

**GOLD, JADE, AND EMERUBY: THE VALUE OF NATURALNESS FOR
THEORIES OF CONCEPTS AND CATEGORIES**

Charles Kalish
University of Wisconsin-Madison

In press, Journal of Theoretical and Philosophical Psychology

Correspondence to:

Charles Kalish Email: cwkalish@facstaff.wisc.edu

Until 6/30/02

Department of Psychology Ph: (212) 998-3794

New York University

6 Washington Place, room 408

New York, NY 10003-6634

After 6/30/02

University of Wisconsin Ph: (608) 262-0840

1025 W. Johnson St. Fax: (608) 262-0843

Madison, WI 53706-1796

ABSTRACT

Researchers studying the psychology of concepts frequently draw distinctions between artificial and natural concepts. Unfortunately, there is a lack of consensus regarding the foundations and implications of the distinction. This paper provides a review and evaluation of the different ways researchers have approached the question of conceptual naturalness. Accounts may be divided into two approaches described as psychologically or externally based. These characterizations motivate distinctive sets of research questions. In addition to the particular implications, I also consider the general significance of a distinction between natural and artificial concepts.

Gold, Jade, And Emeruby: The Value Of Naturalness For Theories Of Concepts And Categories

Under the rubric of post-modernism it is possible to read all manner of challenges to commonsense intuitions about the natural order of things. Whether focusing on particular distinctions we make (such as racial or gender groupings), or on our entire system of beliefs about the kinds of things there are in the world, the argument often goes that we are misguided in thinking that our concepts are products or reflections of an objective nature (e.g., Fish, 1996; Rorty, 1999, cf. Hacking, 1999). There is nothing natural or inevitable about any particular system of concepts. This question seems a basic issue for scientific research on concepts. One of the goals of Psychology and Cognitive Science is to answer the questions of which (if any) concepts are natural and why. Unfortunately, researchers use the designations “natural” and “artificial” in different ways. It seems that the field has no unequivocal answer to the question of which concepts are natural. As a consequence there is no consensus regarding the significance of the distinction for theories of concepts and categories. The goal of this essay is to review the different senses in which concepts may be deemed natural and to draw out the research and theoretical implications of different senses of naturalness.

Early researchers often employed meaningless stimuli in concept learning experiments (e.g., Hull, 1920). A typical experiment required participants to learn a concept such as SMALL OR BLACK (e.g., Goodnow, Bruner, & Austin, 1957). Although these procedures were occasionally criticized as artificial and non-representative of the actual process of concept acquisition (Kakise, 1911; Smoke, 1932) the assumption was that the concepts derived did not differ in meaningful ways from concepts derived outside the experimental setting. In the 1970's Rosch (1973, 1978) posed two, linked, challenges to this view. First, she argued the materials were unrepresentative of naturally occurring objects. Second, the psychological processes involved in categorizing such materials were not the ones typically used in forming concepts.

Rosch's theory involves claims about psychological naturalness and external naturalness. The human mind is disposed toward concepts with a family resemblance structure; construction and use of such concepts is natural. But such dispositions are no accident. Objects in the world actually do form clusters around correlated attributes; the groups are natural. In Rosch's theory it is possible to discern two senses of conceptual naturalness as well as a hypothesis about their

interrelation. Moreover, the research implications also seem clear. It is natural concepts that should be the focus of research. The psychology of concepts is advanced by studies of how people form and use natural concepts, such as BIRD, rather than by studies of artificial concepts, such as SMALL OR BLACK. In the remainder of this review I will discuss additional psychological and external accounts of naturalism as well as further suggestions about their interrelations and the implications for the psychological study of concepts.

PSYCHOLOGICAL CRITERIA FOR CONCEPTUAL NATURALNESS

Much psychological work begins with the intuition that some concepts are more psychologically natural than others. The idea is that we tend to form some concepts rather than others. For example, Osherson (1978) characterizes natural concepts as those that are acquired during the course of normal experiences. On this perspective, concepts are orderings imposed on experience by the mind. Cognitive principles determine which concepts are formed (how experience is mapped onto concepts). Those principles define the set of concepts natural within a given cognitive system. The operation of the principles of conceptualization will have consequences indicative of naturalness, for example some concepts will be more easily learned than others. Ease of acquisition may be only one of a number of effects of naturalness. A theory of concepts will describe the set of cognitive principles responsible for those effects, those processes that make some concepts more natural than others. Within the psychological literature a number of such principles have been proposed.

Structural constraints

The strongest psychological formulations of naturalness involve claims that there are structural constraints on the types of concepts which may be represented. Natural concepts have some formal properties that artificial ones lack. Structural constraints are generally thought to derive from the format with which concepts are represented. Thus if concepts are represented in a propositional form, artificial concepts would include those involving logical contradictions (e.g., COLORLESS GREEN STONE). Similarly, the representation of spatial knowledge may be incommensurate with some concepts (e.g., 6 DIMENSIONAL CUBE). A specific proposal along these lines is the suggestion that natural concepts fit into a hierarchical organization of ontological

types (Keil, 1979, 1981; Osherson, 1978). A consequence of this model is that the predicates applicable to a concept must apply to all concepts below it in the hierarchy (its daughters). If weight is a predicate of OBJECT, then it is also a predicate of all concepts below OBJECT (kinds of OBJECT such as LIVING THING). Concepts involving cycles in the ontological tree structure are unnatural. For example, objects and waves seem to be ontologically different sorts of concepts. Concepts of objects accept predicates of mass and weight but not of frequency. Concepts of waves accept predicates of frequency but not mass. A concept for which both sorts of predicates were appropriate (e.g., ELECTRON) would be unnatural. Either all objects have a frequency or none do (see Carey, 1986 for other counter-examples). Representational format constraints are often involved in developmental claims. For example, the concepts formed by young children may be limited to those with structures that are: complexive (vs definitional, Bruner, et al., 1956; Keil & Batterman, 1984; Vygotsky, 1962), wholistic (vs. analytic, L. Smith, 1989), or organized in a single level (vs. taxonomically organized, see Johnson, Scott, & Mervis, 1997). In the developmental cases the constraints are often (but not always) described as absolute; it is impossible for children to form unnatural concepts. In contrast, adults do seem able to construct unnatural concepts (at least those of the sorts given as examples above).

