perceptual, and obvious qualities. As I will detail in the chapters that follow, children incorporate a variety of nonobvious features into their concepts, including internal parts, functions, causes, and ontological distinctions. By extension, this portrait suggests a shift in views of knowledge development—what is most basic, what is derived, and how knowledge develops. If unobservable constructs are present from the start, then observable surface features cannot be privileged, simpler, or more basic. These are the “good implications” of essentialism for human reasoning.

Third, essentialism seems to motivate and underlie stereotyping. These are the “bad implications” of essentialism for human reasoning. To put it bluntly: stereotyping borrows the language and conceptual framework of essentializing. Different groups of people are treated as distinct in deep, nonobvious ways, and social group differences are assumed to be innately determined and fixed. To the extent that

people buy into this way of thinking, they will have a basis for treating social group differences as central to an individual's identity, for drawing inferences about an individual based on the group to which the individual belongs, and for attributing different motivations and explanations to those from different social groups. The stereotyping individual treats social groups as natural kinds (Rothbart and Taylor, 1990).

Fourth, the study of essentialism calls into question several core assumptions that guide how cognitive scientists think about and study word meaning and concepts. These assumptions include a focus on known (especially perceptual) properties, the belief that a single model suffices for all concepts, the notion that categorization is a single, unitary process, and implicit segregation of categorization from other high-level cognitive processes. Questions of essentialism have also inspired researchers to use a broader range of tasks to study categorization: not just identification and naming, but also induction and causal reasoning. These new tasks enrich our understanding of category functioning over development and suggest ways in which a wide range of distinct phenomena result from essentialist presumptions.

Finally, studies of essentialism have educational and social implications. Some scholars suggest that essentialist assumptions impede attempts to teach evolutionary theory (Evans, 2000, 2001; Mayr, 1982). More generally, much of our knowledge of the world is arrived at by means of inferences rather than being directly taught. Thus, any full account of knowledge acquisition must consider the conditions that promote or discourage inferential reasoning in children. I will argue that an essentialist assumption about categories, and essentialist language about categories, strongly influence children's inferences.

BACKGROUND CONTEXT FOR THE THEMES OF THE BOOK

I turn now to sketch out a bit of the theoretical context for the three themes of the book: essentialism as an early, domain-specific cognitive bias; children's concepts as embedded in theories; and language as an influence on cognitive development. Each of these themes will receive more detailed discussion throughout the volume. My modest aim here is simply to point out some of the theoretical controversies that exist, as background for the more extensive treatment that follows.

Essentialism as an Early, Domain-Specific Conceptual Bias

Essentialism is pervasive, but why? I will be arguing that essentialism is a cognitive predisposition that emerges early in childhood, particularly for understanding the natural world. This position is at odds with two alternative accounts, which I call "historical accident" and "inherent consequence of naming." On the historical accident position, essentialism is the by-product of modern Western philosophy, cultural and political traditions, or technology (Fuss, 1989; Guillaumin, 1980). For ex-

ample, some have argued that we are essentialists at this point in history because we can view the scientific enterprise fairly close-up and know about unobservable entities such as DNA and molecules (Fodor, 1998). However, attributing essentialism to historical accident does not easily explain why preschool children essentialize.

The contrasting argument says that essentialism is an inherent consequence of naming (specifically, a consequence of count nouns). By giving distinct entities the same name, we imply there is some unchanging, underlying sameness that they share. On this view, essentialism is a logical consequence of language use (Carey, 1996; Hallett, 1991; Mayr, 1991). The problem with this position is the domain-specific application of essentialism. It cannot explain why we essentialize more in some domains than others.

My view is distinct from both the alternatives just sketched out. Unlike the historical accident account, essentialism is a universal habit of the mind. People are deep-down essentialists even without the benefit of science or Plato's *Republic*. However, unlike the inherent-consequence-of-naming account, essentialism applies more to some domains than others. Naming in and of itself does not lead children to essentialize.

