Mind & Society, 3, 2001, Vol. 2, pp. 1-30

On the Functional Origins of Essentialism

H. Clark Barrett

Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195, Berlin, Germany

Current affiliation:
UCLA Department of Anthropology, 341 Haines Hall, Box 951553, 90095-1553 Los Angeles, CA, Tel. (w) 310 267 4260; (h) 310 652 2043; e-mail: barrett@ucla.edu
http://www.sscnet.ucla.edu/anthro/faculty/barrett

(Received July 2000, accepted June 2001)

Abstract *This essay examines the proposal that psychological essentialism results from a history of natural selection acting on human representation and inference systems. It has been argued that the features that distinguish essentialist representational systems are especially well suited for representing natural kinds. If the evolved function of essentialism is to exploit the rich inductive potential of such kinds, then it must be subserved by cognitive mechanisms that carry out at least three distinct functions: identifying these kinds in the environment, constructing essentialized representations of them, and constraining inductive inferences about kinds. Moreover, there are different kinds of kinds, ranging from nonliving substances to biological taxa to within-species kinds such as sex, and the causal processes that render these categories coherent for the purposes of inductive generalization vary. If the evolved function of essentialism is to support inductive generalization under ignorance of true causes, and if kinds of kinds vary in the implicit assumptions that support valid inductive inferences about them, then we expect different, functionally incompatible modes of essentialist thinking for different kinds. In particular, there should be differences in how biological and nonbiological substances, biological taxa, and biological and social role kinds are essentialized. The functional differences between these kinds of essentialism are discussed.*

Keywords *Essentialism; evolution; induction; concepts; cognitive development; folk biology; domain specificity; evolutionary psychology; ecological rationality*

Are people predisposed to interpret the world as if it contained natural kinds or classes of things, whose nature is determined by unobservable "essences" (Medin, 1989; Medin & Ortony, 1989; Gelman, Coley, & Gottfried, 1994)? And if so – given that there is widespread skepticism in the philosophical and scientific communities that anything resembling essences exist – why? In other words, is psychological essentialism merely an accident, or could it have a purpose, a function of its own?

A considerable amount of debate and disagreement persists in the literature on essentialism. Some doubt that it exists or that it is a useful theoretical construct (Braisby, Franks, & Hampton, 1996; Strevens, 2000), and others suggest that it is merely a byproduct of language or categorization in general (e.g., Carey, 1995). On

the other hand, many seem to agree that essentialism may be, in some sense, a "rational", or at least functionally effective, way of representing and thinking about certain natural kinds, even if true essences do not exist (Atran, 1998; Coley, Medin, and Atran, 1997; Gelman & Markman, 1986, 1987; Gelman & Wellman, 1991; Keil, 1994, 1995). In fact, it has been suggested that essentialism may have an evolved basis (Atran, 1994, 1998; Sperber, 1994). Expert opinion thus runs the gamut from denying that there is phenomenon to explain, to asserting its status as an evolved feature of human cognitive architecture.

Part of this confusion might reasonably be attributed to a failure to look claims of functionality straight in the eye. The proposal that any particular cognitive phenomenon has a function, if it is to have any value as a scientific hypothesis, carries implications about the design of the cognitive mechanisms that underlie it. In the case of psychological essentialism, it is not enough, ultimately, to claim that essentialism sometimes makes sense. It is easy enough to see the intuitive argument that representing certain kinds of things as sharing a common essence, under conditions of ignorance about the true causes that make those kinds of things what they are, might sometimes be a useful strategy. But to say that people have a propensity to do this, precisely because of its usefulness (rather than as a happy accident), entails making proposals about where such a propensity comes from – in particular, what its functions would have been in the environments in which it evolved – and then to generate hypotheses about how it manifests itself on a cognitive level. The alternative is to rely on intuition: e.g., that there are natural kinds in the world, that people intuitively know what they are, and that people learn that essentializing them is a useful strategy. While the latter approach might seem somehow more parsimonious, in the long run it sweeps ultimate causation under the rug of intuition, rendering hypotheses more difficult to generate and falsify – the question will always come down to "good" or "bad" intuition about what really are or are not, according to experts, natural kinds.

For example, people are said to be wrong when they essentialize races or ethnic groups, because it is generally agreed that these are not true natural kinds (Banton, 1987; Hirschfeld 1994, 1995; Rothbart & Taylor, 1990). However, just why they consistently adopt such a "wrong" strategy depends on both the evolved function of essentializing mechanisms and (consequently) on their input criteria. Some suggest that the essentialization of races results from overextension of a system originally evolved to handle biological taxa (Atran, 1990; Boyer, 1990; Rothbart & Taylor, 1990); others suggest that ethnicity is or has become, in fact, part of the evolved proper domain of an essentializing system originally evolved to handle biological species (Gil-White, 2001); and still others suggest that essentialism is not strictly proper to biological taxa at all (Gelman *et al.*, 1994; Hirschfeld, 1994, 1995). Ultimately, such questions cannot be resolved without theories of adaptive function that generate hypotheses about mechanism design specific enough to distinguish between the possible explanations. Unless it can do this, an evolutionary approach scarcely improves upon an intuitive one.

The purpose of this essay is not to examine the question of whether or not essentialism exists; that question has been investigated in detail elsewhere. Rather, it is to examine the proposal that essentialism has an evolved function, in order to see what

that proposal entails, not simply as an afterthought or as a post hoc explanation for essentialism, but as a source of hypotheses about cognitive design. The first section of the essay will review the most plausible evolutionary theory of essentialism, namely, that it evolved to support inductive inference under ignorance of the real causal processes that structure natural kinds. Section 1 will lay out the two main assumptions, proposed in the literature, which distinguish essentialist representational systems from others and that support valid inductive inferences for essentialized kinds. It will also examine the three functions that essentializing mechanisms must carry out in order to reap the inductive benefits of essentialism. The second section of the essay will ask how these insights about function may help to resolve debates about the proper domain of essentialism. The principle claim will be that in order to understand mechanisms specialized for inference about a particular kind of entity, be it a substance, an artifact, or an organism, one must first examine the causal processes that render a particular grouping of things coherent for the purposes of inductive inference. Because there are multiple processes that generate inductive "kinds", the underlying assumptions necessary to generate the sorts of inductive inferences that would have been valid and useful in human ancestral environments differ for different kinds. Moreover, the assumptions that are valid for one kind category (e.g., substances) may be invalid for others (e.g., whole organisms). Thus, there may be not one essentialism but many. Section 2 examines what the functional properties of these modes of essentialist thinking might be, and what kinds of kinds may fall within their proper domains.

1. What is essentialism for?

1.1. The minimal account of essentialism

Since the seminal philosophical work of Kripke (1972), Putnam (1975), Quine (1977) and others on natural kinds and their role in semantics and conceptual structure, there has been considerable interest in the phenomenon known as psychological essentialism. Essentialism is a stance one may take towards objects in the world, a way of representing or construing them (Gelman, Coley, & Gottfried, 1994; Gelman & Diesendruck, 1999; Gelman & Hirschfeld, 1999; Keil, 1989, 1994; Medin, 1989; Medin & Ortony, 1989). To be essentialist is to treat objects as if they "have essences or underlying natures that make them the thing that they are" (Medin, 1989), and to treat them as if they have properties that result from these essences. This latter point is important, because it distinguishes essentialized representations from mere tabulations of correlated features: being a member of a particular natural kind is construed as causally entailing having certain features, and not the other way around. Although there are undoubtedly many ways of taking such a stance towards objects – and many ways of instantiating cognitive systems that embody essentialist assumptions – it is useful to identify the minimal requirements of essentialism, i.e., what is minimally necessary in order to qualify as essentialist in some meaningful sense.

The core of a minimalist account of essentialism is the notion of what Medin and Ortony (1989) have called an "essence placeholder": that part of a natural kind concept to which various kind-specific features or properties are tied. On this account the essence is represented as causing kind-specific properties, which may in turn be diagnostic of the essence's presence, but are not necessary for the essence to exist. Medin and Ortony proposed that people might have very few explicit beliefs about what essences are and yet still act *as if* the properties of objects were the result of essences. Gelman and Diesendruck (1999, p. 88) explain this notion as follows:

> Essentialism does not entail that people know (consciously or unconsciously) what the essence is. Medin and Ortony (1989) referred to this unknown-yet-believed-in entity as an "essence placeholder". People may implicitly assume, for example, that there is some quality that bears share that confers category identity and causes identifiable surface features, and they may use this belief to guide inductive inferences and 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 theory.

We might contrast, then, an "essentialized" kind concept with a kind concept that merely lists features associated with a kind: whereas being a member of a kind defined in the latter way is a matter of possessing a sufficient number of necessary features, being a member of an essentialized kind simply requires having the essence (whatever that might be). Note that there is currently a debate over how "minimal" the representation of an essence needs to be (if indeed such representations exist), in order to account for the empirical data. Strevens (2000) has recently proposed that one need not postulate the existence of an essence placeholder to account for the data (especially the data on children's judgments). Instead, he makes the even more minimalist claim that category membership alone will do, along with the assumption that being a member of a category entails having certain properties.[1] On either account, what is important is that essentialism entails a distinct representational style, or format, that ties various essentialized features *causally* to a central representational node; otherwise, a kind concept is merely a list of necessary and / or sufficient features.