That people use and form unnatural concepts is a challenge to theories of structural constraints. One way to meet the challenge is to posit multiple representational systems with different structural constraints. For example, mathematical or linguistic representational systems may generate concepts violating natural principles. A capacity to think mathematically may allow formation of concepts that could not be represented otherwise. Such 'externally' generated concepts would always retain their unnatural character, perhaps never completely integrated with other, natural, concepts (unless representational formats change, see Carey & Spelke, 1993). This line of reasoning raises the possibility that people may have multiple sets of incommensurable concepts. Concepts violating natural principles need not be completely excluded from cognition. Nonetheless, this relatively strong construal of naturalness does suggest that unnatural concepts will be rare and formed only under exceptional circumstances (see Sperber, 1990; Osherson, 1978). Note that the converse does not necessarily hold: There are likely many concepts satisfying structural constraints on naturalness that are not commonly acquired in the course of experience

(e.g., EMERUBY, an emerald examined before time T or a ruby thereafter, Goodman, 1955).

Thus structural constraints may provide necessary but not sufficient criteria for naturalness.

Processing principles

An alternative perspective on psychological naturalness characterizes the property in terms of optimality rather than as accordance with constraints. A concept either does or does not violate some structural constraint. In contrast, many psychological characterizations present naturalness as a matter of degree; concepts are more or less natural relative to each other. In such accounts naturalness is a consequence of the operation of psychological processes of concept formation and use. For example, one straightforward determinant of naturalness may be familiarity. The more frequently people use a concept to make inferences, the more natural it seems (see Goodman, 1955). Thus while any concept involving a feature such as NON-RED would seem artificial, familiarity could lead people to accept the concept SAPPHIRE (defined as NON-RED CORUNDUM). It is also possible to reconceptualize the structural constraints described above as matters of degree or preference. For example, some cycles in the ontological tree may be less problematic than others (Kelly & Keil, 1985). In practice it may be difficult to distinguish real structural constraints from strong preferences (see Kalish, 1998; Keil, 1981).

Psychological research has produced a number of candidates for fundamental processes of conceptualization. A longstanding proposal has been that the most natural cases of concept construction and representation involve computations of similarity (for reviews see Medin, 1989; Smith & Medin, 1981). There continues to be debate, however, regarding the relative roles of similarity and more rule-like processes for concept formation and use (Hampton, 1998; Rips, 1989; Smith & Sloman, 1994). Within similarity-based accounts many researchers suggest that exemplar-based representations of concepts are most natural, while others favor a process of prototype abstraction (Medin & Shaffer, 1978; Nosofsky, 1992; J. Smith & Minda, 1998). Other factors that have been thought to make one concept relatively more natural than another include: extensional simplicity (Osherson, 1978), salience of discriminant features, directness or simplicity of linguistic expression (one formulation of the Sapir/Whorf hypothesis), basis in perceptual rather than abstract/theoretical features (Tversky, 1985), and conjunctive rather than disjunctive

combinations of features (Bourne, 1974). The suggestion is that in the typical course of events people will form concepts with one of these structures rather than (or more easily than) the other.

As naturalness is a matter of degree it is expected that people's repertoire of concepts will include a range of values on these attributes. For example, the theory that concepts are formed by a process of prototype abstraction suggests that linearly separable concepts (those based on additive combinations of features) will be more natural than non-linearly separable concepts (see Wattenmaker, Dewey, Murphy, & Medin, 1986). It is not an embarrassment to this theory that people have and use non-linearly separable concepts. Rather the question is whether linearly separable concepts are preferred (e.g., learned more easily, Wattenmaker, et al., 1986). Rosch's proposal that basic level concepts are the most natural is another example (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). People do form concepts at varying levels of generality, with greater or lesser within group similarity and between group distinctiveness. Because of this variability, researchers studying principles of concept formation and use do not typically make a strong distinction between natural and artificial concepts. Rather it is the principles or processes which may be natural or not.

Assuming that some alternative structures or processing of concepts have been identified, why should one alternative be deemed the more natural one? Naturalness seems linked to the conditions under which various conceptualization processes are activated. Psychometric data may be used to ground claims of naturalness. For example Smith and Sloman (1994) argued that similarity-based processes are relatively faster and more automatic than rule-based processes (see also D. Nelson, 1984; Smith, Patalano, & Jonides, 1998). Different ways of forming concepts may be favored in linguistic and non-linguistic contexts (Markman & Hutchinson, 1984). Nonlinguistically mediated conceptual processes are plausibly more natural than verbal processes (Vygotsky, 1962). Finally some concept formation principles are developmentally prior to others, both phylogenetically and ontogenetically. These, and other, signals of naturalness may all identify a common set of principles (roughly similarity relations among perceptual features), but they may not. There is no guarantee that the measures will converge rather than diverge (see Keil, Carter Smith, Simons, & Levin, 1998, and discussion below). Even if a common set of principles is arrived at, it may be more apt to characterize these as the most primitive or basic, rather than

most natural. Some psychological processes are the earliest developing, least demanding of cognitive resources, etc. However, people are not normally limited to forming or using concepts organized by the most basic principles. The natural conditions that humans live and develop under clearly support the acquisition of a large number of non-primitive concepts. It may be that the most common circumstances of concepts acquisition and use call upon non-basic processes. At minimum, the diversity of human concepts suggests that the criteria for designating one processing principle as more natural than another require some careful justification. Which processes of concept formation are really the most natural ones?