My view is closest to an evolutionary adaptation position, which posits that humans evolved a universal essentializing tendency, because it is beneficial for interactions with the world (Atran, 1998; P. Bloom, 2000; Gil-White, 2001; Pinker, 1994, 2002). This position is appealing in accounting for the recurrence of essentialism across cultures, epochs, and developmental ages. However, I underscore three caveats.

1. I will argue that essentialism is not a *single* adaptation, but the result of several distinct cognitive biases that emerged for varying purposes. In other words, essentialism per se was not specifically selected, but components of essentialism were. I will have more to say about that in chapter 11.

2. Whereas evolutionary positions tend to emphasize adaptive benefits, I am equally struck by the costs—most notably with categories that are socially constructed, such as race. How do we understand the errors and perils of essentialism? Are they simply to be expected, because any adaptation is only approximate (i.e., useful but not wholly prescient)? Or do they suggest that what evolved was not essentialism per se, but rather other capacities that result in essentialism as a side effect? I will be arguing for the latter.

3. The view I lay out supplies an important role for language (see the following section). Language guides children to notice and essentialize some categories more than others. As such, essentialism is not a distinct, encapsulated module that gets triggered by perceptual inputs alone. Although an adaptationist position does not specifically exclude language from playing a role, neither does it easily explain how it would do so.

It is difficult to obtain evidence on these issues by studying adults, because with adults the effects of schooling and scientific training cannot easily be factored out.

Children, however, provide a more compelling test case, because they have scant knowledge of either Western philosophy or scientific theories. In this respect, the younger the child under investigation, the more powerful the test case.

The Role of Language in Essentialism

One reason for suspecting essentialism is that the meanings of certain words seem to depend on something other than known, superficial properties. Kripke (1971) and Putnam (1970, 1973) argue that meanings of proper nouns (Kripke) and natural kind terms (Putnam) are based not on a list of known properties, but rather on “deeper” properties—what we could call theory-laden properties—including those that might not even yet be known. For example, the name “Harry Potter” is not defined by a set of features such as “wears glasses, flies on a broom, and has two best friends named Ron and Hermione.” If Harry Potter had died at birth, he would have none of these features, so they cannot be determinative of being Harry Potter. The only feature that seems *necessarily* linked to the name “Harry Potter” is that he was born of certain parents. According to Kripke, proper names refer but do not describe. Any description associated with a name merely helps us pick out the referent; it does not define the referent (see also Schwartz, 1977, 1979).

Kripke (1971, 1972) and especially Putnam (1970, 1973) extended this analysis of proper names to natural kind terms. They argue that, although a set of known features may be used to identify members of a natural kind category, the features do not serve as necessary and sufficient criteria. For example, whales are shaped like fish and live and swim in water as fish do, but they are not fish. Likewise, to use Putnam’s example, most of us cannot distinguish an elm from a beech, yet nonetheless maintain that the words “elm” and “beech” differ in meaning. We seem to assume that elms and beeches are different kinds of things, that the differences are there in the world for us to discover, and that experts could tell us which is which (again signaling that the distinction is real). Putnam thus forcefully argues for a sociolinguistic division of labor, according to which the average speaker need not know how to recognize whether or not something is an “elm” (for example), but experts in the community have the ability to make such a determination. As Putnam famously insists, “‘meanings’ just ain’t in the *head*!” (1975, p. 227). Meanings may not be in the head, but the conceptual underpinnings to such a system imply a kind of essentialism (in the head).

This brief review illustrates that language works in accordance with certain essentialist assumptions. However, it still leaves open the question of whether language *per se* contributes to essentialist thinking. It could be, instead, that essentialist reasoning contributes to how words are used.