As Gelman and Diesendruck (1999) point out, people do not normally, if ever, rely on direct evidence of an "essence" itself – however they might represent it – in order to identify an entity as a member of a particular natural kind category. But the use of correlated features to diagnose kind membership does not imply that those

[1] From a functionalist perspective, it is not entirely clear how Strevens (2000) and Medin and Ortony (1989) differ in their account of essence placeholders, because if one considers whatever marks category membership to be the placeholder, the accounts seem quite similar. Note, however, that another potential point of difference is that Strevens argues that the data do not compel us to assume that children represent a single causal pathway as accounting for the properties characteristic of a given natural kind. Instead, there may be multiple possibilities for how kind-specific properties may be tied to kind membership.

features are construed as all there is to belonging to a kind. One can use properties such as shape, fur, growling, and so on, to identify something as a bear, without believing that these properties alone are what makes the bear a bear (people think there is a difference, for example, between something that merely appears to be a bear and something that really is). Here it is useful to note the distinction, made by Locke, between real and nominal essences: real essences determine the nature of things of a particular kind, while nominal essences are properties that can be used to distinguish or recognize members of a kind (Locke, 1689).

There has been some confusion, in the literature, over the term "essential properties". The central thesis of psychological essentialism as proposed by Medin and others (Medin, 1989; Medin and Ortony, 1989; Gelman *et al.*, 1994) is that some properties of things will be represented as resulting from internal essences. While some might call these "essential properties", it is important to note that people might accept that possibility that some properties, which would normally be expressed as a result of having an essence, might not always be expressed, even in cases where the essence is present – if so, possession of the properties themselves is not necessary for kind membership, and so calling them "essential" may be misleading. A better term for properties represented as resulting from an essence is "essentialized". For example, the ability to fly might be an essentialized property of birds. While people might feel that normally the ability to fly results from possession of the bird essence, they might still agree that while some birds are unable to fly (and thus lack what some would call an "essential property" of birds), they are still birds and possess the bird essence, but simply fail to express all of its usual manifestations[2]. It is thus important to distinguish between the essence itself (Locke's "real" essence) and the essentialized properties that are construed as arising from it (Locke's "nominal" essences).

1.2. What are the functional features that distinguish essentialism?

Psychological essentialism, as described so far, sounds rather mystical, and even irrational in a world where scientifically discovered principles are available for sorting things into real kinds and explaining their properties. Indeed, scientists – those most likely to have expert opinions about whether essences really exist or not – have been scolded for imagining, or acting as if, certain natural kinds have essences (see Dupré, 1993; Mayr, 1982). In general, there exists widespread skepticism that things strictly qualifying as real essences, in Locke's sense, exist (Keil, 1995). If this is

[2] Compare this point to the discussion in Strevens (2000, p. 152), where he discusses a passage from Medin and Ortony (1989, p. 184) in which flying is considered "part of the represented essence of bird". Here I suggest that flying is not likely to be represented as "part of" the essence, but as a *product* of it. Thus, flying should be referred to as an *essentialized* property rather than an *essential* property because, not being part of the essence itself, flying is not strictly necessary for an entity to be considered a bird. To collapse essentialized properties into the essence itself is to reintroduce necessary and sufficient features, and thus to confuse essentialized concepts with classical ones. Apparently, Medin and Ortony make this confusion themselves in considering flying to be part of the bird essence, though the confusion may simply result from imprecise use of language.

true, how could it ever be rational to be essentialist? Or, in evolutionary terms, if the assumptions behind essentialism are patently false, how could essentialist predispositions ever have been favored by natural selection?

One way of approaching this question is to disregard the semantics of the term "essence", to look strictly at the formal properties essences are supposed to have – i.e., at the formal nature of the assumptions underlying essentialism – and to ask whether there is a fit between these assumptions and entities in the world. If so, one could say that essentialism might be an ecologically rational strategy for inference and decision-making (Gigerenzer, 2000). When essentialist assumptions about a particular object lead to correct inferences because of a fit between these assumptions and the structure of the inductive domain, we can say that the object is validly essentialized, even in cases in which a real essence might not exist. This is the theory that has been pursued by those who have argued that essentialism is an evolved feature of human cognition (Atran, 1990, 1994, 1998; Coley, Medin, & Atran, 1997; Gil-White, 2001; Sperber, 1994). The adaptationist approach to essentialism focuses on two core assumptions, implicit in the design of essentialist representational systems, which render essentialism an ecologically rational mode of construal: the assumptions of *executive causation* and *rich inductive potential* (Atran, 1998; Gelman, 1988; Gelman *et al.*, 1994; Gelman & Coley, 1991; Gelman & Diesendruck, 1999; Keil, 1989, 1994; Medin, 1989; Medin & Ortony, 1989).

1.3. Executive causation

The first assumption – implicitly instantiated in a representational system that ties properties to essence placeholders – is that there exists an "executive" or central cause that is responsible for (some of) an object's properties, i.e., just those properties that are represented as kind-specific (Gelman *et al.*, 1994; Gelman & Diesendruck, 1999; Medin and Ortony, 1989)[3]. Although from this assumption is sometimes derived a corollary, namely, that a specific (internal) location of the causal agent is represented (Gelman & Wellman, 1991; Keil, 1989; Simons & Keil, 1995) this is not strictly mandated by the essence placeholder theory. It could be the case, for example, that the essence is construed as being distributed throughout the object in question, or that its spatial qualities are not represented at all. Furthermore, the functionalist theory of essentialism does not mandate any *explicit* representation of a causal link between the essence placeholder and essentialized properties. It does, however, require that the properties of an essentialized object not be represented as occurring together merely by coincidence. Thus, the executive causation assumption, implied by the placeholder representational format, differs from an assumption that natural kinds merely happen to have correlated sets of features, or that natural kinds are defined by the possession of certain features.

[3] The "essences" discussed in this paper are thus *causal* essences, as opposed to *sortal* or *ideal* essences, as distinguished by Gelman & Hirschfeld (1999).

1.4. Rich inductive potential

A second core assumption of essentialism is the assumption of what Gelman and Coley (1991) have called "rich inductive potential". This is an assumption that the properties caused by the essence are not exhausted by what is immediately observed or currently known; instead, it is assumed that the number of properties resulting from the essence is potentially inexhaustible, and that these properties pre-exist, independent of the observer, waiting to be discovered (Gelman, 1988; Gelman *et al.*, 1994; Gelman & Diesendruck, 1999; Keil, 1989, 1994). The assumption of rich inductive potential can be contrasted with an assumption of finite properties: an assumption that the properties of a particular kind or class of things may be confined to those things that are immediately apparent upon inspection, or to those properties that are used to diagnose kind membership itself. One way of restating this latter assumption, in Locke's terms, is that the real and nominal essence of the kind are equivalent, or very nearly so. As an example, Locke refers to triangles, whose nature as a class of things is exhausted by the very properties that define it (Locke, 1689, Book 3, Chapter 6). The assumption of rich inductive potential, by contrast, is that there is more to the nature of a kind than that which we use to recognize it; in other words, that the executive cause has a potentially limitless number of regular effects.

Essentialism, then, can be seen, functionally, as a way of representing kinds that makes two core assumptions, albeit implicitly, about the nature of the kinds being represented: that their properties result from a single, executive cause (implicitly, the "essence"), and that the number of these properties is large and potentially infinite. In this sense it can be distinguished from other representational formats, for example, classical formats that consist of lists or sets of features. The question then arises: what do such representational commitments offer, in functional terms? In other words, what would be the adaptive benefit of representing things in this way?

1.5. The adaptive problem: inductive inference under ignorance of true causes

The most promising proposal for the function of essentialism is that it serves to guide inductive inference (Atran, 1994, 1998; Coley, Medin, & Atran, 1997; Gelman & Markman, 1986, 1987; Gelman & Coley, 1991; Gelman & Wellman, 1991). This is related to a key property of essentialist representational systems, as defined above: they are not driven entirely by appearance and similarity (Keil, 1989, 1994; Keil *et al.*, 1998; Medin, 1989; Medin & Ortony, 1989). Essentialist representational systems allow that appearances and superficial properties are associated with kind membership, and even diagnostic of it, but not that they define it. A large body of empirical research has shown that children appear to follow these assumptions when reasoning about natural kinds (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Gelman & Wellman, 1991; Keil, 1989; Simons & Keil, 1995).

A crucial benefit of such a way of construing kinds is that it allows one to exploit the causal structure of the world (of natural kinds, in particular), without necessarily knowing anything about the causes themselves. In other words, it can be useful

to assume an executive cause for certain natural kinds, especially when one has no perceptual access to the executive causal agent itself, nor to the processes whereby it produces its multiple effects. Here, one can begin to see the outline of a role for natural selection acting on human representational and inference systems: because of the advantages they offer for inductive inference, mechanisms that cause people to construct essentialist representations of certain kinds, i.e., to "essentialize" them, may have an advantage over systems that merely compile lists of correlated features, and thus may be favored by selection. This advantage would have obtained to a greater degree in ancestral environments, before the advent of systems of formal science which attempt to rigorously determine what kinds of things are validly essentializable and why (Atran, 1990).

It is widely agreed that inductive inference presents thorny epistemological problems that cannot be solved without some pre-existing assumptions or constraints (Goodman, 1954; Quine, 1977; Markman, 1989). We are constantly faced with problems of induction, every time we form an expectation or make a prediction. In human ancestral environments, some kinds of inductive inference would have had potentially extreme fitness consequences, and selection could have acted on the design of mechanisms for making such inferences and for organizing the information on which the inferences would be based. For example, the decision to eat a particular kind of food – a mushroom, say – would require the application of acquired knowledge about mushrooms, and a judgment about the kind of mushroom about to be eaten and its likelihood of being poisonous. Judgments about kind-related properties would have been important upon encounter with potential predators, when making decisions while hunting prey, when using plants for medicinal purposes, and even when interacting with another person. What is the nature of the cognitive mechanisms that have evolved to guide our intuitions and help us solve problems of induction in such situations?

Theorists who adopt a functionalist approach to essentialism have identified several classes of problem that essentialist inference systems must solve. To reap the potential benefits of essentialism – to leverage the rich inductive potential of natural kinds – at least three functions must be carried out by the cognitive mechanisms underlying an essentialist inference system: identifying essentializable kinds, building the appropriate representations of them and organizing knowledge about them, and making inferences based on those representations (Atran, 1990, 1994, 1998; Gelman & Markman, 1987, 1987; Gelman & Wellman, 1991; Keil, 1989, 1994; Medin & Ortony, 1989).