Inferential Role

A third type of psychological account defines naturalness in terms of a concept's inferential role. Concepts are embedded within larger cognitive structures, often described as theories (Murphy & Medin, 1985; Wellman & Gelman, 1993, also Lewis, 1970). Natural concepts, usually referred to as 'natural kinds,' are those that are particularly central or fundamental within a theory (Fodor, 1975). Natural kind concepts are organized around theoretically important properties; members of the same natural kind share patterns of causal interaction with other objects. This notion of natural kind is similar to that found in the philosophical literature (Boyd, 1991; Reznick, 1987) with the caveat that naturalness is a function of beliefs about the structure of the world rather than a direct function of that structure. Theory-based conceptual naturalness is thus distinct from external or content-based construals of naturalness discussed below.

Markman (1990; also Gelman, 1988; Keil, 1989) describes several characteristics of natural kind concepts including: rich inductive potential (members of natural kinds are assumed to resemble each other in unknown as well as known ways), centrality to identity (a thing's natural kind membership tells us what it is, rather than just what it is like, and thus cannot change), essential structure (some underlying quality or property determines membership and more characteristic features), and mutual exclusivity (at least at similar levels of abstractness an individual will be a member of only one natural kind concept). Other characteristics of natural kind concepts, derived from their central role in theories, include shared patterns of change or development (e.g., growth, decay) and possibility of anomalous instances (where deeper, theoretical relations contrast with, and outweigh, less central similarities; Keil, 1989). Because

natural kind concepts are central to theories about the structure of the world, they will be perceived as relatively objective matters of fact, rather than as invented or conventional groupings (Kalish, 1998; Malt, 1990). Although often presented as common consequences of theoretical significance, a given concept may have more or fewer of these attributes, or have them to a greater or lesser degree. There is a continuum from extremely well entrenched and motivated concepts to more peripheral and arbitrary ones. It also remains an open question how well theoretical role predicts each of the qualities of natural kind concepts (e.g., essential structure, inductive potential).

Relations between Psychological accounts of naturalness

What are the relations between the three sorts of psychological criteria for naturalness discussed above? One possibility is that they are hierarchically organized. Structural constraints might pick out the broadest class of natural concepts. Within those well-formed concepts, some will have more optimal values along processing dimensions (e.g., better cue/category validity, higher feature salience). Finally, among those concepts consistent with processing principles, some will be relatively central to intuitive theories. Thus the broadest characterization identifies concepts such as COLORLESS GREEN STONE as non-natural while admitting GOLD but also concepts such as EMERUBY or SAPPHIRE (NON-RED CORUNDUM) as natural. Processing principles might rule out EMERUBY and SAPPHIRE while identifying concepts such as GOLD, JEWELRY, and ICICLE as natural. Intuitive theories will further refine the set of natural concepts by including those with rich causal and inferential relations, such as GOLD while excluding more peripheral concepts such as JEWELRY and ICICLE.

Although a hierarchical arrangement may be plausible for many cases, the relations between the three psychological characterizations of naturalness are often more complex. Carey (1986) has argued that the concepts central to theories need not be those identified as natural on formal or processing principles. Rather theoretical considerations may 'naturalize' a concept that would be unnatural on other considerations (e.g., make some features or distinctions more salient, Gelman & Kalish, 1993; Keil, 1989). Thus some natural kind concepts such as INVERTEBRATE or ACID appear to be poor concepts in the absence of a particular theory. It is unclear, however, how flexibly theories may create psychologically natural kinds. Carey and Spelke (1993) consider the possibility that structural or representational constraints may prevent some theoretically defined

natural kind concepts from becoming truly part of people's conceptual systems. Based on data from judgments of concept objectivity, Kalish (1998) proposed a threshold model of naturalness. If a concept is sufficiently well structured (e.g., salient, familiar) it will be natural even if it has only a weak underlying theoretical rationale. However, theoretical considerations may naturalize otherwise poorly structured concepts. Naturalizing influences would also seem to run in the other direction. That a concept is natural on formal or processing principles may promote or entrench its significance within intuitive theories (Keil, 1989; Medin, 1989; Gelman & Medin, 1993).

EXTERNAL CRITERIA FOR CONCEPTUAL NATURALNESS

An alternative to defining naturalness in terms of psychological qualities is to identify natural concepts by their referents or extensions. Naturalness would designate natural occurrence, as opposed to construction by people. One proposal along these lines is that natural concepts are concepts of naturally occurring objects. An alternative is to designate as natural concepts of objectively existing classes or groupings of objects. A third sort of external construal of naturalness involves the ways semantic properties of concepts are determined. Natural concepts are those with externalist semantics. In each of these senses, naturalness is not a psychological property; it is facts about the world, not facts about minds, that makes a concept natural. Thus the psychological significance of conceptual naturalness is an empirical question.

