to relate problems appearing in recent debates in cognitive science to problems addressed by the great philosophers of past centuries.

Acknowledgement. This issue is a result of the project 'Concepts: representation and evolution' of the lab for Applied Epistemology at the University of Ghent. During this project, the editor was supported as a senior research associate by the Flemish Fund for Scientific Research. Participants in this project, in alphabetical order, were: Michael Arbib, Andy Clark, Terry Dartnall, Liane Gabora, Ben Goertzel, Yozef Grodzinsky, Mark Hubey, Markus Peschl, Steven Pinker, Henry Plotkin, William Robinson, Lee Spector, Paul Thagard, Wolfgang Tschacher, Lucia Vaina, Fernand Vandamme and Philip Van Loocke. Papers not included in this issue appear in Philosophica (1996, nr57), and in CC-ai (1996, nr13).

Communication & Cognition
Vol.29, Nr.3/4 (1996), pp.307-362

THE NATURE OF HUMAN CONCEPTS/ EVIDENCE FROM AN UNUSUAL SOURCE

*Steven Pinker**
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, Mass.

*Alan Prince**
Department of Linguistics
Rutgers University
New Brunswick, New Jersey

Abstract

What kind of categories do human concepts represent? *Classical* categories are defined by necessary and sufficient criteria that determine whether an object is in a category or not in it. A popular contemporary view is that concepts correspond instead to *prototype* categories. *Prototype* categories lack necessary and sufficient conditions; their members need not be absolutely "in" or "out of" the category but can be members to greater or lesser degrees; their members display family resemblances in a number of characteristic properties rather than uniformly sharing a few defining properties; and they are organized around "prototypical" exemplars. This distinction raises several questions. Is one type of category psychologically real, the other an artifact (of formal schooling, or of the experimental methodologies used to study them)? If both are psychologically real, do they serve different functions in cognition? Are they processed by the same kind of computational architecture? And do they correspond to fundamentally different kinds of things in the world? We suggest that some answers to these questions can be found in an unusual place: the English past tense system.

This paper is about an extensive parallel we have discovered between a part of language and a part of cognition, and about the possibility that the parallel is not a coincidence. The parallel involves the difference between a *classical category* and a *prototype* or *family resemblance category*, a topic of controversy for many years in cognitive psychology, philosophy, linguistics, and artificial intelligence.

Classical categories are defined by necessary and sufficient criteria, and membership in them is all-or-none. Examples include squares, grandmothers, odd numbers, and the vertebrate class *Aves*. Family Resemblance categories differ from classical categories in a number of ways:

- They *lack necessary and sufficient conditions* for membership. For example, the category "chair" includes objects that have legs and that lack them (e.g. beanbag chairs), and objects that can be sat upon and that cannot (e.g. delicate museum pieces).

- They have *graded degrees of membership*. A robin is a better example of the family resemblance category "bird" than an eagle is; and a penguin a worse example.

- The category can be summarized by an ideal member or *prototype*, sometimes but not always an actual exemplar of the category. The more similar other members are to the prototype, the "better" examples they are. The sparrow, which is used to illustrate the entry for "bird" in many dictionaries, might be a prototype of the bird category.

- There can be *unclear cases* — objects which may or may not be members of the category at all. One example is the fossil species Archaeopteryx, characterized by one paleontologist as "a piss-poor reptile, and not very much of a bird" (Konner, 1982). Garlic is an unclear example of the category "vegetable," as is ketchup, as we see in the famous controversy that followed the proposal of the Reagan administration that ketchup be classified as a vegetable in meeting nutritional guidelines for school lunch menus.

- They often display a *family resemblance* structure (Wittgenstein, 1953).[1] The members of a family of people generally do not have a single feature in common. Instead, a pool of features such as hair color, mouth shape, or nose size is shared by various sets of family members. Similarly, the members of family resemblance categories have different features that run through different subsets: green color is shared by spinach, celery, and broccoli, but not carrots or cauliflower; stems and bunches of florets are shared by broccoli and cauliflower but not carrots.

- Good members tend to have *characteristic nondefining features*. For example, gray hair and a domestic lifestyle characterize many grandmothers, but someone can be a grandmother without possessing either property, such as Elizabeth Taylor.

Evidence for Family Resemblance Categories

Human concepts pick out categories of objects; what kind of category do they pick out? There is a large body of evidence, summarized in Brown (1976), Smith and Medin (1981), and Rosch (1972, 1977, 1988), that has been taken to show that human concepts correspond to family resemblance categories. First, semanticists and philosophers have generally failed in their attempts to find necessary and sufficient conditions for most natural concepts that are labeled by words (see Fodor, Garrett, Walker, and Parkes, 1980). Second, psychologists have found that subjects can give ratings of the goodness of membership of a list of exemplars with respect to a category that are reliable and in close agreement with one another. Similarly, there is good agreement about prototypes and unclear cases. Third, these judgments are not unanalyzable gut feelings but can be predicted in a systematic way using a feature calculus, in which the features possessed by a given exemplar (assessed independently, for example, by asking subjects to list the attributes of the object) are compared with those possessed by the other members of the category. Fourth, judgments of goodness of membership have strong effects on performance in various psychological tasks. For example, people can verify that prototypical members are members of a category faster and more accurately than peripheral members, and when asked to recall instances of a category, they name prototypical members first. Fifth, developmental psychologists have found that children often learn the names for prototypical exemplars of a category before learning other exemplars, and apply superordinate terms such as *bird* to its prototypical members first. Sixth, linguists have found that certain adverbials called *hedges* are sensitive to prototypicality: one can say that a sparrow, but not a penguin, is a bird *par excellence*, and that a penguin, but not a sparrow, is *technically* or *strictly speaking* a bird.