1.6. Identification function: picking out essentializable kinds in the environment

In order to benefit at all from the rich inductive potential of living kinds, it is crucial to identify these kinds properly. It is also important to be able to reidentify them in different situations and to retain the proper identity groupings across the lifespan, to conceptually track kinds across time and space (Keil, 1989; Millikan, 1998, 2000). Thus, the first function that essentializing mechanisms must carry out is to pick out essentializable kinds from the (initially) undifferentiated soup of things in

the environment: to identify things that can be validly essentialized, and to group things with "common essences" together. If essentialism is to play a role in guiding inductive inferences, essences must be represented not as aspects of individuals, but of kinds or classes of things. As Locke observed,

> Essence, in the ordinary use of the word, relates to sorts, and … it is considered in particular beings no further than as they are ranked into sorts … Other creatures of my shape may be made with more and better, or fewer and worse faculties than I have; and others may have reason and sense in a shape and body very different from mine. None of these are essential to the one or the other, or to any individual whatever, till the mind refers it to some sort or species of things.
> (Locke, 1689, Book 3, Chapter 6)

In other words, if essentializing mechanisms exist to exploit the rich inductive potential of natural kinds, then individual entities must not only be identified as essentializable things, but also as members of essentializable kinds. And once such entities have been identified, the problem alluded to by Locke must also be solved, i.e., the problem of determining which of the properties of individuals can validly be attributed to their kind-specific essence (see below).

Just how the identification function of essentializing mechanisms might be carried out, cognitively, is the subject of much empirical research. Clearly, what is needed is some sort of sorting mechanism or mechanisms, i.e., mechanisms that not only categorize objects in the world into essentializable and non-essentializable kinds during development, but that also instruct the developing representational system to treat them differently, perhaps by activating further specialized mechanisms in the case of essentializable kinds. A promising avenue of work in this direction has focused on infants' abilities to distinguish animate from inanimate objects, which may be the basis of a major distinction in the kind categorization system (Gelman, 1990; Gelman & Spelke, 1981; Leslie, 1994; Mandler, 1992; Premack, 1990). The perceptual systems of infants and children seem to be tuned to those sorts of cues that would have been most reliable in picking out these kinds in natural environments, such as motion (Leslie and Keeble, 1987; Premack, 1990) and surface features such as texture or shape (Mandler, Bauer, & McDonough, 1991; Mandler & McDonough, 1998). Functional response to intentional stimulus may be another important kind of cue: for example, following eye gaze, looking in response to pointing, running in response to chasing, and so on (Gelman & Spelke, 1981; Gergely, Nádasdy, Csibra, & Bíró, 1995; Leslie, 1994; Mandler, 1992). Moreover, developmental work by Keil (1989), Gelman and Markman (1986, 1987), Gelman and Wellman (1991), Simons and Keil (1995), and others has shown that knowledge of origins (specifically, origins in reproduction by another member of the same kind) of a particular object may be a crucial factor in essentializing it. These distinctions between kinds of kinds that emerge early and reliably during development may provide the basis for the different modes of essentialist representation and inference discussed in Section 2.

1.7. Conceptual organization function: learning about kinds and building essentialized concepts

It is important here to remember Locke's distinction between real and nominal essences. Locke held that we never perceive true essences, but only their effects; i.e., the properties that they tend to produce in the kinds of things that have them (Locke, 1689). Thus, while we might identify a tiger by its stripes, whiskers, tail, teeth, claws, shape, texture, and so on, these things are not the "true" essence of the tiger, only part of its "nominal" essence, the set of properties caused by its true essence, and by virtue of which we identify the essential kind. Given that humans have no direct access to the true executive causes of the richly correlated features of living kinds, we might expect essentializing mechanisms to respect Locke's distinction between real and nominal essences: kinds might be identified or individuated by certain properties, but these properties are not then represented as what *causes* individuals that possess them to be members of their kind. For example, we do not expect mere surface changes to affect kind membership, even if they alter certain properties normally diagnostic of kind membership (see, e.g., Keil, 1989, 1994; Simons & Keil, 1995).

After having sorted entities in the world according to kind, a second critical function of essentializing mechanisms would be to represent membership in these kinds as being more than simply a matter of possessing some set of necessary and sufficient features. These features must be represented as tied to an executive cause, even if it is as minimal as an essence placeholder. If kind concepts consisted merely of descriptions of the properties that tend to be associated with the kind, there would be no assumption of rich inductive potential; the discovery of a new kind of organ in an animal, for example, would have the same status as the discovery that pencils tend to be made of wood.

This is not to say, however, that feature correlations cannot be used by conceptual development mechanisms to identify true kinds in nature, and to find the "joints" between them. Indeed, as Locke pointed out, we expect the use of such nominal essences for identification of true kinds, and also for representation of their natures. But for a system that captures the nature of living kinds, none of these "nominal essences" should be confused, representationally speaking, with the true essence itself, even if it is represented only by a placeholder (Gelman *et al.*, 1994; Gelman & Diesendruck, 1999; Keil, 1989). Thus, we expect conceptual development mechanisms to be sensitive to the correlated clusters of features that co-occur across individual living things in the environment, and to use these feature clusters to sort living things into kinds. However, these features should be represented as resulting from kind membership, not definitive of it.

1.8. Inference constraining function: constraining extensions of properties

Athird function of essentializing mechanisms would be to guide and constrain the generation of inferences themselves, once essentialized kind concepts were in place. As mentioned before, inductive inferences – such as categorical inferences from kind to property, or from property to kind – would have been particularly facilitated by the

essentialization of kind concepts. A key problem here would be that of avoiding inductive promiscuity, or the unfettered explosion of inductive inferences arising from the assumption of rich inductive potential. Indeed, the assumption of rich inductive potential is a double-edged sword: while rich inductive potential is the key feature of natural kinds that essentialism exploits, it can also be taken too far. The problem is to restrict induction only to those sorts of properties that can be validly essentialized, and avoiding the inductive generalization of accidental ones (Gelman, 1988; Gelman & Wellman, 1991; Keil, 1989; Keil *et al.,* 1998). If one assumes that, once one has found the proper grouping of living things into kinds, there exist a potentially unlimited number of properties that result from kind membership, what prevents the extension of *any* discovered trait across members of the kind? For example, if we encounter one dog with wet fur and four legs, and another dog with wet fur and four legs, what is to prevent us from assuming that all dogs have wet fur and four legs?

What is needed is some kind of constraint, such that inductive inferences would be restricted, at least on average, to the sorts of properties valid for kind membership. If the benefits of exploiting rich inductive potential are not to be cancelled by the negative effects of inductive promiscuity, one would expect essentialist inference systems to possess evolved constraints that restrict inductions to, or at least bias them towards, classes of traits that are validly essentializable, while excluding accidental ones. Experimental studies have shown that children, in their inductive extensions of properties across kinds, obey distinctions between essential properties (such as having four legs, in the case of dogs) and accidental properties (such as being wet) (Gelman, 1988; Keil *et al.*, 1998; Springer, 1992). The nature of the mechanisms that lead children to reliably make these distinctions, however, remains to be elucidated.

In summary, there seems to be general agreement in the literature that at least three functions must be carried out by evolved essentialist representational systems: identifying essentializable kinds, constructing representations of them that embody the assumptions of executive causation and rich inductive potential, and constraining inferences to validly essentializable properties for a given kind. This leaves unanswered one important question: what kinds of kinds did essentialism evolve to handle?

2. How specialized must essentialism be?

What essentialism is, and what its functions are, have been well defined in the cognitive science literature, as reviewed in Section 1. There seems to be general agreement (at least among those who adopt a functionalist approach to cognition) that to essentialize certain kinds of things can be an adaptive, ecologically rational strategy for purposes of inference. There is considerable disagreement, however, over the kinds of things that essentialism might have evolved to help us make inferences about. In other words, there is disagreement about essentialism's proper domain (Gelman *et al.,* 1994; Hirschfeld, 1994; Gelman & Hirschfeld, 1999)[4].

[4] For discussions of the distinction between proper and actual domains, see Millikan (1984) and Sperber (1994).

Gelman and Hirschfeld (1999) summarize one of the key questions of this debate well in an essay entitled "How biological is essentialism?" Although biological taxa are uncontroversial examples of natural kinds that satisfy the core essentialist assumptions of executive causation and rich inductive potential (Atran, 1994, 1995; Carey, 1995; Keil, 1995), as outlined above, the evidence suggests that biological taxa are not the only kinds that people are prone to essentialize. For example, there exist data showing that people are predisposed to be essentialist about a variety of categories of things that are by no means uncontroversial examples of natural kinds. The most contentious of these are "social" (i.e., human) categories such as sex (Fuss, 1989; Gelman, Collman, & Maccoby, 1986; Taylor, 1994; Taylor & Gelman, 1991), personality types (Gelman, 1992; Yuill, 1992), race and ethnicity (Allport, 1954; Banton, 1987; Gil-White, 2001; Hirschfeld, 1994, 1995; Stoler, 1992, 1995), kinship (Hirschfeld, 1986; Springer, 1995, 1996), and even other social categories such as profession (Boyer, 1990, 1994; Hirschfeld, 1994, 1995).

There are a variety of ways to account for extensions of essentialism across kind domains. Gelman *et al.* (1994) summarize these as: (1) borrowing from a base domain, e.g., extending a properly "biological" mode of construal to non-"biological" kinds (see below for a discussion of the meaning of "biological" in this context); (2) domain specificity, but in a proper domain that is broader than just biological taxa; (3) multiple domain-specific notions (e.g., separate "essentialisms" for biological versus social kinds); (4) domain generality with different, domain-specific instantiations.