Concepts of naturally occurring objects

Many studies in the psychological literature distinguish between concepts of naturally occurring and artificially constructed objects. Often, concepts of naturally occurring objects are referred to as 'natural kinds' (Gelman, 1988; McGill, 2000; Smith, 1995). The discussion of category-specific deficits is also framed in terms of a distinction between concepts of naturally occurring and humanly constructed objects (see Forde & Humphreys, 1999, for review). Concepts of animals are the most commonly discussed examples of natural kind concepts (e.g., Barton & Komatsu, 1989; Diesendruck & Gelman, 1999; Kalish, 1995; 1998; Gelman & Markman, 1986). Some research has included concepts of plants and chemical elements (e.g., Braisby & Franks, 1996; Gelman, 1988; Malt, 1990). Many naturally occurring things, such as planets or geographical formations (e.g., LAKE, MOUNTAIN, see Kalish, in press), are typically

omitted from sets of natural kind concepts. Concepts of artifacts are somewhat more widely sampled, frequently including concepts of weapons, vehicles, tools, and clothing (see Bloom, 1996). However, there are many kinds of human made objects that do not appear in sets of stimuli, for example, scientific instruments (e.g., CYCLOTRON, though see Kalish, in press) or social institutions (e.g., SCHOOL). In the literature on natural and artificial concepts (indeed, in most research on concepts) there is a focus on tangible objects, rather than processes or events. However natural and artificial concepts should not be limited to concepts of objects based on the origins of their instances.

A fundamental problem with the strategy of identifying natural and artificial concepts by the origins of their instances is that the same object may be an instance of multiple concepts. A single individual may be a DIAMOND, a JEWEL, and an ASSET. One reaction to this flexibility in conceptualization is to admit psychological characterizations of naturalness. For example, it may be argued that the most frequent, basic, or automatic way of thinking about a diamond is as a DIAMOND. Yet, this characterization of the naturalness of DIAMOND relies not just on the origins of its instances, but also on the psychology of concepts formation and usage. The same is true of stipulations that natural concepts be limited to those involving intrinsic properties, taxonomic organization, or presumed biological basis (e.g., Atran, 1987). An alternative is see naturalness not as dependent on the origins of the instances, but on the origins of the class or kind referenced by the concept. Natural concepts are not concepts of naturally occurring objects, but rather are concepts of naturally occurring kinds.

Concepts of natural classes

In the philosophical literature, 'natural kind,' is often used to refer to a grouping or distinction that is objective rather than invented by people (Kripke, 1972; Putnam, 1975; Reznick, 1987, Schwartz, 1979). Thus we discover that certain kinds of things exist in the world (e.g., that diamonds are distinct from quartz). The naturalness of the class or kind is independent of the origins of its instances. DIAMOND could be natural even if all diamonds had been grown in laboratories. One approach is to rely on scientific theories to provide objective identifications of natural kinds. As scientists describe the structure of reality they will determine which kinds or groupings of entities have a basis in nature and which are 'illusions' or inventions of the human

mind. Of course, whether science will provide an unequivocal and purely objective picture of the world is a matter of debate. Nonetheless, natural kind concepts may still be identified as those concepts which turn out to match the kinds identified by science.

The match between concepts and scientifically-determined classes is best thought of as occurring in extensions. The class of objects included in a natural kind concept is the same as a class identified as a natural kind. The intension, the basis for determining inclusion, would likely differ between scientists and lay people. Thus it is because people's concept of DIAMOND refers to diamonds that the concept is natural, not because lay people know the scientific basis of the kind. Yet this distinction raises the question of how extensions of concepts might be determined. Recent philosophical accounts of concepts have challenged existing psychological theories of concepts on just this issue. Importantly for the purposes of this discussion, such philosophical theories may be understood as involving a distinction between two different ways of fixing reference and identifying concepts. These different views, externalist and internalist characterizations, may imply a distinction between natural and artificial concepts.

Concepts with externalist semantics

Recently a number of philosophers have argued that existing psychological accounts of concepts are fundamentally flawed (Fodor, 1998; Margolis, 1998; Millikan, 1998). Either psychologists have been borrowing from discredited philosophical theories (e.g., inferential role semantics) or have been mischaracterizing the philosophy. For whatever reason, psychologists have limited their investigations to a subset of possible accounts of concepts. While psychologists have pursued descriptivist theories, many philosophers favor nondescriptivist or externalist theories of concepts. Because externalist approaches to concepts may be unfamiliar, and because the implications of such philosophical theories for psychological research are not always clear, I will present such accounts in some detail in this section. Viewing externalism through the lens of natural and artificial concepts may clarify the significance of this approach.

Despite their apparent diversity, all psychological theories share a commitment to descriptivism (Margolis, 1998). Concepts are constituted from mental representations of properties or attributes (e.g., clusters of features or inferential connections). Thus it is people's beliefs that give identity to particular concepts; a concept is GOLD (rather than LEAD) because it is

represented using the features DUCTILE, FOUND IN WEDDING RINGS, and ATOMIC NUMBER 79. In contrast nondescriptionalist or externalist theories (Fodor, 1998; Millikan, 1998) identify a concept based on its informational or causal connections to the external world. Thus what makes a concept GOLD (rather than LEAD) is, in part, that the concept is activated (tokened) in the presence of gold. The mental representation of the concept is irrelevant to its identity (except in supporting the connection to the referent, see below). This is why it is possible for people who don't think that GOLD is ductile, in wedding rings, or element 79 (e.g., children) to nonetheless possess the concept GOLD; all they need is some representation reliably connected to the gold by other means. GOLD does not provide a description (or definition), but rather provides a means of re-identifying instances and accumulating facts about the gold. The intuition that people with different representations can nonetheless have the same concepts (and that people with the same representations could have different concepts) are compelling reasons to support nondescriptionalist theories, at least as accounts of meaning and reference.¹

The implications of externalism for psychological work on concepts are a matter of debate (see commentaries following Millikan, 1998). For this review it is significant that externalist theories of concepts make characteristic use of a distinction which may be glossed as natural versus artificial (Margolis, 1998; Millikan, 1998). This sense of naturalness stems from the observation that descriptionalism is true for some concepts. Externalism may be taken as the claim that nondescriptionalist concepts are natural, while artificial concepts are descriptionalist. Such a distinction is similar to (at least some interpretations) of Locke's (1661/1707) contrast between concepts of substance and mixed modes. Artificial concepts are combinations of primitives (other concepts or features). Natural concepts are those that track or index a real kind of thing that exists in the world (a substance).

