Psychological essentialism is the idea that certain categories, such as 'lion' or 'female', have an underlying reality that cannot be observed directly. Where does this idea come from? This article reviews recent evidence suggesting that psychological essentialism is an early cognitive bias. Young children look beyond the obvious in many converging ways: when learning words, generalizing knowledge to new category members, reasoning about the insides of things, contemplating the role of nature versus nurture, and constructing causal explanations. These findings argue against the standard view of children as concrete thinkers, instead claiming that children have an early tendency to search for hidden, non-obvious features.

Why do preschool children often insist that mothers can’t be firefighters? Why do adults who were adopted in infancy sometimes search for their birth parents? Why do art collectors pay more money for an original painting than for an exact copy? These examples, although widely varying in content, can be understood within a framework of psychological essentialism.

Essentialism is the view that certain categories have an underlying reality or true nature that one cannot observe directly but that gives an object its identity, and is responsible for other similarities that category members share [1,2]. In the domain of biology, an essence would be whatever quality remains unchanging as an organism grows, reproduces, and undergoes morphological transformations (baby to man; caterpillar to butterfly). In the domain of chemistry, an essence would be whatever quality remains unchanging as a substance changes shape, size, or state (from solid to liquid to gas).

Essentialist accounts have been proposed and discussed for thousands of years, extending back at least to Plato’s cave allegory in The Republic. Numerous fields, including biology, philosophy, linguistics, literary criticism, and psychology, stake claims about essentialism. Here we are concerned with essentialism as a psychological claim (see Box 1). Although there are serious problems with essentialism as a metaphysical doctrine, recent psychological studies converge to suggest that essentialism is a reasoning heuristic that is readily available to both children and adults [3]. This article reviews such evidence, and discusses the implications for human concepts.

Evidence for psychological essentialism

Fodor argued that essentialism is the outgrowth of modern scientific and technological knowledge, with its corresponding focus on non-obvious constructs such as genes and chromosomes [4]. Children provide a critical test case of this idea precisely because they lack detailed scientific knowledge. If essentialism requires knowledge of modern science and technology then it should emerge relatively late in development. By contrast, if essentialism can be found in preschool children then this would imply that the doctrine is likely to have a more fundamental basis.

What would be evidence for essentialism, in children or adults? Medin and Ortony suggest that essentialism is a ‘placeholder’ notion: one can believe that a category possesses an essence without knowing what the essence is [2]. For example, a child might believe that there exist deep, non-visible differences between males and females, but have no idea just what those differences are. The essence placeholder would imply that categories permit rich inductive inferences, capture underlying structure (in the form of causal and other non-obvious properties), have innate potential, and have sharp and immutable boundaries. I have detailed at length elsewhere the evidence that preschool children expect certain categories to have all of these properties [3]. It is beyond the scope of this article to review the full range of evidence here. However, I briefly summarize below some of the major findings.

Inductive potential

Induction is the capacity to extend knowledge to novel instances, for example, inferring that a newly encountered mushroom is poisonous on the basis of past encounters with other poisonous mushrooms. This capacity is one of the most important functions of categories [5]. Categories serve not only to organize the knowledge we have already acquired but also to guide our expectations. Young children’s category-based inferences are consistent with essentialism in two respects. First, children readily infer properties that concern internal features and non-visible functions from one category member to another. Second, children draw inferences even when category membership competes with perceptual similarity.

Figure 1 provides an example from a set of studies conducted some years ago by Gelman and Markman [6,7]. The leaf (Figure 1a) and the leaf-insect (Figure 1c) have overall similarity: both are large and green, with striped markings, and share overall shape. However, if told the category membership of each of the three items – (a) ‘leaf’,
Box 1. Defining essentialism

The words ‘essence’ and ‘essentialism’ are used loosely by a range of scholars to mean a wide range of things. Three factors jointly serve to map out the various types of essentialism [43].