Although these are four distinct possible ways to account for observed patterns of essentialist thought, their predictions may be quite difficult to tease apart, empirically. What is needed, in order to do so, is to examine the empirical consequences of the evolutionary claim that essentialism evolved specifically due to the benefits of being able to make reliable inductive inferences about particular kinds of things in ancestral human environments. In order for there to have evolved an essentialism particular to a specific kind of thing, three conditions must have been met: 1) that kind of thing had to exist in human ancestral environments; 2) inferences and decisions about exemplars of that kind had to have had fitness consequences; and 3) there must be principles of valid inference, specific to that kind, that selected for an essentialist architecture specific to it. In principle, these requirements allow us to judge what kinds of kinds can be considered candidate targets for an evolved essentialist architecture. The third requirement is particularly important, because it allows us not only to determine whether a particular might be a candidate target of evolved essentialism, but also whether there is likely to be a kind of essentialism *specific* to that kind of thing. Indeed, the following analysis will suggest that there are *a priori* reasons to expect the existence of multiple kinds or modes of essentialism, because a single set of essentialist assumptions is unlikely to produce valid inferences for all kinds, from nonbiological substances such as water and gold to biological kinds such as predators. It is this functional incompatibility (Sherry & Schachter, 1987) that leads one to expect specialized mechanisms for dealing with particular kinds of kinds.

2.1. Is essentialism strictly biological?

It has been suggested that the kinds of things that best satisfy the core assumptions of essentialism are "living kinds", particularly, biological taxa such as tigers and goldfish (Atran, 1994, 1998; Coley, Medin, & Atran, 1997; Gelman & Wellman 1986, 1987; Keil 1989, 1995). Living things do satisfy the two core essentialist assumptions discussed in Section 1, namely, the assumptions of executive causation and of rich inductive potential. This is because the processes whereby organisms develop are under the executive control of DNA and cellular regulatory machinery, which have cascading effects that lead to a large and potentially infinite number of inducible properties, by virtue of the fact that members of the same taxon share many nucleotide sequences, or many functionally equivalent nucleotide sequences. If one wants to make an inductive generalization about some phenotypic property from one organism to another, then, using taxonomic relatedness as an axis of generalization is a safe bet (Atran 1990, 1994, 1998; Coley *et al.*, 1997; Sober, 1988). The fact that people show a declining willingness to inductively generalize a trait between two organisms as the taxonomic distance between them increases, as shown by Coley *et al.* (1997), is consistent with this analysis, and shows that people closely follow the predictions of a functionalist account of essentialism in the case of biological taxa. Moreover, a variety of experiments have shown that children spontaneously use taxonomic relatedness (rather than, for example, perceptual similarity *per se*) as an axis for the generalization of phenotypic traits (Gelman, 1988; Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Gelman & Wellman, 1991; Keil, 1989).

There seems little disagreement, in the face of such facts, that if essentialism has an evolved target domain, it is likely at least to include the domain of category-based inductions about biological taxa. But there are other kinds of natural kinds, apart from biological taxa, that are often said to be ideal candidates for essentialization as well. For example, nonliving substances, such as gold and water, are as uncontroversial as examples of natural kinds as are tigers and goldfish, and figure just as prominently in discussions of essentialism (Carey, 1985; Gelman, 1988; Keil, 1989; Malt, 1984; Millikan, 1998, 2000; Putnam, 1975). But the principles that make substances such as gold and water natural kinds are quite different from the principles that make tigers and goldfish natural kinds. Would the same set of essentialist assumptions be valid across these diverse kinds of kinds?

2.2. Nonbiological substances[5]

In a classic paper, Putnam (1975) claimed that what makes water what it is, and what is responsible for its various properties (e.g., that it is a clear, odorless liquid),

[5] The term "substance", as used here, follows the folk usage employed by authors such as Au (1994). Millikan (1998, 2000) uses the term in the Aristotelian sense, to refer to any entity that "retains its properties" over time, including kinds (e.g., cats), individuals (e.g., Bill Clinton), event types (e.g., Beethoven's 5th Symphony), and "stuffs" (e.g., ice) (Millikan, 2000, p. 2-3). Substances, as referred to in this paper, correspond to Millikan's stuffs.

is a sort of essence that we can approximately designate with the term H_2O, and that determines the referential extension of the word "water". According to Putnam's account, substance kinds such as water, gold, and so on, are natural kinds by virtue of having such essences (as are biological kinds such as lemons and tigers). On this view, substances such as chemical compounds are good candidates as members of the proper domain of essentialism.

There also exist substances which are of biological origin, even if they are not organisms in and of themselves, such as milk, juice, meat, wood, and so on. Might these be candidates for essentialization as well? Several studies have shown that children'thinking about both biological and non-biological substances shows some aspects of essentialism: for instance, they assume that substances maintain their properties even when chopped into pieces, transformed into different shapes, and so on (Au, 1994; Dickinson, 1987; Smith, Carey, & Wiser, 1985)[6].

Let us first consider nonbiological substances in relation to living things. We have already seen that biological kinds are excellent candidates as targets of the essentialist cognitive architecture sketched in Section 1. Do nonbiological substances such as water and gold satisfy the core assumptions of essentialism in the same way as living organisms?

Not to the same degree, because the nature of the causal pathways between "essences", or executive causal agents, and inducible properties are much different for nonbiological substances and for living things, making the inductive potential of living things far richer. Consider the kinds of inference problems that might have faced an ancestral human with regard to water. Water has a variety of properties that are relevant to human behavior and decision-making: it can and must be consumed to hydrate the body; food can be boiled in it; it is clear and odorless unless something has been dissolved in it; and so on. But the inductive potential of water is not terribly rich compared to, say, the inductive potential of lions. Once one has learned the functional properties of water that are relevant for everyday decision making – a list of which is probably short – there will not be a large and potentially infinite number of properties waiting to be induced. Moreover, the inductive advantage of assuming an executive causal agent – in Locke's terms, a real essence distinct from a nominal essence – is not clear in the case of water. What does it get us, for example, to assume that the surface properties of water are in fact caused by some underlying, invisible essence known as H_2O? Why not just assume that the observable properties of water are simply those of the "essence" itself, rather than being produced by it via some unknown causal pathway?

The case for whole organisms is much different. Unlike water, each exemplar of which shares a relatively small and finite number of interesting properties with every other exemplar (wet, odorless, drinkable, etc.), a lion shares an immense number of properties, many of which remain to be discovered, with every other lion. Aside from obvious ones, such as ferocity and stripes, consider, for example, the complexities of lion behavior, not to mention the practically limitless number of finer but still reliably inducible details such as the organization of the lion's retina,

[6] But see Malt (1994) for an argument that people do not essentialize substances.

arterial branching patterns, and so on. In fact, a lion also shares a huge number of properties with other phylogenetically related organisms that are not lions, including housecats and crocodiles, and the reliability of inducing any given trait varies systematically with phylogenetic relatedness (Sober, 1988). The number of traits that can be induced from water to other substance kinds, on the other hand, is relatively small, because water does not reproduce and thereby transmit complex arrays of properties to other individuals.

There is an even more crucial difference between substances and whole organisms: while substances are homogeneous, such that they can be divided and recombined and still retain their properties, organisms are not. For example, the kind-inducible properties of gold hold for any quantity of gold, in any particular shape or configuration, but what is true of a tiger is not necessarily true of a piece of a tiger. And many of the inducible properties of tigers that are important for human behavior and decision making are whole-body properties. The most important of these, arguably, are their psychology and behavior, and aspects of morphology that are relevant to behavior, such as teeth, claws, and so on. Interestingly, Gelman and Wellman (1991) have found that for living things, children tend to essentialize behavioral traits even more than morphological ones. This is consistent with the idea that, for human-decision makers in ancestral environments, behavioral dispositions may have been among the most important whole-body properties to generalize from member to member of a particular living kind. This is not true of substances[7].

We expect, then, if not two wholly different kinds of essentialism, at least two distinct manifestations. The first is an essentialism specific to substance kinds, which is inductively "shallow" in that it does not assume extremely rich inductive potential, and which includes a homogeneity assumption, supporting inductive generalization of properties across portions of substances, no matter how they are divided or combined (Au, 1994; Carey, 1985; Smith, Carey, & Wiser, 1985). As Au (1994) and others have shown, peoples' essentialist intuitions reflect this assumption: even young children assume, for example, that if you cut a cube of sugar in half, the resulting portions will still retain the property of sweetness that characterized the entire cube (see also Carey, 1985; Smith, Carey, & Wiser, 1985). If and when people apply the homogeneity assumption to animals, they are likely thinking of them under a different mode of construal, as substances rather than as organisms. In other words, if you think of a property of a tiger that holds for any portion of the tiger, you are probably thinking of it as meat (see below for discussion of biological substances).

On the other hand, because of the unique inductive properties of whole-body living organisms, we expect a second, different kind of essentialism specific to them.

[7] In fact, if it is true that the whole-organism mode of construal would have been most useful for thinking about *behavioral* traits, it may be that there would have been relatively few uses in ancestral environments for whole-organism construal of plants. Rather, plants may more often be thought of under a substance mode of construal, which would be relevant for making inferences about them as building materials, food, sources of medicinal compounds, and so on.

Whole-organism essentialism should reflect an assumption of rich inductive potential to a degree not seen for substances. Moreover, whole-organism essentialism should not incorporate the homogeneity assumption. Instead, one expects an implicit assumption that many if not most whole-body properties will not be inducible to body parts. This would have been true for many of the kind-specific aspects of organisms that would have been most important for decision-makers in ancestral environments, most notably, behavior.