Extant theories regarding the role of language on thought are wildly polarized, ranging from the claim that language is the lens through which we organize reality, leading different languages to adopt different worldviews (Hill and Mannheim, 1992; Lucy, 1997; Whorf, 1956), to the claim that language has no substantive effect

whatsoever on human cognition (aside from very local influences, e.g., memories can be encoded in a verbal format) (Pinker, 1994). After years of rejection of language as an influence on thought, there is renewed interest, particularly from a developmental perspective. Important work on these issues is now emerging from scholars including Bowerman (1996, 2000); Danziger (2001); Gopnik and Choi (1990); Gopnik, Choi, and Baumberger (1996); Gumperz and Levinson (1996); Imai and Gentner (1997); Levinson (1996); Martinez and Shatz (1996); Naigles and Terrazas (1998); Sera, Rettinger, and Castillo-Pintado (1991); Shatz (1991); Tardif and Wellman (2000); and Waxman, Senghas, and Benveniste (1997). At the very least, the suggestion of linguistic influences on thought is respectable.

I will take a middle position. I do not think that language drives the basic phenomenon of essentializing. It does not, on my view, create the urge to essentialize. Nor do I think that different languages essentialize to radically different degrees (though there may be subtle differences; see Nisbett, Peng, et al., 2001). Nonetheless, the extent to which a *particular* category is essentialized (or even essentialized at all) is open to linguistic influence. In other words, language helps determine when essentialism is used. Specifically, I will propose that two linguistic forms (common nouns and generic noun phrases) subtly convey to children an essentialist perspective on categories. This position will be presented in chapter 8.

ORGANIZATION OF THIS BOOK

The remainder of this book is organized into three main parts. In part I, I review the empirical basis for arguing that essentialism is an early-emerging reasoning bias in children. Chapter 2 reports children's category-based inductive inferences. Chapter 3 reports a variety of studies reflecting the importance of nonobvious properties in children's categories. In chapter 4, I consider how children reason about nature/nurture conflicts, and particularly their intuitive belief that category members share an innate potential. Chapter 5 examines the role of causation in children's categories. Chapter 6 summarizes and integrates the findings from the previous four chapters, and considers (and refutes) some alternative interpretations of these results. Altogether, the five chapters of this section review a wealth of evidence that jointly undermines the view of children's concepts as perceptually driven, concrete, or atheoretical.

In part II, I turn to the question that naturally arises if one accepts the conclusions of part I: namely, how do essentialist beliefs emerge, and by what mechanisms are they acquired or transmitted? Chapter 7 reports a detailed study of maternal input to young children and argues that parents say very little that is explicit or concrete to endorse or promote essentialism. Nonetheless, I argue that essentialist assumptions are implicitly conveyed to young children by means of certain linguistic forms of maternal speech (namely, generic noun phrases). The forms of language that are and are not argued to promote essentialism are detailed in chapter 8, along with evidence from children's language interpretation. In chapter 9, I

directly take up the question of what sort of developmental account best explains the acquisition of essentialized kinds in young children. To what extent do children rely on simple associative learning strategies as opposed to theories?

In part III, I address more speculatively the implications of parts I and II. In chapter 10, I raise and discuss unanswered questions that arose throughout the book, including essentialism across cultures, developmental change, individual variation, and how to disconfirm essentialism experimentally. Chapter 11 tackles several issues related to why children essentialize, including the scope of children's essentialist reasoning and how essentialism relates to other phenomena involving potent, nonobvious properties (including contagion, contamination, authenticity, and fetishes). Altogether, childhood essentialism has important lessons for our understanding of concepts, cognitive development, and language and thought.

PART I

THE PHENOMENA

This page intentionally left blank

Introduction to Part I:

Notes on Research Methods

In order to make a convincing study of essentialism as a fundamental folk notion, it is obviously crucial to provide evidence regarding the beliefs of ordinary folk. (Here I distinguish between “ordinary folk” and such luminaries as Aristotle, Plato, and Locke.) The notions sketched in chapter 1 may seem at first arcane and counter to common sense. What is commonsensical about invisible qualities that one can never know completely? Here it is important not to confuse the direct observability of the central construct with its status as common sense. Religious concepts provide an apt analogue: God is a mysterious concept, yet one that is readily embraced in folk theories (Boyer, 1994a, 1994b).