Evidence Against Family Resemblance Categories

On the other hand, there is also evidence that certain aspects of human concepts do not correspond to family resemblance categories. Some of the empirical effects that have been interpreted as demonstrating family resemblance classes also occur for categories that people clearly treat as being classical. Armstrong, Gleitman, and Gleitman (1983) have found that subjects show a great deal of agreement with one another in rating the degree of membership of exemplars of categories like "female" and "odd number". For example, they agree that a mother is a better example of a female than a comedienne is, and that 13 is a better example of an odd number than 23. Similarly, Armstrong, et al. found that people take less time and are more accurate at deciding that 13 is an odd number than that 23 is, and that a mother is a female than that a comedienne is. Since these subjects surely knew that 'female' and 'odd number' in reality have sharp boundaries and all-or-none membership (and Armstrong, et al. discovered, in an independent questionnaire, that their subjects believed as such), it calls into question whether the analogous results that Rosch and others obtained for 'bird' or 'tool' really tell us anything about people's representations of those concepts.

Moreover, most judgments of membership in family resemblance categories based on characteristic features are highly corrigible when people are asked to engage in careful reasoning about it. For some purposes people are willing to consider a penguin as a full-fledged bird and Elizabeth Taylor a full-fledged grandmother. In fact characteristic non-defining features can be quickly abandoned, even by young children. Children say that three-legged dogs are dogs, that raccoons with stripes painted down their backs are raccoons, not skunks (Rey, 1983; Armstrong, et al, 1983; Keil, 1989).

People not only shelve their perception of similarity when pressed as to what an object "really is," but also when making everyday inferences about the object based on its category membership. For example, Gelman and Markman (1987) showed three-year old children a picture of a flamingo, a picture of a bat, and a picture of a blackbird, which looked a lot more like the bat than the flamingo. They told the children that a flamingo feeds its babies mashed-up food but a bat feeds its babies milk, and asked them what the blackbird feeds its babies. With no further information, children go by appearances and predict "milk." But all it took was a mention that flamingos and blackbirds were "birds," and the children lumped them together and predicted "mashed-up food."

Similar demonstrations with adults have shown that inference is often not driven by the similarity criteria that define family resemblance categories (see Murphy, 1993; Medin, 1989; Kelly, 1992; Smith, Langston, & Nisbett, 1992; Rips, 1989; Rey, 1983). For example, when people are asked which two out of three belong together — white hair, gray hair, black hair — they say that black is the odd hair out, because aging hair turns gray then white. But when asked about a white cloud, a gray cloud, and a black cloud, they say that white is the

odd cloud out, because gray and black clouds give rain. In another experiment, subjects were asked whether a a three-inch disk is more similar to a quarter or a pizza, and whether it is more likely to *be* a quarter or a pizza? Most said it is more similar to a quarter but more likely to be a pizza, presumably because quarters have to be standardized but pizzas can vary. Most people, upon being presented with a centipede, a caterpillar that looks like it, and a butterfly that the caterpillar turns into, feel that the caterpillar and butterfly are "the same animal," but the caterpillar and centipede are not, despite appearances to the contrary.

Possible resolutions

There are several possible resolutions of this conflicting evidence.

First, human concepts could basically pick out family resemblance categories. Classical categories would be special cases or artifacts resulting from explicit instruction, such as in formal schooling. Alternatively, human concepts could basically pick out classical categories. Family resemblance categories would be artifacts of experimental tasks asking subjects for graded judgments or asking them to make categorization decisions under time pressure. A third, compromise position, would say that human concepts correspond to both classical and family resemblance categories. Classical categories are the "core" of the concept, used for reasoning. Family resemblance categories are "identification procedures" or "stereotypes," used for identification of category exemplars on the basis of available perceptual information, or for rapid approximate reasoning. Although most theorists have tended toward compromise positions, something close to the mainly-family-resemblance view can be found in Lakoff (1987), Rosch (1978), and Smith, Medin, and Rips (1984); something close to the mainly-classical view can be found in Rey (1983), Fodor (1981), and Armstrong, et al. (1983), and tentative proposals favoring the core-plus-identification-procedure compromise can be found in Smith and Medin (1981), Armstrong, et al. (1983), and Osherson and Smith (1981).

This leads to several open questions. (1) Is one type of category psychologically real, the other an artifact or special case? (2) If both are psychologically real, can they be distinguished by function (e.g., reasoning versus categorization)? (3) If both are psychologically real, are they handled by the same kind of computational architecture? (4) If either or both are psychologically real, do they correspond to ontological categories? Rey (1983) stresses the importance of distinguishing the "metaphysical" problem of what kinds of categories the world contains (as characterized by the best current scientific characterization of that aspect of the world), and the "epistemological" or psychological question of what kinds of categories people use to understand the world. That is, are classical (or family resemblance) categories incorrectly imposed by people on the world because of limitations of the way the mind works, or is there some sense in which the world contains classical (or family resemblance) categories, which people can accurately represent as such, presumably because the mind evolved to grasp aspects of the world accurately?

We will attempt to shed light on these questions by examining an unusual source of evidence: English past tense forms.

An Unexpected Test Case: English Past Tense Forms

English verbs come in two types: those that have regular past tense forms, and those that have irregular past tense forms. Consider them as two categories: "regular verbs," such as *walk/walked*, *talk/talked*, *jog/jogged*, *pat/patted*, *kiss/kissed*, and *play/played*, and "irregular verbs," such as *hit/hit*, *go/went*, *sleep/slept*, *make/made*, *ring/rang*, *bring/brought*, *stink/stunk*, and *fly/flew*.