2.3. Biological substances

A key inductive quality that distinguishes nonbiological substances is homogeneity: their inducible properties obtain for any portion of the substance, and are not dependent on any particular whole-object configuration. This is easily seen for compounds that are homogeneous on the level of molecular composition, such as water or gold. But what about substances that are biological in origin, such as meat, wood, and milk?

Under a microscope, these substances do not appear homogeneous at all. However, for the kinds of inferences that are important for everyday decision making, meat, wood and milk do qualify for substancehood, because their important macroscopic properties are homogeneously distributed. Consider, for example, the properties of milk and meat that are important for cooking and consumption, and the properties of wood that are important for building fires or constructing artifacts. These are certainly not whole-body properties, as they apply to any portion of milk, meat, or wood, as long as they are within the macroscopic size range[8]. Thus, meeting the homogeneity assumption is one way, in terms of inductive principles, in which biological substances bear a closer resemblance to nonbiological substances than to whole-body living kinds.

What about the assumption of rich inductive potential with regard to biological substances? It has been argued that the sorts of inducible properties of organisms that might have been most important for human decision-making in ancestral environments were complexly caused whole-body properties, such as aspects of psychology and behavior. Because biological substances such as milk, meat, and wood do not have such properties, they clearly lie closer to nonbiological substances such as water and gold in this regard. While biological substances may have some complexly caused properties – consider, for example, the potential pharmacological properties of substances derived from plants – it seems likely that a single mode of construal for substances could handle both biological and nonbiological substance kinds[9].

[8] There may be interesting exceptions to this; consider sawdust, which people may think of as a different "kind" than wood, because it does not have many of the functional properties that characterize wood in its rigid form; see Au (1994).

[9] An exception to this may be that taxonomic relatedness of the sources of biological substances may support induction in ways that are not true for nonbiological substances (consider properties inducible between cow's milk and goat's milk). It also bears noting that it is not entirely clear where plants might fall under the scheme presented here; while they do have whole-body properties, most of their proper-

If it is true that there are at least two kinds of essentialism, one for substances and one for whole organisms, and that biological substances such as milk, meat, and wood fall in the proper domain of substance essentialism, then it would be the case that the line between the substance and organismal modes of essentialism does not lie strictly between the biological and the nonbiological. In fact, if the most important class of whole-body properties that distinguishes organisms is behavior, the line may lie instead between the animate and the inanimate[10]. Under this scheme there is a commonly occurring class of event in which entities would regularly make the transition from one mode of construal to another: death. When an animal is killed and processed to become food, inferences about it must switch from focusing on whole-body properties such as behavior – which are not only no longer important, but also no longer apply – to focusing on homogeneous properties of its constituent substances (e.g., meat). Thus, it may not be the case that once a particular essentialist construal is applied to an entity, it necessarily remains permanently attached. In the case of death, and perhaps for other kinds of events or situations as well, we expect people to be able to shift their representation of a single entity from one major ontological class and mode of construal to another.

2.4. Artifacts

Artifacts have long been of interest to those who study categorization and mental representation, precisely because of their curious status as kinds (Bloom, 1996; Keil, 1989; Matan & Carey, 2001; Rosch & Mervis, 1975; Smith & Medin, 1981). While it is generally agreed that artifacts such as chairs and wastebaskets are not natural kinds in the same sense as gold and tigers, membership in artifact categories is neither entirely arbitrary nor rule-based (Rosch & Mervis, 1975). Might the inductive assumptions of essentialism apply to artifacts?

Bloom (1996) proposed that the "kind" status of artifacts is rooted, from the point of view of mental representation, in the intentions of the designer (the "intentional-historical" theory, see also Dennett's (1987) "design stance" theory). According to this theory, the representation of a particular artifact as being a member of a particular kind category, such as CHAIR or WASTEBASKET, is driven by a knowledge of origins, which outweighs factors such as perceptual similarity to other artifacts. For example, upon knowing that an object was designed to be a flowerpot – despite its strong perceptual similarity to a coffee pot – we would be inclined to call it, and represent it as, a member of the artifact category FLOWER POT. On this theory, then, our system of representing artifacts would share a common feature with

ties that are useful for everyday human decision making could be handled by a substance mode of construal. Finally, contagious diseases and other transmissible entities of biological origin may constitute an additional essentialist domain, as the assumptions necessary to support valid inferences about them might be incompatible with both the substance and whole-organism modes of construal. See Rozin, Millman, and Nemeroff (1986) for an outline of the inductive principles of this domain.

[10] There is, in fact, evidence that the animate / inanimate distinction may be ontologically prior to the living / nonliving distinction, and that young children, lacking the latter distinction, may use the term "alive" to mean "animate" (see, e.g., Carey, 1985; Leslie, 1994; Mandler, 1992; Piaget, 1951).

biological essentialism: knowledge of origins would trump perceptual similarity. Gelman & Bloom (2000), Keleman (1999) and Matan & Carey (2001) have provided empirical evidence that intended function is indeed what guides peoples' intuitions about artifact category membership.

Matan and Carey (2001), while noting that artifacts "do not naturally fall in the realm of psychological essentialism nor in the realm of framework theories", suggest that, if the assumptions of the design-stance theory are correct, in the case of artifacts, "The intended function is the factor which determines the artifact's properties, the actual functions it can serve (the intended function as well as others) and its kind. In that sense, the original intended function is the artifact's essence" (Matan & Carey, 2001, p. 2). Matan and Carey are thus suggesting that people should exhibit certain symptoms of the essentialist mode of construal with regard to artifacts, and they provide experimental evidence that people indeed do so: their data show that people privilege intentional origins over appearance when representing kind membership, just as they privilege reproductive origins for biological kind membership. But there is more to essentialism than merely privileging origins. How would artifacts fit into the evolutionary-functionalist framework outlined above? Do we expect the same kind of essentialism that is applied to biological kinds (but not, by hypothesis, to nonliving substances) to be applied to artifacts?

Artifacts are interesting because, while they are not living things in and of themselves, they are biological in origin. As Dawkins (1982) has pointed out, they are products of the evolutionary process, albeit of a special type: they are part of the "extended phenotype" of our species, just as beaver dams, spider webs, and bee hives are parts of the extended phenotypes of beaver, spiders, and bees, respectively. Thus, they do have biological origins, and their properties can be thought of as goal-directed and functionally designed in the same way that the properties of living things can be construed [11]. Moreover, there are potentially many properties of an artifact that might be induced from the knowledge of its origins in the intentions of its designer. As for any biological entity, the origins of artifacts matter, though their reproduction occurs through different processes than the reproduction of whole organisms.

Reproduction is a key aspect in which artifacts differ from living things: unlike organisms, artifacts do not reproduce themselves. In biological reproduction, organisms transmit huge numbers of properties from themselves to their offspring, because offspring carry a nearly identical copy of the genetic material of their parents, and the genetic material influences a vast number of phenotypic traits. These include not only traits that have been acted on by selection, but any genetically influenced trait (consider the number of dimensions along which two leopards are the same). The case with artifacts, on the other hand, is different: while artifacts may

[11] Of course, while at least some of the design features of artifacts are intentional in orgin, the design features of organisms are not. Under the theory presented here, it is not necessary to represent a "designer" to reap the benefits of an essentialist stance, because essentialist assumptions hold true even under ignorance of true causes (see Section 1). Nevertheless, the fact that artifacts do have a clear designer whereas organisms do not is an interesting one, and may bear on the design of evolved inference systems as well as on religious intuitions.

indeed be copied, the copying mechanism is quite different. Copying of artifacts is under the control of human cognition, and therefore, for a variety of reasons, we do not expect the same degree of copying fidelity for artifacts as we do for genetically controlled replication. For example, chairs have a variety of functional features that we expect, in general, to be preserved across exemplars, such as various aspects of fit to the human body form. These invariant properties result largely from the fact that designers of chairs seek a common functional target, or at least a common region in functionality space (Keil, 1995). However, there are many dimensions along which chairs might vary, and yet still achieve their intended functionality (consider rocking chairs, desk chairs, beanbag chairs, lazyboys, and so on). In fact, we expect designers to exploit these degrees of freedom in expressing creativity and innovation. Moreover, while artifacts have a history, artifact kinds can spring into being quickly and disappear quickly, at the whims of their designers. Thus, artifacts do not satisfy the assumption of rich inductive potential to the same degree as living things. If I tell you that X is a chair, and ask you to induce properties of X, the list of induced properties that you can be sure about will not be nearly as long as if I told you that X were a leopard, and asked you to do the same.

There is, however, a caveat with regard to these points about artifacts. Recently in human evolutionary history, there have emerged artifact kinds that are so complex, and so carefully reproduced, that they approach the rich inductive potential of true living kinds. Consider an example from Millikan (1984, 2000): an automobile, the 1969 Plymouth Valiant 100:

> … in 1969 every ’69 Valiant shared with every other each of the properties described in the ’69 Valiant’s handbook and many other properties as well. And there was a good though complicated explanation for the fact that they *shared* these properties. They all originated with the selfsame plan – not just with identical plans but with the same plan *token* … [Thus they] had such and such strengths, dispositions, and weaknesses … placement of distributor … size of piston rings … shape of door handles … the fenders of the ’69 Valiant that has not been garaged tend to rust out whereas the body stands up much better; the ball joints are liable to need replacing after relatively few thousands of miles whereas the engine … is not likely to burn oil until 100,000 miles.
> (Millikan, 1984, pp. 279-80; cited in Millikan, 2000)

Thus, to the observer, a 1969 Valiant has nearly the inductive status of a living thing; a huge number of things I may discover about it (e.g., the size of the piston rings) are likely to hold for other 1969 Valiants as well. Indeed, I am not likely to exhaust through casual inspection the number of kind-specific inducible traits. Thus, 1969 Valiants, and a whole class of complex high-tech artifacts, approach the rich inductive potential of living kinds such as leopards, and satisfy the assumption of executive causation (in the form of their manufacturing plans) (see Keil, 1989, for a discussion of complex artifacts). But artifacts of such complexity are evolutionarily novel. They may indeed be processed, and validly so, by essentializing mechanisms

originally evolved to handle living kinds[12], but they cannot be part of the evolutionarily proper domain of such mechanisms. On the other hand, humans have been making simple artifacts, such as tools, clothes, houses, and so on, for a long enough period of evolutionary time for these to be considered as possible targets of evolved essentialism. Based on all these considerations, then, do we expect (simple) artifacts to be part of the proper domain of essentialism?