Direct evidence of essentialism is difficult to obtain. As I mentioned earlier, essentialism does not entail that people know what the essence of a given category is. Instead, it can be placeholder notion (Medin and Ortony, 1989). People may implicitly assume, for example, that there is some quality that bears have in common that confers category identity and causes their identifiable surface features, and they may use this belief to guide inductive inferences and produce explanations—without being able to identify any feature or trait as the bear essence. This belief can be considered an unarticulated heuristic rather than a detailed, well-worked-out theory.

Furthermore, an essence would rarely be consulted to determine category membership, for the simple reason that people often do not know (or cannot readily access) the relevant information (S. Gelman and Medin, 1993). In such instances, people use other features instead. Gender provides a useful example: although we typically assess someone’s gender based on outward (clothed) appearance and voice, even young children acknowledge that genital information is more diagnostic (Bem, 1989), and in our technological society we even use chromosomal information in certain contexts (e.g., amniocentesis and Olympic Games com-

mittees). (My arguments here are backed up by the Kansas Supreme Court, which on March 15, 2002, ruled that chromosomal information determines not only biological gender ["a post-operative male-to-female transsexual is not a woman"], but also ordinary language use ["The words 'sex,' 'male' and 'female' in everyday understanding do not encompass transsexuals"]! See report in the *New York Times*, March 16, 2002.) This means that people's use of salient observable features cannot be taken as evidence against essentialism.

These initial points of clarification imply that essences are difficult (perhaps impossible) to access directly. Psychological essentialism entails that people believe in the existence of essences, not that people have detailed knowledge regarding the content of essences, nor that essences exist. Accordingly, some results that might at first appear to contradict essentialism (as when people classify instances based on nonessential features, or cannot specify an essence, or have concepts that conflict with scientific concepts, e.g., Dupré, 1993) are not evidence against psychological essentialism. They are valuable for examining what kinds of information are used in certain tasks, but they do not constitute tests of psychological essentialism as a folk theory of concepts.

As a result, we have relied on indirect, converging methods. These are sketched out in Figure 1.2. At times the evidence focuses on realist assumptions about categories, and other times on underlying essences. These shifting criteria may frustrate some readers; I will at least do my part to try to keep clear which sense of essentialism I am addressing. The evidence in chapters 2 through 6 comes from several related lines of research: inductive potential, incorporation of nonobvious properties, stability over transformations, sharp category boundaries, beliefs about the relative role of nature versus nurture, and incorporation of causal features into categories. Together, these studies show that young children's categories are richly structured and extend beyond surface features (chapters 2 and 3), incorporating nonvisible properties (chapter 3), beliefs about innateness (chapter 4), and causality (chapter 5). In chapter 10, I return to the question of to what extent these features jointly constitute an essentialist framework.

CATEGORY DOMAINS

I assume that categories differ substantially from one another in structure, that not all categories are essentialized, and that serious confusion will result from positing a theory of "concepts" without specifying the type of concept under consideration. To render this assumption intuitively plausible, contrast the animal kind "bird" with the artifact category "tchotchke" (Yiddish for knickknack or inexpensive trinket) (see also Markman, 1989, for an incisive discussion of Mill's example of the category of "white things"). Most laypeople reading these words would probably agree to most or all of the following essentialist propositions about birds (placed in quotes to indicate that these are folk beliefs rather than metaphysical certainties):