In fact the irregular verbs are not a single class but a set of subclasses, which can be subdivided according to the kind of change that the stem undergoes to form the past tense (Jespersen, 1942; Curme, 1935; Bybee and Slobin, 1982a; Pinker and Prince, 1988):

- **No-Change:** *hit, fit, slit, split, quit, knit, spit, shit, bid, rid, forbid, shed, spread, wed, let, set, upset, wet, cut, shut, put, burst, cast, cost, thrust, hurt*
- **Lax the Vowel:** *bleed, breed, feed, lead, mislead, read, speed, plead, meet, hide, slide, bite, light, shoot*
- **Devoice the Final** ***d*****:** *bend, send, spend, lend, rend, build*
- **Lax the Vowel, Add a** ***-t*****:** *lose, deal, feel, kneel, mean, dream, creep, keep, leap, sleep, sweep, weep, leave*
- **Change the Vowel, Add a** ***-d*****:** *flee, say, hear, sell, tell, foretell, do*
- **Change the Rhyme to** ***-ought*****:** *buy, bring, catch, fight, seek, teach, think*

- **Change** ***E*** **or similar vowel to** ***O*****:** *freeze, speak, steal, weave, get, forget, swear, tear, wear, bear, forbear, forswear, awake, wake, break, choose*
- **Change** ***-ing*** **to** ***-ang*** **or** ***-ung*****:** *ring, sing, spring, drink, shrink, sink, stink, swim, begin, cling, fling, sling, sting, string, swing, wring, stick, dig, win, spin, stink, slink, run, hang, strike, sneak*
- **Change the Vowel to** ***u*****:** *blow, grow, know, throw, draw, withdraw, fly, slay*
- **Miscellaneous Vowel Changes:** *take, mistake, forsake, shake, partake, bind, find, grind, wind, rise, arise, write, smite, ride, drive, strive, fall, befall, hold, behold, come, become, eat, beat, see, give, forgive, forbid, sit, spit, lie*
- **Complete Replacement (Suppletion):** *be, go, undergo*

Let us consider some properties of the irregular subclasses.

Properties of the Irregular Subclasses

1. Characteristic Nondefining Features

The irregular subclasses tend to be characterized by phonological properties other than those that define the change from stem to past form. Consider the subclass that changes an *o* vowel to *u*, I[blow, grow, know, throw, draw, withdraw, fly, slay]. In principle, any verb with an *o* or similar vowel could be included in the subclass. In fact, all the verbs in the subclass end in a vowel, usually a diphthong, and most begin with a consonant cluster.

Similarly, the subclass that changes *ay* to *aw* — *bind, find, grind, wind* — could include any verb with the vowel *ay*, but in fact, all the verbs happen to end in *-nd*. The subclass that changes a final *d* to *t* — *bend, send, spend, lend, rend, build* — could include any word ending in *d*, but in fact, most of the verbs rhyme with *-end*. Finally, the subclass that changes the vowel *e* to *U* — *take, mistake, forsake, shake* — could include any word with an *e*, but in fact all the verbs rhyme with *-ake* and begin with a coronal consonant.

Note that the characteristic nondefining features are arbitrary, not lawful, with respect to the sound pattern of English. No rule of phonology excludes *loon* as the past tense of *loan* or *choud* as the past of *chide*.

2. Family Resemblance

Irregular subclasses display a family resemblance structure (Bybee and Slobin, 1982a; Bybee and Moder, 1983). Consider the subclass that changes an *I* to an ^. Most of the verbs end with velar nasal consonant: *shrink, sink, stink, cling, fling, sling, sting, string, swing, wring, slink*. Some end in a consonant that is velar but not nasal: *stick, dig, sneak, strike*. Others end in a vowel that is nasal but not velar: *win, spin, swim, begin*.

Similarly, within the subclass that changes a final diphthong to *u*, some begin with <consonant-sonorant> cluster, and contain the diphthong *ow*: *blow, grow, throw*. But one member, *know*, contains the *ow* diphthong, but does not begin with a consonant cluster. Others begin with a consonant cluster, but have a different diphthong or no diphthong at all: *draw, withdraw, fly, slay*.

3. Prototypicality

Bybee and Moder (1983) point out that for many of the subclasses, one can characterize a prototype, based on the kinds of characteristic phonological properties that go into defining the family resemblance structure. According to Bybee and Moder, the prototype of the *ing* --> *ung* subclass is:

S C C i [velar]
[nasal]

where 'C' stands for a consonant. This prototype is maximally similar to the most members of the existing subclass, but more interestingly, it predicts subjects' generalization of the *I* --> ^ change to novel verbs. Bybee and Moder asked subjects to rate how natural a variety of putative past tense forms sounded for each of a set of nonce stems. The independent variable was the similarity of the stem to the prototype listed above. They found that subjects were extremely likely to accept the vowel change for stems like *spling, strink*, and *skring*, which match the schema for the prototype exactly. They were only slightly less willing to accept *struck* and *skrum* as the past of *strick* and *skrim*, which differ from the prototype in one feature. Somewhat lower in acceptability were *spruv* for *spriv*, and similar past forms for *sking, smig, pling* and *krink*. *Glick, krin, plim, shink* were even less likely to admit of the vowel change, and *trib, vin*, and *sid*, the forms furthest from the prototype, were the least acceptable of all. The results have been replicated by Prasada and Pinker (1993), and with analogous German forms by Brinkmann, Clahsen, Wiese, & Pinker (1995).