Where is the essence located?
Is it in the world (metaphysical essentialism) or in human representations (representational essentialism)? The former is a philosophical position concerning objective reality; the latter addresses how people construe reality (in their belief systems, language, and cultural practices). Essentialism faces difficulties as a characterization of the natural world [2]. Social categories, such as race and caste, have no true underlying essence [17,21]. Biological species also are without an essence because they evolve and are population-based rather than reflecting properties inherent in each individual [44,45]. Furthermore, whereas essentialism implies that there is a single appropriate classification for each organism, there might in fact be numerous valid classifications [46]. The essentialist view therefore seems to be a human construction rather than a perceived reality.

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)?

The sortal essence is the set of defining characteristics that all and only members of a category share. On this view the essence of a grandmother would be the property of being the mother of a person’s parent (rather than accidental properties such as having gray hair). The viability of this account has been called into question by models of concepts that stress the importance of probabilistic features, exemplars, and theories in concepts [38].

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. For example, the causal essence of tigers might be something like tiger DNA, which is responsible for various observable properties that tigers have. The cluster of properties ‘striped, ferocious, and large’ is not a causal essence of tigers, despite being true of all (or most) members of the category, 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 quality that is imperfectly realized in real-world instances of people performing good deeds. 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.

What degree of specificity is entailed?
Are essences specific (their particulars known and identified) or placeholder (their particulars unknown and perhaps unknowable)? Specific essentialism posits that a category essence is known and contributes to the meaning and use of a category label (e.g., H_2O for water), whereas placeholder essentialism suggests that a person believes that there is some causal essence that holds a category together, without knowing just what that essence is (e.g., that all samples of water share some inherent, non-obvious property). Although placeholder essentialism would typically be insufficient to determine word extensions, it has implications for people’s beliefs regarding the depth and stability of a concept.

In this article, the focus is on representational, causal, placeholder essentialism.

(b) ‘bug’, (c) ‘bug’ – and asked to draw novel inferences about the leaf-insect, children rely on the category membership conveyed by the label. Once children learn a new fact about one member of a category, they generalize the fact to other members of that category, even if the two category members look substantially different. This effect holds up for animals (bird, fish, rabbit), for natural substances (gold, cotton), and for social categories (boy, girl, smart, shy) [8,9]. It is found when using three-dimensional objects as well as drawings [10]. The effect is found with adults as well as children [6,11]. Even 1- and 2-year-old children draw category-based inferences about atypical instances [12,13]. Thus, the appreciation that words can signal non-obvious properties seems in place at the very start of word learning.

By 4 years of age, children display subtlety and flexibility regarding when they do and don’t make category-based inductive inferences [3]. They do not use a simple matching strategy, in which they extend properties only when two items share identical labels. The effect emerges not only with familiar labels, although not uniformly with all labels. The explanation that accounts most satisfactorily for the varied patterns of data is that children assess the extent to which entities are members of the same category (often conveyed via a label or phrase, although not necessarily), and independently assess the extent to which the property in question is a relatively enduring (versus temporary or accidental) feature (but see [14] for debate). Category-based induction results only when the entities belong to a category and the property is relatively enduring. When other causal mechanisms are known and available, however, category-based induction seems to be less frequent [5].

Innate potential

One of the most important kinds of evidence for essentialism is the belief that properties are fixed at birth, that is, that an organism displays innate potential. Details vary, but the basic paradigm to test this notion is as follows: children learn about a person or animal that has a set of biological parents, but is switched at birth to a new environment and a new set of parents. Children are then asked to decide whether the birth parents or the upbringing parents determine various properties.

For example, in one item set, children learned about a newborn kangaroo that went to live with goats, and were asked whether it would be good at hopping or good at climbing, and whether it would have a pouch or no pouch [15]. Preschool children typically reported that it would be good at hopping and have a pouch. Even if it cannot hop at birth (because it is too small and weak), and is raised by goats that cannot hop, and never sees another kangaroo, hopping is inherent to kangaroos. Therefore this property...
effects. Furthermore, features that are causally mere associations; causes are also more important than inherent cause. By 4 years of age, children appeal the animal itself, rather than to an external force share. Even 3-year-olds attribute an animal's actions to explanations for the properties that members of a category children provide consistent, domain-specific causal adults' concepts.