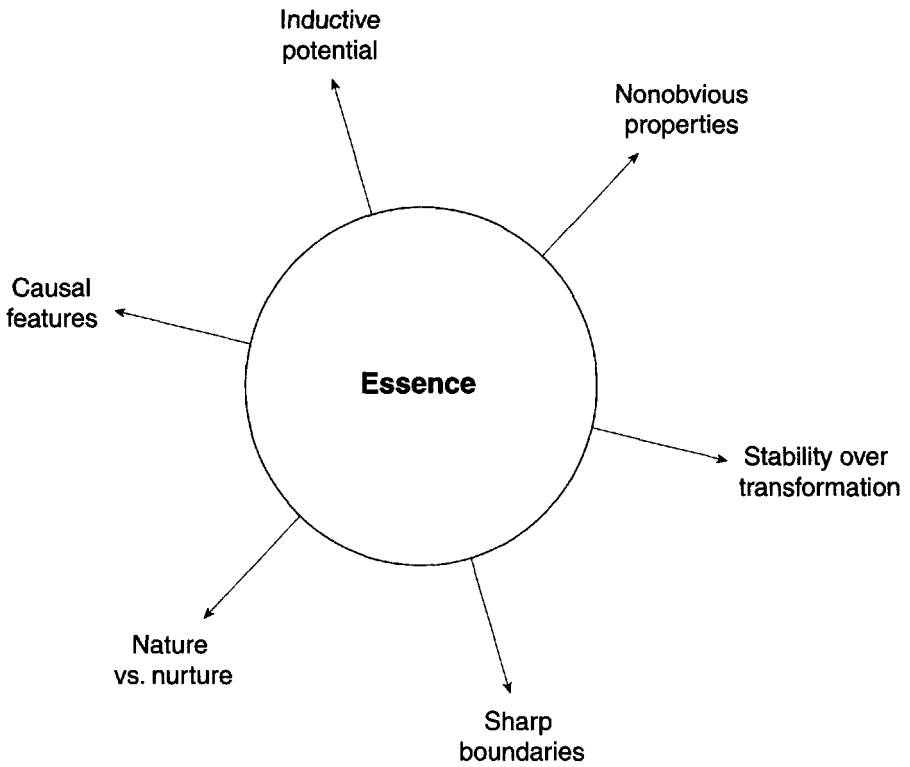


Figure 1.2. Implications of essentialism.

1. “The category of ‘birds’ is not an artificial grouping invented simply for the convenience of humans. Rather, birds belong together in some natural sense. We *discovered* the category of birds.”
2. “There are many nonobvious properties that birds have in common with one another, including properties that people haven’t yet discovered but will discover in the future”
3. “There is some underlying property (maybe genetic code? maybe evolutionary history?) that causes birds to be alike.”
4. “Many commonalities that birds share are biologically determined.”
5. “Throughout its existence, an entity that is once a bird is *always* a bird—it cannot be turned into some other kind of thing.”
6. “Something either is or is not a bird—it can’t be ‘kind of’ a bird or ‘partly’ a bird.”

Now try replacing “birds” with “tchotchkes” in each of these statements, and the differences between these two sorts of categories become quite clear.

It is interesting that much past work on categorization, continuing to the present, has focused on categories structured more like “tchotchkes” than like “birds” (Johnson-Laird and Wason, 1977). In contrast, most of the work I will review focuses on basic-level natural kind categories, as these are particularly likely to reveal essentialism (Markman, 1989; Keil, 1989). Artifact categories (including vehicles, furniture, and tools) and social categories (including gender, race, and psychological traits) are also occasionally included, to test the boundaries of essentialist thought. I will have more to say about the mapping of essentialism onto domains in chapters 6 and 11.

RESEARCH PARTICIPANTS AND DEVELOPMENTAL ISSUES

The research participants in these studies are, for the most part, preschool children of middle-class families in the United States. The choice of middle-class English-speaking children is one of convenience only, and in chapters 7, 8, and 11 I take up the complex and important issues of cultural variability. The choice of preschoolers, however, is quite deliberate, and I say a few words here to explain that choice.

As noted in chapter 1, preschool children seem to rely on different bases for classification than adults. Two- to five-years-olds make errors that an older child would find laughable. They tend to be captivated by salient appearances, as Piaget’s conservation errors demonstrate. So preschool children provide either an especially apt choice for studying essentialism (because a demonstration of an essentialist bias in children would provide particularly strong evidence), or a decidedly foolish choice (because we might not find essentialism in children so young). Quine (1977), for example, has suggested that children are at first limited to “intuitive kinds” that reflect our innate sense of similarity and only gradually move beyond them to form theory-based categories. Neisser likewise assumed that similarity-based categories and theory-based categories “correspond to points on a developmental continuum” (1987, p. 6).