Consider the acquisition of an artificial concept, such as FOO (Millikan, 1998). What makes the concept artificial, from a nondescriptionalist perspective, is that what is acquired is a particular description. An expert or teacher decided that FOO means BLUE OR SQUARE. Somehow the concept learner reproduces that description; he or she infers which combination of attributes satisfies the teacher's criteria. What is learned is that it is blueness or squareness that constitutes FOO. This could be the way we naturally go about acquiring concepts. However, the

nondescriptionalist position seems to imply that this is not the case. For example, Millikan (1998) suggests that such a process does not characterize the first concepts we acquire nor, seemingly, the bulk of our conceptual repertoire.

Natural concept learning may tend to produce nondescriptionalist concepts. In acquiring a concept, people are not attempting to discern the criteria upon which some expert (experimenter, parent) determines the extension of a concept. Instead they are attempting to identify the real kind or substance referenced by the concept (Millikan, 1998). Part of learning a concept is figuring out those features diagnostic of category membership (characteristic features, Keil, 1989; Medin & Ortony, 1989). However, such features are understood to be only more or less useful heuristics for identifying examples. Concepts with this type of structure are often described as essentialist (Medin & Ortony, 1989; Wellman & Gelman, 1993) because they involve the belief that there is some underlying structure in the world (the essence) responsible for the manifestation of the characteristic features. It is because concepts are supposed to capture or identify substances that they are amenable to an externalist semantics (Margolis, 1998; Millikan, 1998). It is because concepts are taken to track substances in the world that facts about the nature of those substances determine the content of the concepts.

Thus externalist semantics involves an important psychological distinction between natural and artificial concepts. Natural concepts are attempts to track substance and acquire more information about them. Artificial, or nominal, concepts are attempts to accurately represent distinctions or definitions represented by others. Consider the difference between RUBY and JADE. At some point it was discovered that two different types of minerals were being identified as RUBY. If RUBY is taken as tracking a substance, this implies there have been some mistakes; some things we thought were RUBY are not. Alternatively, if RUBY is taken as a description (e.g., CLEAR HARD RED GEM) the interpretation is instead that there are two chemical substances which may be RUBY. For RUBY people settled on the first interpretation (and some pretenders, including the Black Prince's Ruby in the royal crown of England, were 'unmasked' as SPINEL). However, in much the same circumstances there was a different outcome for JADE. Instances of JADE may be either NEPHRITE or JADEITE, but all are JADE because they satisfy the description DENSE GREENISH TRANSLUCENT STONE.

Before considering the implication of accounts of naturalness for research on the psychology of concepts, it is worth noting that the relation between conceptual structure and externalist semantics is somewhat more complicated. There is a large middle ground between concepts taken to refer to real, fundamental substances or kinds in the world (e.g., GOLD, DOG) and those taken to be arbitrary constructions of properties (e.g., FOO). Concepts of human artifacts (e.g., CHAIR, AIRPLANE), social categories (e.g., DOCTOR, FRENCH), and non-scientific kinds (e.g., JEWELRY, ICICLE) seem neither completely objective nor simply conventional descriptions. Millikan includes these intermediate types as substances (real kinds). They meet the criteria of providing the ability to reidentify or track instances and accumulate knowledge. One way of reading Millikan is as offering a developmental hypothesis; these concepts are first acquired as substances but become descriptive 'when confidence is lost in the reality...or univocity' (Millikan, 1998, p.93). Such a position is similar to the suggestion that young children are biased to treat concepts as indicating natural kinds (Gelman & Kalish, 1993) and only later come to appreciate more arbitrary groupings.

An alternative perspective is to see substancehood (and also externalism?) as a matter of degree. This is the approach that Millikan seems to favor. GOLD is a relatively good substance useful in tracking the structure of the world; JEWELRY is a relatively poor substance primarily useful as a way of classifying things. But how do descriptive (serving classification) and external (serving identification) aspects combine? At least in some cases it seems that descriptions may be constitutive of substance. It is in virtue of satisfying a description that something becomes JEWELRY and, thereby, acquires properties that one might learn about and exploit. Thus the function of the concept JEWELRY is to successfully track or characterize a description. Among the best examples of such 'artificial substances' are social categories (cf. Hacking, 1995; Searle, 1998). In virtue of satisfying the description of TEACHER (or GEMOLOGIST) a person acquires a host of properties (privileges, responsibilities). Thus one may learn new ways to (fallibly) identify teachers and appreciate more inferences supported by the concept.² Thus there seem to be more substances than natural classes (as described above). Whether all substances have a nondescriptionalist character seems an open question.³

Summary

A distinction between natural and artificial appears throughout the literature on concepts. However, rather than reflecting a commitment to a shared set of assumptions or problems, researchers frequently mean different things by the designations. A review of the literature suggests two broad types of characterizations of conceptual naturalness. A concept may be natural (or artificial) because of the psychological aspects of its acquisition or use. Alternatively, a concept may be natural because of its content or relation to external phenomena. Within psychological approaches there are different specific proposals. A concept may be deemed natural if it is representable by the human mind; naturalness is a structural constraint on possible concepts. Alternatively, a natural concept may be one that is a product of some psychological process (e.g., prototype abstraction) rather than others (e.g., rule learning); naturalness is a degree of optimality in cognitive operations. Finally, of the psychological construals, natural concepts may be those that are central to intuitive theories; naturalness is an inferential role. External accounts also come in three variants. Natural concepts may be those with naturally occurring instances, those that are representations of naturally occurring classes, or those with externalist semantics. One purpose of this essay has been to delineate these different approaches to conceptual naturalness. A second focus is to draw out some of the implications of these construals for research on concepts. Different ideas of naturalness motivate different research questions. Two researchers both studying natural concepts may have quite divergent goals.