Underlying structure
Underlying structure can be seen in the importance children grant to causality, and to non-obvious features. By 2 years of age children view causes as vital to what something is. Features that are causes are more core than non-causal features. Causes are more important than mere associations; causes are also more important than effects. Furthermore, features that are causally coherent (meaning that they fit together in an explanatory framework) are weighted more heavily than features that are equally available but not participating in a network of causal explanations. Recent work with adults details more formally the importance of causality in adults' concepts.

Causality is further central to children's categories in that children provide consistent, domain-specific causal explanations for the properties that members of a category share. Even 3-year-olds attribute an animal's actions to the animal itself, rather than to an external force (inherent cause). By 4 years of age, children appeal to 'energy' as a causal force. By 8 years of age, children attribute actions and growth to an animal's 'insides'.

Children also privilege internal, non-obvious properties in their categories (see Figures 2 and 3). By 4 or 5 years of age, children often recognize that an animal cannot be transformed into another kind of thing (for example, a raccoon cannot become a skunk). Instead, category membership is stable over striking transformations. Also, non-obvious properties, especially internal properties, appear to be salient to young children and are privileged in their determinations of what things are. This finding appears not only among middle-class children in the US, but also in middle-class and Favela-dwelling children in Brazil.

Summary
The studies reviewed here suggest that both preschool children and adults from a variety of cultural contexts expect members of a category to be alike in non-obvious ways. They treat certain categories as having inductive potential, an innate basis, stable category membership, and sharp boundaries between contrasting categories. These beliefs are not the result of a detailed knowledge base, nor are they imparted directly by parents, although language might play an important indirect role (see Box 2). Instead, they appear early in childhood with relatively little direct prompting.

What psychological essentialism implies about children
Childhood essentialism poses a challenge to more traditional theories of children's concepts, which have emphasized their focus on superficial, accidental, or perceptual

Figure 2. Sample item used to study what procedures children deem relevant when determining ambiguous category membership (Lizotte and Gelman, reported in [3]). Five-year-olds and college students saw a series of trials depicting a pair of items that looked nearly identical. They were told that the two items differed in some respect (e.g. one was a dog and one was a wolf; or one was an animal and one was a toy), and that their job was to figure out which item was which. They were asked to decide (yes/no answers) whether checking any of the following would be informative procedures: insides, origins/parentage, behaviors, or age. The results indicated that children and adults believe that the items can be distinguished by a wide range of means: not just by external behaviors but also by internal properties and origins. Five-year-olds and adults reported that origins and insides are important clues as to which of two seemingly identical animals is a dog or a wolf, although it is unlikely that they could say specifically how the insides of dogs and wolves differ.

Figure 3. Sample items used in a study examining the role of internal parts in directing children's word learning ([31]. [The original items were in color.] a) Brazilian and US 4-year-old children learned a new word referring to a novel sub-type of a familiar category; for example, 'zava' for the snake in the bottom panel of (a), and were then asked to find the snake(s) and the zava(s) from among the three pictures in (b). Children who heard that the two animals in (a) shared internal similarities (e.g. have the same kind of bones, blood, muscles and brain) were more likely to learn the new word correctly (as a sub-type of snake) than children in a control condition who heard that the two animals had similar origins and insides. (a) and (b) were intended to be substitutable for one another (i.e. one was a dog and one was a wolf; or one was an animal and one was a toy).
Standard concepts have two distinct, although interrelated, levels: the level of observable reality and the level of explanation and cause.