The analysis one adopts here depends on how one accounts for children’s categorization errors. Let us consider three main candidate explanations (not that these exhaust the range of possibilities): lack of knowledge or expertise, tendency to focus on salient features of a task, and general limitations in logical capacity. I agree that children’s categorization errors are due at least in part to a relative lack of knowledge about most domains (Gobbo and Chi, 1986; Bedard and Chi, 1992; K. Johnson and Eilers, 1998; Keil, 1987), and to a general tendency to focus on the most salient aspects of a task or event (Inhelder and Piaget, 1964). As Chi has demonstrated so strikingly in the domains of chess and dinosaurs, even a preschooler can become expert in a domain and shift to using less obvious, more causally central features (Chi and Koeske, 1983; see also Chi, Hutchinson, and Robin, 1989). Where I disagree is with the further claim that preschool children lack the

logical structures that older children and adults use to form concepts (Inhelder and Piaget, 1964).

Young children pose particular challenges to designing a task to study essentialism. If we ask children to sort objects into categories of their own accord, chances are they will sort objects into categories that do not look anything like essentialized natural kinds. A hat and a birthday cake, for example, share little in the way of nonobvious commonalities. Likewise, on Inhelder and Piaget's free-form classification task (where a child is given an array of toys to group as they wish), children tend to be guided by the spatial configuration of objects in the array and "allow themselves to be guided by what they can perceive" (1964, p. 45) rather than making use of a coherent concept. For instance, they might group together a square and a triangle to make a house, rather than grouping a square with other squares, or a triangle with other triangles.

I agree that the focus on *what* is in a child's category yields a misleading portrait. What is critical instead is an examination of *how* the category functions. The studies that I report in chapters 2–6 typically supply children with a classification and then examine what sorts of inferences they make from the classification. These studies contrast with the standard sorting task, which requires children to form categories on their own. I will elaborate on this methodological issue at various points throughout the book.

In brief, by focusing on young children, I have placed my theoretical bets on a certain optimistic position regarding children's early concepts. I suggest that children are not limited to intuitive kinds, that they can look beyond similarity-based categories, and that they can form theory-based categories. I describe these capacities in the following chapters.

Chapter 2

The Inductive Potential of Categories

Classifications are theories about the basis of natural order, not dull catalogues compiled only to avoid chaos.

Stephen Jay Gould, *Wonderful Life*

Preschool children are not scientists, but they would (implicitly) agree with Gould's assessment of categories, at least in certain domains. With their classifications, children too are attempting to discover the "natural order" in the world. My goal in chapters 2 through 5 is to lay out the evidence for this claim. In this chapter I focus on children's capacity to use categories as the basis for novel inferences about the world. Children's category-based inferences are essentialist in two important respects: they involve reasoning about nonobvious properties (including internal parts, novel behaviors, and causal effects) and an appreciation that appearances can be deceiving when it comes to category membership.

To illustrate the centrality of categories in induction, consider the following real-world example. According to National Public Radio's *Morning Edition* news program (November 21, 2000), a large number of Magellan penguins were washing up on the beaches of Rio de Janeiro. Many of the people who found them took them home and put them in their freezers to keep them cool. However, this species of penguins lives only a few hundred miles south, in year-round above-freezing temperatures. As a result, when the Brazilians eventually asked for help from the staff at the local zoo, many of the penguins were on the verge of dying from hypothermia. (Thanks to Nicola Knight for providing this example.)