IMPLICATIONS FOR RESEARCH ON CONCEPTS

The primary focus of research on natural and artificial categories has been to describe psychological correlates of conceptual naturalness. For example, naturally occurring individuals may be arrayed into distinct classes and be represented using taxonomically organized concepts, while classes of artifacts tend to overlap (Atran, 1987; Barr & Caplan, 1987; Diesendruck & Gelman, 1999). Natural kinds may be believed to be characterized by compositional properties or essences (in contrast to relational properties of artifacts, Ahn, 1998; Barton & Komatsu, 1989; Bloom, 1996; Gelman & Medin, 1993; Keil, 1989; Malt & Johnson, 1992), have all-or-none membership (Diesendruck & Gelman, 1999), and promote many inductive inferences (Markman, 1990). The two types of concepts might be represented differently in the brain (Caramazza, Hill, &

Leek, 1994; Warrington & Shallice, 1984 see Forde & Humphreys, 1999 for review). In each case research focuses on whether concepts of natural kinds have the proposed psychological property and whether they differ from concepts of artifacts⁴.

Consider results demonstrating a correlation between conceptual type and cognitive structure, for example, some sorts of brain damage impair naming instances of natural kinds but not artifacts. The logical next question is to ask what it is about natural versus artificial objects that accounts for this difference. It is not possible for natural status to have a direct effect on mental representation. Perhaps it is discovered that the category-specific deficits depend on whether the concept has a perceptual or functional basis (see Forde & Humphreys, 1999). At this point an external construal of naturalness is no longer useful; instead of characterizing the deficits in terms of natural kinds and artifacts, it is more accurate to refer to perceptual and functional concepts. Research may uncover correlations between psychological variables and natural status, but in each case, naturalness must reduce to some psychological property.

Research based on a psychological construal of naturalness is in a different position. The properties that define natural concepts are determined by prior theoretical or epistemological commitments. The goal is to demonstrate that the psychological properties constitutive of naturalness cause or determine other psychological properties. For example, a researcher starts from the belief that computation of similarity relations among perceptual features is the most natural process of concept formation. The empirical question then becomes which other properties of concepts follow from, or are associated with, the defining property of naturalness. The motivation for identifying some psychological property of concepts as “natural” is that it has important consequences. Critically, the properties used to identify natural concepts (the independent measure) must be different than those found to correlate (the dependent measure). For example, if the researcher finds that natural kinds are organized around perceptual similarity, he or she cannot have used perceptual basis as a criteria for selecting natural kinds. Unfortunately, the basis for selecting natural kinds and artifacts is rarely discussed in research on natural concepts (see Kalish, in press, for discussion of issues of sample selection). For researchers starting with an external construal of natural kinds, any psychological correlation is an empirical finding. For researchers

beginning with a psychological construal it is critical to be clear what is a logical consequence of the definition of natural concept and what is an actual empirical result.

External and psychological construals of naturalness have different implications for research. These differences extend to the relative importance of natural and artificial concepts. Psychological construals privilege natural concepts in a way that external construals do not. In particular, any given research study must choose whether to focus on natural or artificial concepts (or both). What are the consequences or motivations for these choices?

For researchers operating with an external construal it will be important to study both natural and artificial concepts. Clearly people form and use concepts of naturally occurring and artificially constructed objects and classes. Any psychological differences between the two kinds of concepts need not carry implications of priority. Suppose it turns out that natural concepts are organized around intrinsic properties while (most) artificial concepts are based on relational properties (Barton & Komatsu, 1989). It is not therefore the case that intrinsically-based concepts are somehow more fundamental or cognitively privileged than relationally-based concepts.

In contrast to the external case, structural constraints on conceptual naturalness, such as the ontological hierarchies (Keil, 1979; Osherson, 1978), make strong claims about the infrequency of artificial concepts. Thus it is an embarrassment to such theories that they treat as violations some easily learned and commonly used concepts, and that they treat as acceptable some apparently unlearnable concepts (see Carey, 1986). On the processing principles and inferential role accounts it is expected that among the set of easily acquired, common concepts many will be artificial, at least to some degree. Given such relaxed standards, it may nonetheless be that artificial concepts are less easily learned, or are rarer, than natural concepts. Researchers expect that concepts according with principles of categorization will be preferred (e.g., learned more easily) over those that do not. A complicating factor is that there may be several, somewhat independent, sets of principles of categorization (e.g., similarity-based and rule-based, Smith, et al., 1998). These contrasting sets might yield contrasting preferences. Nonetheless, an attribute that characterizes only few commonsense concepts is a poor candidate for a natural principle of concept formation.

Inferential role (theory-based) accounts of naturalness need not even hold themselves to the relaxed standard of relative preference. It may well be that only a few, specially acquired, concepts

are central to commonsense theories (have rich inductive potential, essential structure, etc.). People could show no particular propensity for forming theory-based versus atheoretical concepts. Preferences may line up with the proposals, concepts deemed natural could be common and easily learned, but they may not. In contrast to our pretheoretical assumptions, it may be the artificial side of the distinction which is more common. In this case it would be the sense of natural concepts as those discovered groupings which truly exist in the world that could license the labeling of potentially uncommon, difficult to learn concepts as 'natural.'