4. Graded Goodness-of-Membership

Within most of the subclasses, there are some verbs that clearly accept the irregular past tense form, but others, usually of low but nonzero frequency, for which the specified past tense form is less than fully acceptable, being accompanied by a sense of unusualness or stiltedness. Below we contrast, for a variety of subclasses, some "good examples" of the past tense form with "poor examples" of the same kinds of forms. Intuitions vary from person to person for the "poor" examples, as is true for nonprototypical exemplars of conceptual categories; the perceptions of "poorness" we report here are true for most of the speakers of American English we have consulted. and are documented quantitatively by Ullman (1993):

(1)

Good Examples	**Poor Examples**
hit, split	spit, forbid
bled, fed	pled, sped
burnt, bent	learnt, lent, rent
dealt, felt, meant	knelt, dreamt
froze, spoke	wove, hove
got, forgot	begot, trod
wrote, drove, rode	dove, strove, smote, strode

5. Unclear Cases

For some verbs associated with a subclass, the mandated past tense form is so poor in people's judgment that it is unclear whether the verb can be said to belong to the subclass at all. Sometimes these are verbs are restricted to idioms, clichés, or other specialized usages. For example, the expression *forgo the pleasure of*, as in *You will excuse me if I forgo the pleasure of reading your paper until it's published*, sounds fairly natural. Because the verb has a transparent morphological decomposition as [*for* + *go*], the form *forgoed* is clearly unacceptable, but the irregular past tense form, as in *Last night I forwent the pleasure of grading student papers*, is decidedly peculiar if not outright ungrammatical (this intuition has been corroborated by ratings from subject in a study by Ullman and Pinker, in preparation). Likewise, the sentence *the Vietnam War is rending the fabric of American society* is a natural-sounding cliché, but ? *The Vietnam War rent the fabric of American society* is distinctly less natural. One occasionally hears the idiom *That conclusion does not sit well with me*, but many people balk at *That conclusion has not sat well with many people. That dress really becomes you* is a natural English sentence; *When you were ten pounds lighter, that dress really became you* is almost unintelligible.

In other cases grammatical phenomena conspire to make the past tense form of a verb extremely rare. The transitive verb stand meaning "to tolerate" is fairly common but because it is usually used as the complement of a negated auxiliary, as in *She can't stand him*, the verb is almost always heard in its stem form. In constructions where the past is allowed to reveal itself, the verb sounds quite odd: compare *I don't know how she stands him with ?I don't know how she stood him;* similarly, *I don't know how she bears it* versus *?I don't know how she bore it*.

Conclusions about the Irregular Subclasses

Subclasses of irregular verbs in English have characteristic nondefining features, family resemblance structures, prototypes, gradations of goodness of membership, and unclear or fuzzy cases. Since these are exactly the properties that define family resemblance categories, we conclude, in agreement with Bybee and Moder (1983), that the irregular subclasses are family resemblance categories.

This is a surprising conclusion. Linguistic rules are traditionally thought of as paradigm case of categorical, all-or-none operations, and might be thought to correspond classical categories if anything did. The fact that entities subject to grammatical operations can have a clear family resemblance structure thus has far-ranging implications for some theorists. For example, for Rumelhart and McClelland (1986) this phenomenon is part of their argument for a radically new approach to studying language, based on a computational architecture in which rules play no causal role. For Lakoff (1987), it is part of a call for a radically new way of understanding human cognition in general.

It seems clear that at least one kind of linguistic object, English irregular past tenses, fall into family resemblance categories. An important question at this point is: Do all linguistic objects fall into family resemblance categories?

Properties of the Regular Class

More specifically, we might ask, do English regular verbs fall into family resemblance categories? One answer, favored by Bybee (Bybee and Moder, 1983; Bybee, 1991) and by Rumelhart and McClelland (1986), is "yes": the regular class just has more members, and more general characteristic features. Let us examine this possibility.

A Confounding Factor: The Blocking Principle

The regular and irregular classes interact in a specific way, and it is necessary to take account of this interaction so that the properties of the irregular subclasses do not confound our examination of the properties of the regular class. The interaction is governed by what has been called the "Blocking Principle" (Aronoff, 1976) or the "Unique Entry" Principle (Pinker, 1984): if a verb has an irregular past tense form, its regular form is pre-empted or "blocked". Thus the fact that *go* has an irregular past *went* not only allows us to talk of past instances of going using *went*, but it prevents us from using **goed*. The verb *glow*, in contrast, does not have an irregular past **glew*, so its regular past *glowed* is not blocked.

We saw in a previous section how some irregular past forms are "fuzzy" or marginal in their grammaticality. As a result of Blocking, these gradations of goodness can cause the appearance of complementary gradations of goodness of the corresponding regular. Thus because *pled* is a marginal past tense form for *plead* but one that we nonetheless recognize, the regular form *pleaded guilty* sounds fairly good but may be tinged with a bit of uncertainty for some speakers. Conversely, *?wept* is a fairly good past tense form of *weep*, though not maximally natural (compare, for example, *kept* for *keep*). As a result *??weeped* does not sound terribly good, though it is not perceived as being completely ungrammatical either (compare **keeped*). This effect has been documented by Michael Ullman (1993; see also Pinker, 1991 and Pinker & Prince, 1994) who asked subjects to rate the naturalness of irregular and regularized past tense forms for 40 verbs whose irregular pasts were somewhat fuzzy in goodness. The two sets of ratings were consistently negatively correlated.

What we now try to do is put aside this reciprocity effect due to blocking, and see if it is possible to determine whether the regular class has family resemblance-category properties independent of those of the irregular subclasses with which it competes.

1. Independence of the Phonology of the Stem

The first salient property of the regular class is that it has no sensitivity to the phonological properties of its stems. As a result, it has no phonologically characterized prototype, gradations of membership, or characteristic features.