Box 2. Language and essentialism

The language that is used to express membership in a category can influence children’s judgments about that category. Count nouns imply that a category is relatively more stable and consistent over time and contexts than adjectives or verbal phrases. For example, in one study [47], 5- and 7-year-old children first learned about a set of individuals with either a count noun (‘Rose is 8 years old. Rose eats a lot of carrots. She is a carrot-eater.’) or a verbal phrase (‘Rose is 8 years old. Rose eats a lot of carrots. She eats carrots whenever she can.’). They were then probed for how stable they thought this category membership would be across time and different environmental conditions (e.g. ‘Will Rose eat a lot of carrots when she is grown up?’ ‘Would Rose stop eating a lot of carrots if her family tried to stop her from eating carrots?’). Children who heard the count noun were more likely than children who heard the verbal phrase to judge that the personal characteristics would be stable over time and adverse environmental conditions. (For other examples of noun-labeling effects, see also [12,13,48–50].)

**Generic noun phrases** express essential qualities and imply that a category is coherent and permits category-wide inferences [51]. When 4-year-old children hear a new fact in generic form (e.g. ‘Bears have three layers of fur’), they treat this fact as typically true of most or all category members [52]. Generic nouns are plentiful in the speech that children hear [32,53], and children are highly sensitive to cues that mark generality [54–56].

Additionally, there are language-specific devices that convey essentialism. For example, young Spanish-speaking children make inferences about the stability of a category based on which form of the verb ‘to be’ is used to express it (ser versus estar [57]). Although it is unlikely that language is the source of psychological essentialism, it provides important cues to children regarding when to treat categories as stable and having an intrinsic basis [3].

What psychological essentialism implies about concepts

There is an idealized model of categorization that has formed the basis for much work in psychology. Standard theories of concepts have been based on considering which known properties are most privileged, and in what form. By contrast, essentialism tells us that known properties do not constitute the full meaning of concepts. Concepts are also open-ended. They are in part placeholders for unknown properties.

Historically, it has often been assumed that there is a single, unitary process of categorization [38]. Yet essentialism makes clear that categorization cannot be said to be a single thing. Categorization serves many different functions, and we recruit different sorts of information depending on the task at hand. Rapid identification calls for one kind of process; reasoning about genealogy calls for another. Task differences yield different categorization processes [39]. Even when the task is restricted to object identification, people make use of different sorts of information depending on the task instructions [40]. Two separate categorization procedures, rule application and judging similarity to an exemplar, can readily apply to the same categories, although they activate different neural regions in the adult brain [41]. Domain is also an important consideration. Although essence-like construals can be found in concepts of artwork and artifacts [42], essentialism is found most often with natural kinds and certain social categories. Categories such as penguin or apple tree imply inherent non-obvious properties, inductive potential, and innate commonalities not found with categories like window or crayon.

Because essentialism is found early in childhood, all of the points discussed above must be operating in quite basic and fundamental ways. In other words, we cannot simply assume that these are frills added on top of basic or standard categorization.

Conclusion

Recent work has shown that preschool children treat a variety of natural categories as having substantial inductive potential, innate properties, and underlying structure. Related to this, there are several important aspects to children’s concepts from very early in life, including placeholder notions, theory-based properties, category and task variability, and interdependence of categorization and other cognitive processes. More generally, these findings overturn assumptions about what is
simple or basic in human concepts. I have provided a framework of ‘psychological essentialism’ to account for these data (although see Box 3 for alternative accounts and debates). Ongoing research examines developmental antecedents in infancy, how best to understand the relation between perceptual and conceptual information in children’s categories, and modeling these representations in formal terms [24,26,27] (see Box 4).

Acknowledgements
Preparation of this article and some of the research described within were supported by NICHD grant HD36043 to the author. I am grateful to Bruce Mannheim, the Editor, and the anonymous reviewers for very helpful comments on an earlier draft of the manuscript.

References
1 Locke, J. (1671/1959) An Essay Concerning Human Understanding (Vol 2), Dover

Box 3. Alternative accounts
Although the data reviewed in this article are consistent with psychological essentialism in children, they fall short of providing direct evidence. Children do not explicitly articulate an essentialist perspective, and the word ‘essence’ probably does not enter the vocabulary until well into school-age years. Nor can we say with any precision what the essence of, say, a dog or a tree is to a child. Perhaps, then, essentialism is unnecessary, and we should instead refer to the component phenomena discussed earlier (e.g. categories have inductive potential; causal features are central to categories).