It is unlikely, because the inductive benefits of assuming an essence aren't rich enough. While the assumption of executive causation holds for artifacts, the assumption of rich inductive potential does not, at least not to the same degree as for organisms. Nor, in fact, does the homogeneity assumption of substance essentialism hold. It is true that artifacts are objects of biological origin, being part of the extended phenotype of our species. Many of their properties do have an executive cause, namely, the intentions (and manufacturing abilities) of their designers[13]. And they can be grouped into kinds based on their intended functions: chairs are intended to be sat upon, knives are intended to cut things, and so on. Thus, as Bloom (1996) and Matan and Carey (2001) point out, origins should be important for artifacts, just as they are for living kinds, but this simply means that they should be *categorized* according to origin, not necessarily essentialized. Based on the considerations discussed here, we don't expect an evolved essentialism specific to artifacts, nor for them to fit into evolved schemes for other essentialized kinds, except by mimicry (e.g., automobiles, computers, and other complex artifacts).

2.5. Living kinds

So far the notion of "living kind" has been treated unproblematically, yet it is a far from unproblematic notion. While taxonomic kinds such as tigers and ostriches are generally taken, at least by implication, as bona fide examples of natural kinds (e.g., Atran, 1994, 1998; Berlin, 1992; Coley, Medin, & Atran, 1997; Keil, 1989; Keil, 1994; Putnam, 1975; Quine, 1977), there are many other ways of grouping living things, some of which cut across taxonomic boundaries (e.g., predators), others of which are nested within them (e.g. sex, race, personality, profession). In the latter case, the possibility that people have a propensity to essentialize human kinds such as race and sex – along with empirical evidence that they do indeed do so – has generated considerable controversy and debate (Atran, 1990; Banton, 1987; Boyer, 1990; Fuss, 1989; Gelman & Hirschfeld, 1999; Gelman, Collman, & Maccoby, 1986; Gil-White, 2001; Hirschfeld, 1994, 1995, 1996; Rothbart & Taylor, 1990;

[12] Not only do automobiles obey the core essentialist assumptions of executive causation and rich inductive potential, they also emit cues, such as self-propelled motion, goal-directed behavior (responding to changes in road direction, other vehicles, etc.), and even eye-like headlights, which may serve as triggers for processing as living kinds. Interestingly, there is some evidence that vehicles dissociate with living kinds in category-specific cognitive deficits that result from neurological damage (see, e.g., Fig. 3.1 in Caramazza, Hillis, Leek & Miozzo, 1994, p. 72).

[13] This raises the interesting possibility that if an essence were represented for an artifact, it might be represented as internal to the designer rather than internal to the artifact itself.

Taylor, 1996). Strictly speaking, all of these groupings are groupings of living things, and so potentially might come within the domain of "biological" essentialism, at least as defined here. However, the analysis so far has focused on taxonomic kinds. What of these other groupings: are all of them equally valid, and would they have been so in ancestral environments? Is their essentialization merely the result of "overextension" of a mode of construal evolved to deal with taxonomic kinds, as some have suggested (Atran, 1990; Boyer, 1990)? It is important to examine, as we did for substances, the processes that make living kinds cohere for the purposes of induction, thus rendering them good targets for an essentialist mode of construal, and then to ask what potential groupings of living things into kinds would fit this framework.

2.6. Two processes that generate living kinds: descent and design

As we saw in the discussion of substances and artifacts, one of the things that makes living things unique – and that distinguishes them from other kinds, such as substances and artifacts – is that they have enormous numbers of properties which are transmitted directly from parent to offspring via biological reproduction, and which are thus shared with other individuals. The process of biological reproduction is what is responsible for the rich inductive potential that is the characteristic feature of living kinds: because the genome is reproduced with extraordinarily high fidelity, and because it influences the development of such a large number of traits, individuals that share a common ancestor share a potentially vast number of properties. However, because genetic replication is not perfect – mutations and other replication errors sometimes occur, and lineages diverge over time – the genetic similarity of individuals gradually declines as the number of generations since the individuals diverged from their most recent common ancestor increases. These facts account for the nested hierarchical nature of phylogenetic relationships: living things can be grouped according to most recent common ancestor, and the resulting groupings can be nested in taxonomic hierarchies from species to genera, families, orders, and so on (Atran, 1990; Coley, Medin, & Atran, 1997; Sober, 1988). Not only are these groupings convenient to humans for organizing the diversity of living things with which they are confronted, they are also real. Because all traits influenced by the genome are, *ceteris paribus,* transmitted from parents to offspring, the more closely related by descent any two individuals are, the more likely they are to share any given trait: this makes biological taxonomic kinds true kinds and inductively valid ones, satisfying the core assumptions of essentialism outlined in Section 1 (Coley, Medin, & Atran, 1997).

But the transmission of properties via descent entails another conclusion: while membership in kind categories is mutually exclusive for categories at the same taxonomic rank (e.g., an organism cannot simultaneously be a member of two different genera), membership in kind categories of different ranks is not only possible, but necessary, because biological reproduction leads to a nested hierarchy of descent relationships. Species are nested with genera, which are nested within families, and so on, and each of these categories is a bona fide natural kind. For example, a

colobus monkey is at the same time a colobus monkey, a primate, and a mammal, all of which are inductively valid kind categories, although their inductive validity declines with increasing taxonomic rank from colobus monkey to mammal (Coley, Medin, & Atran, 1997). While two colobus monkeys are quite likely to share a particular trait, a colobus monkey and a cow are more likely to share a given trait than are a colobus monkey and a goldfish, because the colobus monkey and the cow are both mammals, and share a more recent common ancestor than does either with goldfish[14]. Thus, an organism can be at the same time a member of more than one taxonomic kind: colobus monkey, primate, and mammal are all true natural kinds, and not mutually exclusive. Moreover, all are validly essentializable under the criteria presented here. It would be reasonable, then, to expect people to essentialize mammals (consider the large number of traits that may be considered to result from the mammal essence: warm blood, fur, lactation, live birth, four limbs, etc.), and at the same time to essentialize tigers (stripes, ferocity, and so on). This conclusion runs contra to the claim that essentialism entails a mutual exclusivity assumption, i.e., that an entity cannot simultaneously be a member of more than one essentialized kind category (Carey 1995; Kalish, 1995)[15].

Another process that generates inductively valid biological kinds is natural selection. Whereas descent by reproduction transmits properties from individual to individual without regard to their function, natural selection causes change by favoring traits that improve the adaptive fit between organisms and their environment. The process of convergent evolution, whereby organisms evolve similar solutions to similar adaptive problems, results in design relationships between organisms that are not due to descent from a common ancestor. For example, ecologists recognize non-taxonomic natural kinds such as predators, herbivores, and so on. The inductive generalizations of scientific fields such as behavioral ecology are made possible by the fact that kinds such as predators have regular, stable properties. This holds for non-technical, common sense inference as well: lions and crocodile share properties by virtue of design that they do not share by virtue of descent. A cow, for example, is more closely related to a lion than a crocodile is, yet the lion and the crocodile share traits not shared by the cow.

It thus appears to be that there are multiple kinds of inductively valid biological kinds, and an individual organism could simultaneously be a member of multiple kind categories, including both *taxonomic* and *role* kinds (e.g., lion, felid, mammal, predator). In keeping with this conclusion, there is experimental evidence that people essentialize organisms at multiple taxonomic levels simultaneously (Coley, Medin, & Atran, 1997), and that they essentialize not only taxonomic categories

[14] In the nomenclature of modern systematics, taxonomic groups such as primates and mammals are real, inductively valid natural kinds because they are monophyletic, i.e., all and only members of the group share a common ancestor (Sober, 1988). There has been a debate in systematics over whether to reorganize taxonomic nomenclature so that all taxonomic names used by systematists would strictly refer to monophyletic groups; under such a scheme, for example, birds would become a category of reptiles.

[15] It does not, however, invalidate the claim that an entity cannot simultaneously be a member of more than one essentialized category in the same rank level.

such as mammal, but functional categories such as predator as well (Barrett, Cosmides, & Tooby, forthcoming). The fact that people can be made to shift between overlapping systems of categorization (e.g. taxonomic vs. functional), for the purposes of different kinds of inference, suggests that people do not place a given organism in a single essentialized kind category, with a single indivisible essence. Rather, essences may be more complicated: people may simultaneously essentialize the category cat, the category lion, and the category predator, and not regard these essences as mutually exclusive at all. In terms of their inferential properties, essences might behave rather like substances: for example, they could be mixed in different proportions in different individuals[16]. We might expect this, if evolved inference systems were not to be confounded by such phenomena as hybridization (and within-species kinship; see below). Indeed, the ability to represent multiple, non-mutually-exclusive essences would seem to be a prerequisite for the ability, which people apparently have, to simultaneously essentialize human nature in general and also within-human categories such as sex and / or race (see below).