First, the phonological conditions that govern the irregular subclasses can be entirely flouted by regular verbs. In the extreme case, homophones can have different past tense forms: *ring/rang* versus *wring/wrung*, *hang/hung* (suspend) versus *hang/hanged* (execute), *lie/lay* (recline) versus *lie/lied* (fib), *fit/fit* (what a shirt does) versus *fit/fitted* (what a tailor does). More generally, there are regular counterexamples to the membership criteria for each of the irregular subclasses:

(2)

shut/shut	jut/jutted
bleed/bled	need/needed
bend/bent	mend/mended
sleep/slept	seep/seeped
sell/sold	yell/yelled
freeze/froze	seize/seized
grow/grew	glow/glowed
take/took	fake/faked
stink/stunk	blink/blinked
ring/rang	ring/ringed

This shows that the phonologically-defined fuzzy boundaries of the irregular subclasses do not create complementary phonological fuzzy boundaries of the regular classes. The effect of the Blocking Principle is that specific irregular *words* block their corresponding regulars. Though most of those words come from regions of phonological space whose neighbors are also often irregular, those regions do not define complementary fuzzy "holes" in the space from which the regulars are excluded; a regular form can occupy any point in that space whatsoever. Moreover, it is not just that there *already* exist regular verbs in the language that live in irregular phonological neighborhoods; regular class can *add* members that violate *any* irregular membership criteria. The reason has been spelled out by Kiparsky (1982a,b), Pinker and Prince (1988, 1992), Kim, Pinker, Prince, & Prasada (1991), Kim, Marcus, Pinker, Hollander, & Coppola (1994). Irregular forms are verb roots, not verbs. Not all verbs have verb roots: a verb that is intuitively derived from a noun (e.g. *to nail*) has a noun root. A noun or an adjective cannot be marked in the lexicon has having an "irregular past," because nouns and adjectives do not have past tense forms at all; the notion makes no sense. Therefore, a verb created out of a noun or adjective cannot have an irregular past either. All such verbs are regular, regardless of their phonological properties:

(3)

He braked the car suddenly. ≠broke
He flied out to center field. ≠flew
He ringed the city with artillery. *rang
Martina 2-setted Chris. *2-set
He sleighed down the hill. *slew
He de-flea'd his dog. *de-fled
He spitted the pig. *spat
He righted the boat. *rote
He high-sticked the goalie. *high-stuck
He grandstanded to the crowd. *grandstood.

This makes it possible, in principle, for *any* sound sequence whatsoever to become a regular verb. There is a lexical rule in English that converts a name into a verb prefixed with *out*, as in *Reagan has finally out-Nixoned Nixon*. Like all verbs derived from non-verbs, it is regular. Since any linguistically-possible sound can be someone's name, any linguistically possible sound can be a regular verb, allowing there to be regular homophones for any irregular. For example:

(4)

Mary out-Sally-Rided Sally Ride.
*Mary out-Sally-Rode Sally Ride.

In grim notoriety, Alcatraz out-Sing-Singed Sing-Sing.
*In grim notoriety, Alcatraz out-Sing-Sang Sing-Sing.
*In grim notoriety, Alcatraz out-Sang-Sang Sing-Sing.
*In grim notoriety, Alcatraz out-Sing-Sung Sing Sing.
*In grim notoriety, Alcatraz out-Sung-Sung Sing Sing.

This effect has been demonstrated experimentally in several kinds of subject. Kim, et al. (1991) asked subjects to rate the regular and irregular past tense forms of a set of verbs that were either derived from nouns that were homophonous with an irregular verb or were derived directly from the irregular verbs. For verbs with noun roots, the regular form was given higher ratings; for verbs with verb roots, the irregular form was given higher ratings. Similar effects have been demonstrated in non-college-educated subjects (Kim, et al., 1991), children (Kim, et al., 1994), and German-speaking adults (Marcus, et al., 1995).

Perfectly natural-sounding regular past tense forms exist not only when the verb root is similar to an irregular, but when it is *dissimilar* to existing regular roots, and hence lacks a prototype that would serve as the source of an analogical generalization. Prasada and Pinker (1993) replicated Bybee and Moder's (1983) study but also presented novel *regular* words of differing similarity to existing English regular words. For example, *plip* is close to one of the prototypes for regular verbs in English, because it rhymes with *slip, flip, trip, nip, sip, clip, dip, grip, strip, tip, whip,* and *zip*, whereas *smaig* rhymes with no existing verb root, and *ploamph* is not even phonologically well-formed in English. Nonetheless people rated the prototypical and peripheral forms as sounding equally natural (relative to their stems), and were produced the prototypical and peripheral forms with the same probability when they had to produce them.

2. No Prototypes, Gradation of Membership, or Unclear Cases Caused by Low Frequency or Restricted Contexts

Unlike irregular past tense forms, regular past tense forms do not suffer in well-formedness on account of frequency, familiarity, idiomaticity, frozenness, or restricted syntactic contexts. Pinker and Prince (1988) noted that though the verb *perambulate* may be of low frequency, it is no worse-sounding in its past tense form than it is in its stem form; there is no feeling that *perambulated* is a worse past tense form of *perambulate* than *walked* is of *walk.* In fact, a verb can be of essentially zero frequency and still have a regular past tense form that is judged as no worse than the verb itself. Though *fleech, fleer,* and *anastomose* are unknown to most speakers, speakers judge *fleeched, fleered,* and *anastomosed* to be perfectly good as the past tense forms of those verbs. These observations have been confirmed experimentally by Ullman (1993), in a study in which people judged the naturalness of hundreds of verbs and their past tense forms. Subjects' ratings of regular pasts correlate highly with their ratings of the corresponding stems, but not with the frequency of the past form (partialing out stem rating). In contrast, ratings of irregular pasts correlate less strongly with their stem ratings but significantly with past frequency, partialing out stem rating.