One reason to frame these results in terms of essentialism is parsimony: a range of phenomena that co-occur within a relatively brief developmental span seem to be instantiations of a single broader principle. This argument, however, requires a closer empirical foundation. More fine-grained analysis will be needed to discover how tightly linked these various phenomena are, both in adult cognition and in development (see [58] for an excellent start). A second consideration is the existence of explicit essence formulations among adults in varied cultures [3]. By treating people’s implicit construals as essentialist, a common framework covers both the explicit and the implicit phenomena.

Even if essentialist phenomena can reasonably be considered interrelated, this does not mean that the concept is indivisible. Indeed, psychological essentialism appears to have two related although separable assumptions: (a) a kind assumption, that people treat certain categories as richly structured ‘kinds’ with clusters of correlated properties; and (b) an essence assumption, that people believe a category has an underlying property (essence) that cannot be observed directly but that causes the observable qualities that category members share.

The major difference between kind and essence is that the latter incorporates the former and adds to it the idea that some part or quality (i.e. the essence) causes the properties shared by the kind. In the literature, the two notions have often been treated as interchangeable. However, categories can be bound together in crucial ways without considering the causal basis of the kind. In an important argument, Stevens [59] suggests that the data taken as evidence for psychological essentialism could instead be accounted for if people simply assume that there are causal laws connecting kind membership with observable properties. He terms such causal laws ‘K-laws’ (kind laws), and his alternative formulation the ‘Minimal Hypothesis’.

Stevens’s account, although eschewing essentialism, overlaps with the current model in emphasizing that people treat surface features as caused and constrained by deeper features of concepts.

Developmentally, a notion of kind might precede a notion of essence. Evidence for use of kinds is present by age 2 years (in children’s inductive inferences), but evidence for appeal to an essence (e.g. with the switched-at-birth method) has so far not appeared below age 4 years. An important direction for future research would be to use converging methods to try to distinguish ‘kind’ from ‘essence’ at different points in development. Nonetheless, something more than the Minimal Hypothesis might be required to account for essentialist-like behavior in children aged 4 years and over [60].

There are additional issues that are beyond the scope of this article but nonetheless important to mention. Some scholars have argued that essentialism cannot account for a variety of experimental findings with adults [61–64]. For example, the extent to which different liquids are judged to be water is independent of the extent to which they share the purported essence of water, H2O. Whether these findings undermine (or even conflict with) psychological essentialism is a matter of current debate [3,39].

Box 4. Questions for future research
• How coherent are essentialist beliefs? Do different strands (e.g. nativism, inductive potential, boundary intensification) all ‘hang together’, or do they develop piecemeal? Is there a single essentialist ‘stance’ [65], or is it the amalgamation of a variety of other tendencies [3]?
  • Why do children (and adults) essentialize? Some existing proposals include the following: an innate, domain-specific module for reasoning about biological species [23]; a domain-general consequence of sorts and count nouns [44]; the convergence of a variety of other early-emerging capacities needed for conceptual growth in childhood [3].
  • We know that non-verbal and preverbal organisms categorize [34], but do non-verbal and preverbal organisms also treat categories as having a non-obvious basis? If so, how could we tell? If not, why not?
• How does essentialism change with the acquisition of more detailed scientific knowledge? Does essentialism disappear with the acquisition of a scientific theory, or do the two frameworks co-exist in adult reasoning? Lay adults’ conceptual difficulties with evolutionary theory and genetics argue against true conceptual reorganization [66].
• Can certain inputs reduce children’s reliance on essentialism? If so, of what sort?
• Are there stable individual differences in essentialism, and if so, what is the source of such differences [67,68]? Are individual differences in essentialism related to individual differences in stereotyping? How does essentialism differ from other related notions, including stereotyping and entitativity [69]?
• Is essentialism found in the same way and to the same extent across cultures, or is there systematic variation [70]?