The Brazilian caregivers were basing their actions on their understanding of penguins as a category. Other than that these creatures were penguins, there was no reason to assume that they should live in a freezer. After all, other creatures that wash up on the beaches of Rio presumably prefer warmer climates. The Brazilians apparently assumed (falsely, as it turned out) that knowing that a bird is a penguin allows you to infer that its habitat and body temperature are equivalent to those of other penguins. They relied on what they already knew about a subset of the category to make inferences about novel category members. Unfortunately, the category of penguins does not cohere as tightly as the Brazilians' naive theories led them to believe.

If we try to model the implicit reasoning involved in this example, we find that three plausible strategies may be involved:

1. All penguins live in cold climates.
Beaky is a penguin.
Therefore, Beaky lives in cold climates.
2. Waddles is a penguin.
Waddles lives in cold climates.
Therefore, all penguins live in cold climates.
3. Waddles and Beaky are penguins.
Waddles lives in cold climates.
Therefore, Beaky lives in cold climates.

The first is a straightforward deductive inference. It is rare, however, that people explicitly learn category statements in the form of universal quantifiers (for example, explicitly learning that all penguins live in cold climates; see chapter 7), and so deductive inferences are not the usual means of learning about the world. The other two kinds of inference are both inductive inferences, because they extend beyond what is already known or what could be known with logical certainty (as opposed to deductive inferences).¹ The reasoning processes in examples 2 and 3 may differ from one another in interesting ways, but both entail reasoning from the known to the unknown, and both rely (in one case more explicitly than the other) on categorical reasoning. The studies I detail below have (implicitly) the logic of example 3 above.

THE BASIC FINDING: CATEGORIES PROMOTE INFERENCES CONCERNING NONOBVIOUS PROPERTIES

For a number of years, my collaborators and I have examined category-based induction, or the inferences people make from one category member to another, especially for hidden, unobservable properties. In this section I will describe in some detail our original basic finding. This work sets the stage for more-recent studies examining the scope and development of children's inferences.

Carey's Induction Task

Carey (1985) pioneered the use of inductive inferences as a tool for examining the nature of children's concepts. In an ingenious set of experiments, children learned new, unfamiliar facts about certain categories (e.g., that a dog has a spleen inside), then were given opportunities to report whether or not these facts generalized to other instances that were more or less similar to the target category (e.g., does a person have a spleen inside? does a flower have a spleen inside?). The focus of

Carey's work was not essentialism, but rather conceptual change in children's biological concepts. Nonetheless, two of her early findings nicely set the stage for the present discussion. Specifically, even the youngest children in her work (age four years) showed violations of similarity-based induction.

One finding was that children (especially four-year-olds, her youngest subjects) displayed a striking tendency to draw more inferences from properties taught about people than from properties taught about other animals. As a result, their patterns of inference conflicted with perceptual similarity. So, for example, four-year-olds more often projected properties from people to aardvarks (76%) than from dogs to aardvarks (29%), despite the greater similarity between dogs and aardvarks. Likewise, four-year-olds more often projected properties from people to stink-bugs (52%) than from bees to stinkbugs (12%), despite the greater similarity between bees and stinkbugs. The second finding was an asymmetry in inductive projections, with more inferences from people (e.g., people to dogs) than to people (e.g., dogs to people).

Carey provided a detailed theoretical explanation for these findings in terms of children's developing understanding of the animal domain. Most important for our purposes, preschool children's inductive inferences were constrained by their knowledge of categories, and did not follow a strict similarity-based pattern. The task revealed conceptual links that could not be accounted for strictly in terms of outward object appearance.

Gelman and Markman Triad Task

Ellen Markman and I conducted a series of studies using a task similar in basic structure to that used by Carey: in each of several picture sets, a child was taught a new fact and then tested on how she generalized the fact to other instances (S. Gelman and Markman, 1986). Because of our interest in the relative power of category membership in guiding children's inferences, each picture set constituted a triad in which the third picture closely resembled one of the first two pictures but was from the same category as the other (see Figure 2.1). In other words, each item pitted perceptual similarity and category membership against one another. Adult similarity ratings confirmed that the picture sets conformed to this principle. Of interest was whether children would draw inferences from one picture to another on the basis of outward appearance or natural kind category membership.