Unlike irregular verbs, when a regular verb gets trapped in a frozen or restricted expression, putting it into the past tense makes it no worse. For

example, the verb *eke* is seldom used outside contexts such as *She ekes out a living*, but *She eked out a living*, unlike *forwent the pleasure of*, does not suffer because of it. Similarly: *He crooked his finger; She stinted no effort; I broached the subject with him; The news augured well for his chances.* The regular verb *to afford*, like the irregular verb *to stand*, usually occurs as a complement to *can't*, but when liberated from this context its past tense form is perfectly natural: *I don't know how she afforded it*. Similarly, both *She doesn't suffer fools gladly* and *She never suffered fools gladly* are acceptable.

The phenomena discussed in this section and the preceding one show why the apparent gradedness of acceptability for regular forms like *pleaded* or *weeped* can be localized to the gradedness of the corresponding irregulars because of the effects of the Blocking principle and are not inherent to the regular verbs per se. The gradedness of certain irregulars generally comes from low frequency combined with similarity to the prototypes of their subclasses (Ullman, 1993). But for regular verbs that do not compete with specific irregular roots, there is no complementary landscape of acceptability defined by phonology and frequency; all are equally good.

3. Default Structure

As we have seen, the regular past tense alternation can apply regardless of the stem's:

- Phonological properties
- Verb-Root versus Non-Verb-Root Status
- Frequency
- Listedness (Familiarity)
- Range of Contexts

Apparently, the regular class is the *default* class. More generally, there is a sense in which the category of regular verbs has no properties; it is an epiphenomenon of the scope of application of the regular rule.

Conclusions about the Regular Class

These phenomena invite the following conclusion: The class of regular verbs in English is a classical category. Its necessary and sufficient conditions are simply the conditions of application of the regular rule within English grammar. Those conditions for membership can be stated simply: a verb, unless it has an irregular root. The category has no other properties.

Psychological Implications

We have shown that by the standard criteria the irregular subclasses are prototype or family resemblance categories, and the regular class is a classical category. If we take this conclusion seriously, it has several immediate implications.

Psychological Reality

First, both family resemblance categories and classical categories can be psychologically real and natural. Classical categories need not be the product of explicit instruction or formal schooling: the regular past tense alternation does not have to be taught, and indeed every child learns it and begins to use it productively in the third year of lie (Marcus, Pinker, Ullman, Rosen, Hollander, & Xu, 1992). The fact that children apply the regular alternation even to high-frequency irregular stems such as *come* and *go*, which they also use with their correct irregular pasts much of the time, suggests that children in some way appreciate the inherently universal range of the regular rule. And like adults, they apply the regular suffix to regular verbs regardless of the degree of the verbs' similarity to other regular verbs (Marcus, et al., 1992), and to irregular-sounding verbs that are derived from nouns and adjectives (Kim, et al., 1994). Gordon (1985) and Stromswold (1990) have shown that children as young as three make qualitative distinctions between regular and irregular plural nouns related to their different formal roles within the grammar, without the benefit of implicit or explicit teaching inputs (see Marcus, et al., 1992, and Kim, et al., 1994 for discussion).

The regularization-through-derivation effect (*flied out, high-sticked*) provides particularly compelling evidence that classical categories do not have to be the product of rules that are explicitly formulated and deliberately transmitted. The use of the regular rule as a default operation, applying to any derived verb regardless of its phonology, is a grass-roots phenomenon whose subtleties are better appreciated at an unconscious level by the person in the street than by those charged with formulating prescriptive rules. Kim, et al. (1991) found that non-college-educated subjects showed the effect strongly, and in the recent history of English and other languages there are documented cases in which the language has accommodated such regularizations in the face of explicit opposition from editors and prescriptive grammarians. For example, Mencken (1936) notes that the verb *to joy-ride*, first attaining popularity in the

1920's, was usually given the past tense form *joy-rided*, as we would predict given its obvious derivation from the noun *a joy-ride*. Prescriptive grammarians unsuccessfully tried to encourage *joy-rode* in its place. Similarly, Kim, et al. (1994) showed that children display the effect despite the fact that most have rarely or ever heard regularized past tense forms for irregular-sounding verbs in the speech of adults.

On the other side, family resemblance categories are not necessarily artifacts of reaction time studies or rating studies, as Fodor (1981) and Armstrong, et al. (1983) have suggested. Children generalize family resemblance patterns of irregular subclasses to inappropriate regular and irregular verbs in their spontaneous speech, as in *brang* for *brought* and *bote* for *bit*, and their generalizations appear to be sensitive to the frequency and family resemblance structure of the subclasses (Xu & Pinker, 1995; Bybee and Slobin, 1982a; Rumelhart and McClelland, 1986; Pinker & Prince, 1988). The irregular subclass structure also affects dialectal variation and historical change in adult speech (Bybee and Slobin, 1982b; Mencken, 1936; Prasada & Pinker, 1993) with new irregular forms occasionally entering the language if their stems are sufficient similar to existing irregular stems.

Psychological Function

A further corollary is that classical categories and family resemblance categories do not have to have different psychological functions such as careful versus casual reasoning, or reasoning versus categorization of exemplars. What is perhaps most striking about the contrast between the regular and irregular verbs is that two kinds of entities live side-by-side in people's heads, serving the same function within the grammar as a whole: regular and irregular verbs play indistinguishable roles in the syntax and semantics of tense in English. There is no construction, for example, in which a regular but not an irregular verb can be inserted or vice-versa, and no systematic difference in the temporal relationships semantically encoded in the past tense forms of regular and irregular verbs.

More specifically, it is difficult to make sense of the notion that family resemblance categories are the product of a set of identification procedures used to classify exemplars as belonging to core categories with a more classical structure. The suggestion that "irregulars are used in perceptually categorizing members of the regular class" is uninterpretable. The irregulars are a class of words that display one kind of category structure; the regulars do not display it.