2.7. Within-species role kinds

There exist inductively valid, within species kinds as well, which, in addition to cross-species kinds such as predators and herbivores, can be thought of as role kinds: these include, for example, males and females, parents, siblings, and so on. Consider sex. Males and females differ in principled ways because of a history of selection to solve different kinds of problems (Symons, 1979; Trivers, 1972). There are thus sex-specific phenotypic properties, including psychological, behavioral as well as morphological traits, which can be inductively generalized within sex (MacCoby & Jacklin, 1974; Symons, 1979). It is important to note that not all sex differences need have a direct genetic basis for them to be inductively valid. Sex-specific patterns of interaction and socialization can also result in robust sex differences, which can be reliably generalized within sex in a particular culture, and sometimes across cultures as well. The fact that not all of these differences may be generalizable across cultures does not pose a problem for an evolutionary account, because what would have been important for ancestral decision-makers would have been inductive generalization within one's local environment. Thus, within-species role categories such as sex may be part of the evolved proper domain of whole-organism, biological essentialism, and a disposition to essentialize sex may not be surprising (Fuss, 1989; Gelman, Collman, & Maccoby, 1986; Taylor, 1996). It is likely that there also exist other validly essentializable within-species ecological roles that lie within the proper domain of essentialism, such as parents, siblings, offspring, and so on.

[16] It is sometimes assumed that rigid category boundaries are a necessary assumption of an essentializing conceptual system. Kalish (1995), for example, assumes that graded category membership and essentialism are incompatible, and takes evidence for graded categories to be evidence against essentialism. The view developed here, however, suggests that essences may be represented, at least implicitly, as dilutable and blendable.

2.8. Non-role social kinds

What about non-role social kinds, such as race and ethnicity? As is the case for sex, there is evidence for disposition to essentialize race and ethnicity (Banton, 1987; Hirschfeld, 1994, 1995, 1996; Gil-White, 2001). There has been considerable debate over why this should be the case, as it is generally agreed that racial and ethnic categories do not correspond to natural kinds. Atran (1990), Boyer (1990), and Rothbart and Taylor (1990), among others, have suggested that the essentialization of race may occur via cross-domain transfer of essentialist assumptions and inference procedures, from the domain of biological taxa to the domain of race. Gil-White (2001) has suggested that while the original proper domain of essentialism was indeed the domain of biological taxa, the use of essentializing mechanisms to represent and make inferences about ethnicity was favored by selection to such a degree that ethnicity has now become part of the proper domain of essentialism. Hirschfeld (1994, 1995, 1996; Gelman & Hirschfeld, 1999) has argued that the available data cannot be accounted for by the theory of cross-domain transfer. Instead, he argues that the kind of essentialist thinking typically applied to race and ethnicity is different in certain ways from that associated with biological taxa. For example, he argues that children rely more on social cues than perceptual differences between category members when learning racial categories than when learning biological taxonomic categories (Hirschfeld, 1993, 1994).

Gil-White (2001) has argued that classification of individuals into "ethnies", or ethnic groups, can indeed support useful inductive generalizations, because of the many culturally transmitted traits that tend to cluster within the boundaries of ethnic populations. For this reason, he suggests a two-stage evolutionary model in which ethnicity was originally represented using mechanisms evolved to handle biological taxa. Ethnic groups initially provided cues that accidentally activated essentializing mechanisms. This accidental activation, however, proved to have beneficial byproducts, according to Gil-White, because of the kind-like rich inductive potential of ethnic groups, which led to selection for ethnic groups to become part of the proper domain of essentializing mechanisms. Gil-White's (2001) theory hinges on the possibility that while races and ethnic groups are not true natural kinds, they may have inductive validity for cultural reasons. Gil-White's (2001) differs from Hirschfeld's (1994, 1995, 1996) in this regard, in that it offers an adaptationist explanation for the essentialization of ethnicity. An important question in distinguishing adaptation theories from byproduct / overextension theories of the essentialization of ethnicity is the degree to which ethnic groups support inductive generalizations, and whether the proposal that they are part of the proper domain of essentialism has any testable implications that can distinguish it from a byproduct / overextension model.

It may well be that the question of what is "essentialism's proper domain" will not prove to be a fruitful one if it turns out that essentialism, as such, is too broad a term to capture the subtle differences in thinking that people apply to different categories of thing, from substances, to biological taxa, to race and ethnicity. From an adaptationist perspective, the most interesting question is whether a particular

"kind" of essentialism shows evidence of special design for reasoning about a particular kind of entity. In the case of race and ethnicity, for example, if Gil-White's (2001) conjecture is true, and "ethnies" do have inductive validity because of real clusterings of culturally transmitted traits, then one might expect, if there really were an evolved disposition to essentialize ethnicity, that peoples' inductive generalizations would show adaptive constraint to the kinds of traits and properties that are culturally transmitted, and therefore inductively generalizable. Because these are precisely not the kinds of traits that are inductively generalizable for animal taxa, the domain-transfer theory (Atran, 1990; Boyer, 1990; Rothbart & Taylor, 1990) and the ethnicity-specific theory (Gil-White, 2001) of essentialism lead to different predictions with regard to the kinds of traits people should be observed to generalize. Only by examining how inductive generalization is constrained for these kinds of categories will we be able to decide between the alternative accounts.

3. Conclusion

The first section of this paper set out to review the implications of the adaptationist claim that psychological essentialism has an evolved function: namely, to aid inductive inference by constructing representations of certain kinds of entities that assume that these entities have richly interlocked sets of features by virtue of being members of particular natural kinds (Atran, 1994; Gelman *et al.*, 1994; Gelman & Diesendruck, 1999; Keil, 1989; Medin & Ortony, 1989). If true, this theory has a variety of implications for the cognitive design of representational systems. For example, there must exist mechanisms that pick out essentializable kinds in the world, that build appropriate representations of them, and that constrain inductive inferences to those sorts of properties that would be inductively generalizable for the kinds in question.

To see what the design of these mechanisms would look like, one must undertake an analysis of the kinds of entities that would have been their targets in ancestral environments, and of their properties. For example, mechanisms that identify essentializable classes of things in the world must contain implicit assumptions about what these kinds of things are like. It is unlikely that a single mechanism could serve to identify nonliving substances such as water and gold, partitioning them into separate essentialized kind categories, and at the same time pick out taxonomic categories of animals and plants, non-taxonomic biological categories such as predators, and even within-species categories such as sex. Moreover, it is unlikely that a single mechanism could serve to constrain inductions for all of these diverse categories of things, because there do not exist overarching, kind-general principles determining which classes of properties are validly generalizable within a kind. Mechanisms instantiating assumptions valid for different kinds of kind may therefore be functionally incompatible (Sherry & Schacter, 1987). While it may be true (and there is evidence suggesting that it is) that people essentialize everything from water to tigers to ethnic groups, the term "essentialism" may be too broad to encompass the subtleties of representation and inference that characterize these different cases.

Many accounts of essentialism assume that essentialism and natural kind status should go hand-in-hand (Gelman, 1988; Gelman & Markman, 1986, 1987; Keil, 1989; Kripke, 1972; Putnam, 1975). On this view, because gold and tigers are true natural kinds, they should both be essentialized, and because ethnicity is not a true natural kind, it should not. But if the evolved function of essentialism is to guide inductive generalization, then we expect the inductive properties of different classes of entities – not simply natural kind status *per se* – to determine both whether and how they are essentialized[17]. What matters is the causal processes that render inductive generalizations valid for a particular category of things, and these causal processes can vary enormously, from the laws of chemistry, to genetic reproduction, to cultural transmission.

The analysis presented here suggests that there should be a major cleavage between substance and whole organism kinds in how they are essentialized. Substances should be *shallowly* essentialized, because these kinds are not characterized by complex causal processes in which an executive causal agents leads to a cascading multiplicity of inducible properties. Substances should be treated as homogeneous, without whole-body properties that cease to apply when the object is partitioned. The kind of essentialism one expects for whole organisms, on the other hand, is a *deep* essentialism, which assumes 1) a multitude of properties caused by an executive causal agent or agents, 2) complex whole-body properties, including behavioral and psychological properties which, unlike the homogeneous properties of substances, do not hold for the constituent parts of the organism, and 3) functionality and purpose in many of these whole-body properties, especially behavioral and psychological ones. Moreover, we expect living kind representational systems to be able to handle 1) multiple, overlapping and / or nested essentialized kind categories, such that a single individual can simultaneously belong to multiple categories (e.g., lion, mammal, predator, female, all at the same time), and 2) non-atomic essences, e.g., essences that can be graded, diluted, and combined.

There is a substantial difference between the proposal that essentialism happens to be a useful strategy and the proposal that it was designed to be useful by natural selection. From an evolutionary point of view, simply stating that essentialism and natural kinds go together is insufficient; one must first develop a proposal about the kinds of kinds we have been selected to be essentialist about, and then examine the inductive structure of these kinds to see what kinds of assumptions would have to be built into an essentialist architecture. Doing so may help us to understand how our thinking even now is guided by mechanisms that evolved to aid inference in a world in which formal science did not yet exist.

[17] As pointed out above, the cue structure of an entity should matter as well. Entities may "mimic" a particular kind by presenting certain cues that trigger a kind-specific, essentialist mode of construal, even when they are not truly kind members.

References

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

Atran, S. (1994) Core domains versus scientific theories, in L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (Cambridge, Cambridge University Press).

Atran, S. (1998) Folk biology and the anthropology of science, *Behavioral and Brain Sciences,* 21, pp. 547-611.

Au, K. T. (1994) Developing an intuitive understanding of substance kinds, *Cognitive Psychology,* 27, pp. 7-111.

Banton, M. (1987) *Racial theories* (Cambridge, Cambridge University Press).

Barrett, H.C., Cosmides, L, & Tooby, J. (in prep.). By descent or by design? Evidence for two modes of biological reasoning.

Berlin, B. (1992) *Ethnobiological classification* (Princeton, Princeton University Press).

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

Boyer, P. (1990) *Tradition as truth and communication* (New York, Cambridge University Press).

Boyer, P. (1994) *The naturalness of religious ideas: A cognitive theory of religion* (Berkeley, University of California Press).