Figure 2.1 shows one of the sets we used. However, it is important to keep in mind: that participants viewed twenty different triads, including a wide range of animal and natural-substance categories (see Table 2.1); that the properties used also ranged broadly, including such things as internal parts ("has little eggs inside"), behavior ("eats grass"), physical transformations ("melts in an oven"), function ("helps make snow melt"), and origins ("comes from inside a sea animal"); and that the results I report held up generally across the various item sets.

In one set, the pictures were a colorful tropical fish, a gray dolphin, and a gray

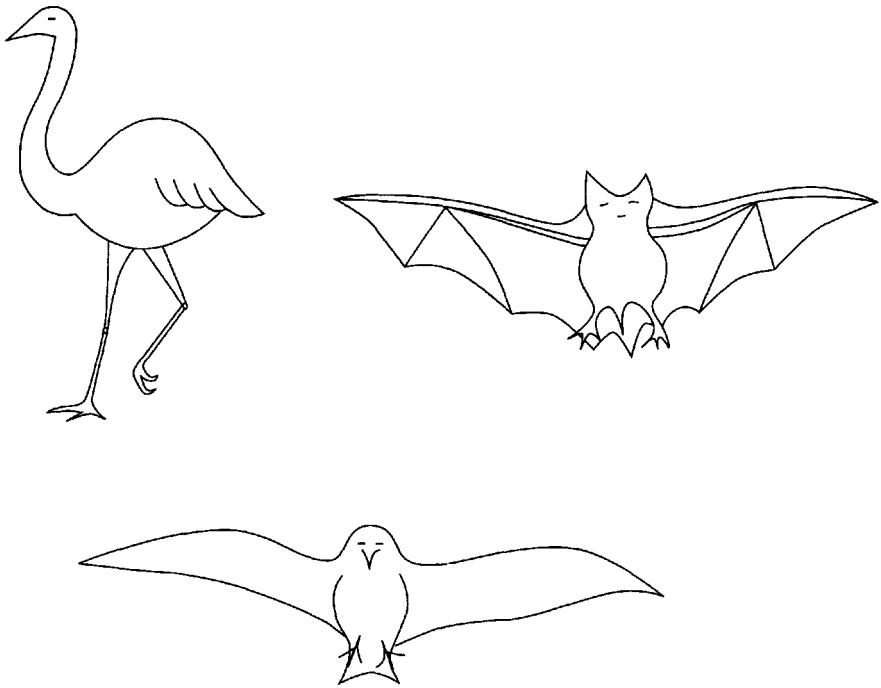


Figure 2.1. Sample item set, S. Gelman and Markman (1986). *Source:* S. Gelman and Markman (1986). Reprinted with the permission of Elsevier Science.

shark (fish). The shark was similar in appearance to the dolphin and dissimilar from the tropical fish, but shared category membership with the tropical fish, not the dolphin. The procedure began with the researcher naming the target pictures (“fish” and “dolphin”), after which children learned an unfamiliar nonobvious property about each. (Pretesting as well as a control condition confirmed that four-year-olds did not previously know which property applied to each animal, on any of the twenty item sets.) In this example, the children were told, “This fish [i.e., tropical fish] stays underwater to breathe” and “This dolphin pops above the water to breathe.” After being shown the picture of the shark and told that it was a fish, the participants were asked whether it stayed underwater to breathe like the fish or popped above the water to breathe like the dolphin.

To determine whether our analysis of category-based induction was sound for adults, we first gave the task to Stanford undergraduates. The task was a paper-and-pencil version of what children saw. The only substantive change was that we used less familiar properties such as “teeth have enamel” versus “teeth do not have enamel” for the tropical fish and dolphin, respectively. This change was needed in order to ensure that the adult participants had minimal preexisting beliefs or knowledge about the properties being tested (a point that was confirmed in a