Perhaps a closer analogy would be between membership conditions for the irregular subclasses and the operation on the stem that generates the past tense form. One might say that a family resemblance structure characterizes the membership of each subclass, but once an item is a member (for whatever reason), it is transformed into a past tense form by a classical all-or-none operation such as laxing the vowel. But even here, the core-identification distinction does not easily apply, because the changes that the member stems of a class undergo, and not just the properties of the stems, have a heterogeneous structure. Within the subclass of irregulars ending in *ing/ink*, *sing* goes to *sang* while *sting* goes to *stung* and *bring* goes to *brought*. Similarly, within the subclass that adds a [d] to the past tense form some verbs have their vowel laxed (e.g. *hear/heard*, some have their final consonant deleted (e.g. *make/made, have/had*, some undergo the *e - o* ablaut that is frequent across the various subclasses (e.g. *sell/sold*), and one undergoes a unique vowel change (*do/did*). Furthermore, within a class there can be graded differences of acceptability among different possible change operations: *?spit/spat* for *spit, ?begot/??begat* for *beget, bid/?bade* for *bid*. In sum, both the membership conditions and the operations of the irregular subclasses display family resemblance category effects. Later we will show that the core/identification distinction does not work well for conceptual categories either.

Underlying Psychological Mechanism

Though classical and family resemblance categories, in the case of the past tense, do not differ in psychological function — what they are used for), they do differ in psychological structure — what mental processes give rise to them. Our main claim is that the psychological difference between regulars and irregulars is a fundamental one, and is of a piece with the psychological difference between classical and family resemblance categories in general, including conceptual categories.

As we have seen, the classical category consisting of regular verbs is defined completely and implicitly by the nature of a rule in the context of a formal system, in this case, a rule within English grammar that applies to any word bearing the part-of-speech symbol "verb" unless it has an irregular root. The category is not a generalization or summary over a set of exemplars; indeed, it is blind to the properties of the exemplars that fall into the category. It falls out of the combinatorial rule system that allows humans to communicate propositions (including novel, unusual, or abstract propositions) by building complex words, phrases, and sentences in which the meaning of the whole is determinable by the meanings of the parts and the way in which they are combined.

Family resemblance categories, in contrast, are generalizations of patterns of property correlations within a set of memorized exemplars. Consequently, factors that affect human memory affect the composition of the irregular class. A well-known example is word frequency. Irregular verbs tend to be higher in frequency than regular verbs (Ullman, 1993; Marcus, et al., 1995), and if an irregular verb's frequency declines diachronically, it is liable to become regular (Hooper, 1976; Bybee and Slobin, 1982b; Bybee, 1985). Presumably this is because irregulars are memorized; to memorize an item one has to hear it; if opportunities for hearing an item are few, its irregular form cannot be acquired and the regular rule can apply as the default. This is also presumably the cause of the fuzziness of the past tenses of irregular verbs that are used mainly in nonpast forms, such as *forgo* or the idiomatic meanings of *stand* or *become*.

A related account could help explain the genesis of the family resemblance structure of the irregular verbs. Rosch and Mervis (1975) found that people find lists of strings that display family resemblance structures easier to remember than lists of strings with arbitrary patterns of similarity. Just as frequency affects the memorizability, hence composition, of the irregular subclasses, so might family resemblance structure. The current subclasses may have emerged from a Darwinian process in which the irregular verbs that survived the generation-to-generation memorization cycle were those that could be grouped into easy-to-remember family resemblance clusters.

In sum, the properties of the regular and irregular classes of verbs in English show that both classical categories and family resemblance categories can be psychologically real, easily and naturally acquired, and not subject to a division of labor by function along the lines of reasoning versus identification of exemplars. Rather, they differ because they are the products of two different kinds of mental processes: a formal rule system, and a memorized partially-structured list of exemplars. We now point out two less obvious conclusions based on properties of the regular and irregular classes: classical and prototype categories are suited to different kinds of computational architectures, and the mental mechanisms giving rise to classical and family resemblance categories are suited to representing inherently different kinds of entities in the world. Finally, we return to human conceptual categories like "bird" and "mother," seeing whether we can gain insight by generalizing our findings about classical and prototype categories.

Computational Architecture

The acquisition of English past tense morphology has recently been implemented in a computer simulation model by Rumelhart and McClelland (1986). The architecture of the simulation, its behavior, and its fidelity to human data have been discussed in detail (Pinker & Prince, 1988, 1992; Lachter and Bever, 1988; Sproat, 1992; Prasada & Pinker, 1993; Marcus, et al., 1992, 1995).

The RM model makes use of a device called a "pattern associator". This device is paradigmatic of Parallel Distributed Processing (PDP) or Connectionist architectures that are currently a central topic of debate in cognitive science (Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986; Pinker & Mehler, 1988). Our discussion will refer to pattern associators in particular, not to all PDP architectures.

A pattern associator is a feedforward two- or three-layer network that is designed to take an input representation and map it onto an output representation; in this case, a stem and a past tense form, respectively. There is a set of input nodes, each corresponding to a possible property of an input form. An input is encoded by a separate device, an encoder, that dissolves it into its properties: a node is turned on corresponding to each property that the input possesses, and the unique identity of the input item itself is not registered on any node (this is sometimes called "distributed representation"). Likewise, the pattern associator has a set of output nodes, each representing a property of the output form when activated. Every input node is connected to every output node by a weighted link. When a set of input nodes is activated, each node sends its activation level, multiplied by the link weight, to the output nodes it is connected to. Each output node sums its weighted inputs, compares the result to a threshold, and turns on if the threshold is exceeded, with a likelihood related to the difference.