Braisby, N., Franks, B., & Hampton, J. (1996) Essentialism, word use, and concepts, *Cognition,* 59, pp. 247-274.

Caramazza, A., Hillis, A., Leek, E.C., & Miozzo, M. (1994) The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits, in L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (Cambridge, Cambridge University Press).

Carey, S. (1985) *Conceptual change in children* (Cambridge, MIT Press).

Carey, S. (1995) On the origins of causal understanding, in D. Sperber, D. Premack, & A. Premack, (Eds.), *Causal cognition: A multidisciplinary debate* (Oxford, Oxford University Press).

Coley, J.D., Medin, D.L., & Atran, S. (1997) Does privilege have its rank? Inductive inferences within folkbiological taxonomies, *Cognition,* 63, pp. 73-112.

Coley, J.D., Medin, D.L., Proffitt, J.B., Lynch, E., Atran, S. (1999) Inductive reasoning in folkbiological thought, in D.L. Medin & S. Atran, (Eds.), *Folkbiology* (Cambridge, MA, MIT Press).

Dawkins, R. (1982) *The extended phenotype* (Oxford, Oxford University Press).

Dennett, D.C. (1987) *The intentional stance* (Cambridge, MIT Press).

Dickinson, D.K. (1987) The development of a concept of material kind, *Science Education,* 71, pp. 615-628.

Dupré, J. (1993) *The disorder of things: Metaphysical foundations of the disunity of science* (Cambridge, Harvard University Press).

Fuss, D. (1989) *Essentially speaking: Feminism, nature, and difference* (New York, Routledge).

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

Gelman, S.A. (1988) The development of induction within natural kind and artifact categories, *Cognitive Psychology,* 20, pp. 65-95.

Gelman, S.A. (1992) Children's conception of personality traits (commentary), *Human Development,* 35, pp. 280-285.

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

Gelman, S.A., & Coley, J.D. (1990) The importance of knowing a dodo is a bird: Categories and inferences in two-year olds, *Developmental psychology,* 26, pp. 796-804.

Gelman, S.A., Coley, J.D., & Gottfried, G.M. (1994) Essentialist beliefs in children: The acquisition of concepts and theories, in L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (Cambridge, Cambridge University Press).

Gelman, S.A., Collman, P., & Maccoby, E. (1986) Inferring properties from categories versus inferring categories from properties: The case of gender, *Child Development*, 62, pp. 396-414.

Gelman, S.A., & Diesendruck, G. (1999) A reconsideration of concepts: On the compatibility of psychological essentialism and context sensitivity, in E.K. Scholnick, K Nelson, S.A. Gelman, & P.H. Miller (Eds.), *Conceptual development: Piaget's legacy* (Mahwah, New Jersey, Lawrence Erlbaum Associates).

Gelman, S.A., & Hirschfeld, L.A. (1999) How biological is essentialism? in D.L. Medin & S. Atran, (Eds.), *Folkbiology* (Cambridge, MA, MIT Press).

Gelman, S.A., & Markman, E.M. (1986) Categories and induction in young children, *Cognition*, 23, pp. 183-209.

Gelman, S.A., & Markman, E.M. (1987) Young children's inductions from natural kinds: The role of categories and appearances, *Child Development*, 8, pp. 157-167.

Gelman, S.A., & Spelke, E.S. (1981) The development of thought about animate and inanimate objects: Implications for research on social cognition, in. J.H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (Cambridge, Cambridge University Press).

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

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995) Taking the intentional stance at 12 months of age, *Cognition*, 56, pp. 165-193.

Gigerenzer, G. (2000) *Adaptive Thinking* (Oxford, Oxford University Press).

Gil-White, F. J. (2001) Are ethnic groups 'species'to the human brain? Essentialism in our cognition of some social categories, in press, *Current Anthropology*, 42.

Goodman, N. (1954) *Fact fiction and forecast* (Cambridge, MA, Harvard University Press).

Hirschfeld, L.A. (1986) Kinship and cognition: Genealogy and the meaning of kinship terms, *Current Anthropology*, 27, pp. 217-242.

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

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

Hirschfeld, L. A. (1995) Do children have a theory of race? *Cognition*, 54, pp. 209-252.

Hirschfeld, L.A. (1996) *Race in the making: Cognition, culture, and the child's construction of human kinds* (Cambridge, MIT Press).

Kalish, C. (1995) Essentialism and graded membership in animal and artifact categories, *Memory and Cognition*, 23, pp. 335-353.

Keil, F.C. (1989) *Concepts, kinds, and cognitive development* (Cambridge, MIT Press).

Keil, F.C. (1994) The birth and nurturance of concepts by domains: The origins of concepts of living things, in L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (Cambridge, Cambridge University Press).

Keil, F.C. (1995) The growth of causal understandings of natural kinds, in D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (Oxford, Oxford University Press).

Keil, F.C., Smith, W.C., Simons, D.J., Levin, D.T. (1998) Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge, *Cognition*, 65, pp. 103-135.

Keleman, D. (1999) Function, goals, and intention: Children's teleological reasoning about objects, *Trends in Cognitive Sciences*, 3, pp. 461-468.

Kripke S. (1972) *Naming and necessity* (Cambridge, MA, Harvard University Press).

Leslie, A.M. (1994) ToMM, ToBy, and Agency: Core architecture and domain specificity, in L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (Cambridge, Cambridge University Press).

Locke, J. (1689) *An essay concerning human understanding* (6th Ed., ILT Digital Classics, 1995).

MacCoby, E.E., & Jacklin, C.N. (1974) *The psychology of sex differences* (Stanford, Stanford University Press).

Malt, B.C. (1994) Water is not H_2O, *Cognitive Psychology,* 27, pp. 41-70.

Mandler, J.M. (1992) How to build a baby: II. Conceptual primitives, *Psychological Review,* 99, pp. 587-604.

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

Mandler, J.M., & McDonough, L. (1998) Studies in inductive inference in infancy, *Cognitive Psychology,* 37, pp. 60-96.

Markman, E.M. (1989) *Categorization and naming in children: Problems of induction* (Cambridge, MA, MIT Press).

Matan, A., and Carey, S. (2001) Developmental changes within the core of artifact concepts, *Cognition,* 78, pp. 1-26.

Mayr, E. (1982) *The growth of biological thought* (Cambridge, MA, Harvard University Press).

Medin, D.L. (1989) Concepts and conceptual structure, *American Psychologist,* 44, pp. 1469-1481.

Medin, D.L., & Ortony, A. (1989) Psychological essentialism, in S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (Cambridge, Cambridge University Press).

Millikan, R.G. (1984) *Language, thought, and other biological categories* (Cambridge, MIT Press).

Millikan, R.G. (1998) A common structure for concepts of individuals, stuffs, and real kinds: More mama, more milk, and more mouse, *Behavioral and Brain Sciences,* 21, pp. 55-100.

Millikan, R.G. (2000) *On clear and confused ideas: An essay about substance concepts* (Cambridge, Cambridge University Press).

Piaget, J. (1951) *The child's conception of the world* (London, Routledge Kegan Paul).

Putnam, H. (1975) The meaning of "meaning", in H. Putnam (Ed.), *Mind, language and reality: Philosophical papers,* vol. 2 (New York, Cambridge University Press).

Quine, W.V.O. (1977) Natural kinds, in S.P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (Ithaca, NY, Cornell University Press).

Rosch, E., & Mervis, C.B. (1975) Family resemblances: Studies in the internal structure of categories, *Cognitive Psychology,* 7, pp. 573-605.

Rothbart, M., & Taylor, M. (1990) Category labels and social reality: Do we view social categories as natural kinds? in G. Semin & K. Fiedler (Eds.), *Language and social cognition* (London, Sage).

Rozin, P. Millman, L., & Nemeroff, C. (1986) Operation of the laws of sympathetic magic in disgust and other domains, *Journal of Personality and Social Psychology,* 50, pp. 703-712.

Sherry, D.F., & Schacter, D.L. (1987) The evolution of multiple memory systems, *Psychological Review,* 94, pp. 439-454.

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

Smith, C., Carey, S., & Wiser, M. (1985) On differentiation: A case study of the development of the concepts of size, weight, and density, *Cognition,* 21, pp. 177-237.

Smith, E.E., & Medin, D.L. (1981) *Categories and concepts* (Cambridge, MA, Harvard University Press).

Sober, E. (1988) *Reconstructing the past* (Cambridge, MA, MIT Press).

Sperber, D. (1994) The modularity of thought and the epidemiology of representations, in L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (Cambridge, Cambridge University Press).

Springer, K. (1992) Children's awareness of the biological implications of kinship, *Child Development,* 63, pp. 950-959.

Springer, K. (1995) Acquiring a naïve theory of kinship through inference, *Child Development,* 66, pp. 547-558.

Springer, K. (1996) Young children's understanding of a biological basis for parent-offspring relations, *Child Development,* 67, pp. 2841-2856.

Strevens, M. (2000) The essentialist aspect of naïve theories, *Cognition*, 74, pp. 149-175.
Symons, D. (1979) *The evolution of human sexuality* (Oxford, Oxford University Press).
Taylor, M. (1996) The development of children's beliefs about social and biological aspects of gender differences, *Child Development*, 67, pp. 1555-1571.
Trivers, R.L. (1972) Parental investment and sexual selection, in B. Campbell, (Ed.), *Sexual selection and the descent of man, 1871-1971* (Chicago, Aldine).
Yuill, N. (1992) Children's conception of personality traits, *Human Development*, 35, pp. 265-279.

Acknowledgements

Thanks to Jennifer Davis, Larry Fiddick, Peter Todd, Francisco Gil-White, Rob Kurzban, and Ruth Millikan, who discussed many of the ideas presented here, and/or read prior drafts. Preparation of this paper was supported by a Postdoctoral Research Fellowship from the Max Planck Society.