The natural-artificial distinction is not primarily useful as a criterion for defining the domain of concept research. Rather, in developing a distinction between natural and artificial concepts researchers are lead to explore dimensions along which concepts may differ. At a higher level of remove, this process leads to questions about the interrelationships between the various dimensions proposed by different researchers. The real utility of the idea of naturalness is in supporting this kind of investigation. Which qualities of concepts co-occur? Do concepts with all-or-none membership have greater inductive potential? Are linearly separable combinations of features easier to learn? The designation 'natural' should reflect a hypothesis about the ways properties of concepts cluster. Empirical research can then proceed to test the hypothesis; do concepts with some of the properties indicative of naturalness have the others as well? A primary challenge for research is to demonstrate the significance of distinguishing between natural and artificial concepts by showing important within group similarities and between group differences.

The primary purpose of this review has been to discuss the kinds of empirical questions about concepts generated by different perspectives on conceptual naturalness. In particular, two programs of research were described. First, research may be directed toward exploring the psychological correlates of natural concepts. Do concepts identified as natural on external criteria share any significant psychological properties? A second program of research focuses on interrelations between psychological characteristics of concepts. To what degree is it possible to describe clusters of properties that make concepts natural or artificial? These questions have been important empirical and theoretical ones within the field. There is every reason to believe that the continued development of accounts of naturalness and artificiality will continue to spur productive advances for research on categorization and concepts.

NOTES

¹ These intuitions are not the only, or even primary, motives for externalist characterizations of concepts. Fodor (1998) argues for nondescriptionalist views of concepts based on considerations of individuation, stability, and compositionality.

² A similar point applies to the idea of externalism supported by deference to experts ('deferential kinds' Margolis, 1998). It seems as often as not what someone is deferring to is an expert's analytic definition of a term (e.g., BRISKET, Burge, 1979) rather than to an expert's superior ability to track real kinds. Such deference seems equally characteristic of artificial concept acquisition. It is only by deference to the experimenter that a subject can learn that FOO means BLUE OR SQUARE.

³ Although a full discussion of this issue is beyond the scope of this paper, Fodor (1998) would regard both the natural and artificial concepts as atomistic in that one may possess an artificial concept (JADE, JEWELRY) without representing the constituents of its description. On Fodor's view it is not clear that even FOO is descriptionalist; subjects may make claims about FOO, have FOO thoughts, without the concepts BLUE and SQUARE. This may happen when the subject is mistaken and thinks, for example, that FOO means GREEN AND CIRCLE. In contrast one cannot think about 'BLUE OR SQUARE' without BLUE and SQUARE. Thus a concept may be artificial in that it involves a reproduction of an experimenter's description without having descriptionalist semantics (at least for the subject of the experiment). However, it is not clear that one may have an artificial concept without some, perhaps mistaken, representation of a constituent description.

⁴ For examples of a debate concerning the psychological attributes of concepts of natural vs. artificial kinds see Diesendruck & Gelman (1999), Malt (1992), Rips (1989) arguing natural kinds have all-or-none membership determined by hidden essences and Braisby & Franks, (1996), Hampton (1996), Kalish (1995, in press) arguing that natural kinds and artifacts do not differ in this way.

REFERENCES

- Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. Cognition, *69*, 135-178.
- Atran, S. (1987). Folkbiological universals as common sense. In S. Modgil and C. Modgil (Eds.), Noam Chomsky: Consensus and controversy (pp. 247-268). London: Falmer Press.
- Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. Memory & Cognition, *15*, 397-418.
- Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. Journal of Psycholinguistic Research, *18*, 433-447.
- Bloom, P. (1996). Intention, history, and artifact concepts. Cognition, *60*, 1-29.
- Bourne, L. E. Jr. (1974). An inference model of conceptual rule learning. In R. Solso (Ed.), Theories in cognitive psychology (pp. 231-256). Washington, DC: Erlbaum.
- Boyd, R. (1991). Realism, Anti-foundationalism, and the enthusiasm for natural kinds. Philosophical Studies, *61*, 127-148.
- Braisby, N. Franks, B., & Hampton, J. (1996). Essentialism, word use, and concepts. Cognition, *59*, 247-274.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.
- Caramazza, A., Hills, A., & Leek, E. C., (1994). The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits. In S. Gelman and L. Hirschfeld (Eds.). Mapping the mind: Domain specificity in cognition and culture. (pp.68-84). New York: Cambridge University Press.
- Carey (1986). Constraints on semantic development. in Demopolous & Marratsos (Eds.), Language learning and concept acquisition: Foundational issues. (pp.154-171). Norwood: Ablex .
- Carey, S., & Spelke, E. (1993). Domain-specificity and conceptual change. In L. Hirschfeld & S. Gelman, (Eds.). Mapping the mind: Domain specificity in cognition and culture. (pp. 169-200). New York: Cambridge University Press.

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

Fish, S. (1996). Professor Skokal's bad joke. New York Times. May 21, p. 23.

Fodor, J. A. (1975). The language of thought, Cambridge, MA: MIT Press.

Fodor, J. A. (1998). Concepts: Where cognitive science went wrong. Oxford: Clarendon.

Forde, E. M. & Humphreys, G. W. (1999). Category-specific recognition impairments: A review of important case studies and influential theories. Aphasiology, 13, 169-193.

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

Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. in R. Pasnak and M. L. Howe (Eds). Emerging themes in cognitive development. (pp. 3-32). New York: Springer-Verlag.

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

Gelman, S. A., & Medin, D. L. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. Cognitive Development, 8, 157-167.

Goodman, N. (1955). Fact, fiction, and forecast, Indianapolis: Bobbs-Merrill.

Hacking, I. (1999). The social construction of what? Cambridge, MA: Harvard University Press.

Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, and A. Premack (Eds.) Causal cognition: A multidisciplinary debate. Symposia of the Fyssen Foundation. (pp. 351-394). New York: Clarendon

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. Cognition, 65, 137-165.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts: An experimental study. Psychological Monographs, 28, No. 1, Series No. 123. Princeton, N.J.: Psychological Review Company.

Johnson, K.E., Scott, P., & Mervis, C. B. (1997). Development of children's understanding of basic-subordinate inclusion relations. Developmental Psychology, 33, 745-763

Kakise, H. (1911). A preliminary experimental study of the conscious control of understanding. American Journal of Psychology, 22, 14-64.

Kalish, C. W. (in press). Essentialist to some degree: The structure of natural kind categories. Memory & Cognition.

Kalish, C. W. (1995). Graded membership in animal and artifact categories. Memory & Cognition, 23, 335-353.

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

Keil, F. C. (1979). Semantic and conceptual development: an ontological approach, Cambridge, MA: Harvard University Press.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. Psychological Review, 88, 197-227.

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

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. Journal of Verbal Learning and Verbal Behavior, 23, 221-236.

Keil, F. C., Carter Smith, W., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism. Cognition, 65, 103-135.

Kelly, M. H., & Keil, F. C. (1985). The more things change . . . : Metamorphoses and conceptual structure. Cognitive Science, 9, 403-416.

Kripke, S. (1972). Naming and necessity. In D. Davidson and G. Harman (Eds.), Semantics of natural language (pp. 253-355). Dordrecht, Holland: Reidel.

Lewis, D. (1970). How to define theoretical terms. Journal of Philosophy, 67, 427-446.

Locke, J. (1961). An essay concerning human understanding, 5th edition, Volume II. New York: E.P. Dutton. (Original work published 1707).

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

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? Journal of Memory and Language, 31, 195-217.

Margolis, E. (1998). How to acquire a concept. Mind and Language, 13, 347-369.

Markman & Hutchinson (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. Cognitive Psychology, 16, 1-27.

Markman, E. M. (1990) Categorization and naming in children. Cambridge, MA: MIT Press.

McGill, A. L. (2000) Counterfactual reasoning in causal judgments: Implications for marketing. Psychology and Marketing, 17, 323-343.

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

Medin, D. L., Lynch, E. B., & Solomon, K. E. (2000). Are there kinds of concepts. Annual Review of Psychology, 51, 121-147.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. in S. Vosniadou and A. Ortony (eds.) Similarity and analogical reasoning. (pp. 179-195). New York: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. Psychological Review, 85, 207-238.

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, 9, 55-100

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

Nelson, D. K. (1984). The effect of intention on what concepts are acquired. Journal of Verbal Learning and Verbal Behavior, 23, 734-759.

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. Healy and S. Kosslyn (Eds.). Essays in honor of William K. Estes, Vol. 2: From learning processes to cognitive processes. (pp. 149-167). Hillsdale, NJ, USA: Lawrence Erlbaum .

Osherson, D. N. (1978). Three conditions on conceptual naturalness. Cognition, 6, 263-289.

- Putnam, H. (1975). 'The Meaning of 'Meaning'' in H. Putnam, Mind, Language and Reality: Philosophical Papers, Vol. 2 (pp.215-271). Cambridge: Cambridge University Press.
- Reznek, L. (1987). The nature of disease. New York: Routledge & Kegan Paul.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 21-59). New York: Cambridge University Press.
- Rorty, R. (1999). Philosophy and social hope. London: Penguin.
- Rosch, E. (1973). Natural categories. Cognitive Psychology, 4, 206-216.
- Rosch, E. (1978). Principles of categorization. In E. Rosch and B. B. Lloyd (Eds.) Cognition and categorization, (pp. 59-77). Hillsdale: Erlbaum.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.
- Schwartz, S. P. (1979). Natural kind terms. Cognition, 7, 301-315.
- Searle, J. R. (1983). The construction of social reality. New York : Free Press.
- Smith, E. E. (1995). Categories and concepts. In E. Smith & D. Osherson (Ed.s). An invitation to cognitive science (2nd Edition), Volume 3, Thinking. (pp. 3-33). Cambridge: MIT Press.
- Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.
- Smith, E. E., & Sloman, S. (1994) Similarity vs. rule-based categorization. Memory & Cognition, 22, 377-386.
- Smith, E. E., Patalano, & Jonides (1998). Alternative strategies of categorization. Cognition, 65, 167-196.
- Smith, J. D. & Minda, J. Paul (1998). Prototypes in the mist: The early epochs of category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 24, 1411-1436.
- Smith, L. B., (1989). From similarities to kinds of similarities. in Vosniadou & Ortony (Eds.), Similarity and analogical reasoning. (pp. 146-178). New York: Cambridge U. Press.
- Smoke, K. L. (1932). An objective study of concept formation. Psychological Monographs, 42, No. 4, Series No. 46. Princeton, N.J.: Psychological Review Company.

Sperber, D. (1990). The epidemiology of beliefs. in C. Fraser and G. Gaskell (Eds.), Social psychological study of widespread beliefs. (pp. 25-44). Oxford: Clarendon Press.

Tversky, B. (1985). The development of taxonomic organization of named and pictured categories. Developmental Psychology, 21, 1111-1119.

Vygotsky, L. S. (1962). Thought and Language. Cambridge, MA: MIT Press.

Warrington, E. K., & Shallice, T. (1984). Category-specific semantic impairment. Brain, 107, 89-854.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. Cognitive Psychology, 18, 158-194.

Wellman & Gelman (1993) Cognitive development: Foundational theories of core domains, Annual Review of Psychology, 43,

ACKNOWLEDGEMENTS

Preparation of this article was supported by a research grant from the National Institute of Health (R01 HD37520 – 02). Thanks to Martha Alibali for helpful comments and suggestions.