The model "learns" by adjusting the weights on its links. During "training" trials, an input is provided to the model, it is allowed to generate an output, and the output pattern is compared to the target "correct" output pattern provided by a "teacher". The discrepancies between actual and target outputs are registered,

and the weights on the links leading to the discrepant output nodes are adjusted to lessen the likelihood of such an error when similar input patterns are presented in the future. For an output node that is off but should be on, the weights on the links from currently active input links are incremented and the node's threshold is lowered. For an output that is on but should be off, the weights on links from active input nodes are lowered, and the threshold on the node is raised. The procedure is repeated for other inputs; node adjustments from different inputs are superimposed in a single set of link weights, which represent the aggregate correlational structure between properties of the inputs and properties of the outputs in the training set.

Two properties of pattern associators are crucial in understanding their behavior: items are represented by their properties, and statistical contingencies between every input property and every output property across a set of items are recorded and superimposed.

Before being applied to the case of learning past tense forms, pattern associators had been studied in detail, including their ability to learn and identify members of conceptual categories (McClelland & Rumelhart, 1985) and they are known to do certain things well. They can often reproduce a set of associations in a training set, and generalize to new cases based on their similarity to existing ones. They are sensitive to input pattern frequencies in ways similar to humans. Furthermore they reproduce many of the effects displayed by people when dealing with family resemblance categories. McClelland & Rumelhart (1985) and Whittlesea (1989) have devised pattern associators that are fed patterns of data concerning properties of a set of nonlinguistic objects. They found that the models do fairly well at duplicating the effects of frequency, prototypicality, family resemblance, gradations of membership, and influence of particular exemplars on human classification times and error rates. Since such effects are known to be related to cooccurrence frequencies among objects' features (Smith & Medin, 1981), this is not surprising.

Thanks to these abilities, the pattern associator that Rumelhart and McClelland applied to learning past tense forms handled the irregular verbs with some success. The model was fed a set of 420 verbs (each one presented as a pair consisting of its stem and its past form), including 84 irregular verbs, about 200 times each. Following this training it was able to approximate the past tense forms for all of them given only the stem as input. Furthermore, it was able to generalize to new irregular verbs by analogy to similar ones in the training set, such as *bid* for *bid*, *clung* for *cling*, and *wept* for *weep*. In addition, it showed a tendency to extend some of the subregular alternations to regular verbs based on their similarity to irregulars, such as *kid* for *kid* and *slept* for *slip*, showing a sensitivity to the family resemblance structure of the irregular subclasses. Finally, its tendencies to overgeneralize the regular *d* ending to the various irregular subclasses is in rough accord with children's tendencies to do so, which in turn is based on the frequency and consistency of the vowel changes that the verbs within each subclass undergo (Pinker & Prince, 1988; Sproat, 1992).[2]

However, pattern associators do not seem to perform as well for other kinds of mappings. In particular, they are deficient in handling regular verbs. For one thing, their uniform structure, in which regulars and irregulars are handled by a single associative mechanism, provides no explanation for why the regular class has such different properties from the irregular classes; it falsely predicts that the regular class should just be a larger and more general prototype subclass.

Moreover, the pattern associator fails to acquire the regulars properly. Pinker & Prince (1988) pointed out that the model is prone to *blending*. Competing statistical regularities in which a stem participates do not block each other, they get superimposed. For example, the model produced erroneous forms in which an irregular vowel change was combined with the regular ending, as in *sepped* as the past of *sip* or *brawned* for *brown*. It would often blend the *t* and *id* variants of the regular past tense form, producing *stepted* for *step* or *typted* for *type*. Sometimes the blends are quite odd, such as *membled* for *mailed* or *toureder* for *tour*.

Furthermore, Pinker & Prince noted that in contrast to the default nature of the regular rule, the RM model failed to produce any past form at all for certain verbs, such as *jump, pump, glare,* and *trail*. Presumably this was because the model could not treat the regular ending as an operation that was capable of applying to any stem whatsoever, regardless of its properties; the ending was simply associated with the features of the regular stems encountered in the input. If a new verb happened to lie in a region of phonological space in which no verbs had previously been supplied in the training set (e.g. *jump* and *pump*, with their unusual word-final consonant cluster), no coherent set of output features was strongly enough associated with the active input features, and no response above the background noise could be made. Pinker & Prince's diagnosis was tested by Prasada and Pinker (1993) who presented typical-sounding and unusual-sounding verbs to the trained network. For the unusual-sounding items, it produced odd blends and chimeras such as *smairf-sprurice, trilb-treelilt, smeej-leefloag,* and *frilg-freezled*.

The model is inconsistent with developmental evidence. Children first use many irregulars properly when they use them in a past tense form at all (e.g. *broke*, then begin to overregularize them occasionally (e.g. *broke* and *breaked*) before the overregularizations drop out years later. Since pattern associators are driven by pattern frequency, the only way the RM model could be made to duplicate this sequence was first to expose it to a small number of high-frequency verbs, most of them irregular, presented a few times each, followed by a large number of medium-frequency verbs, most of them regular, presented many times each. Only when the model was swamped with exemplars of the regular pattern did it begin to overregularize verbs it had previously handled properly. However, the onset of overregularization in children is not caused by a sudden shift in the proportion of regular verbs in the speech they hear from their parents: the proportion remains largely unchanged before, during, and after the point at which they begin to overregularize (Pinker & Prince, 1988; Slobin, 1971; Marcus, et al., 1992). Nor is it caused by a rapid increase in the proportion of verbs in their vocabulary that is regular; the percentage of children's vocabulary that is regular increases quickly when they are *not* overregularizing, and increases more slowly when they *are* overregularizing (Marcus, et al., 1992).

The results support the traditional explanation of overregularization, which appeals not to frequency but to different internal mechanisms: children at first memorize irregular and regular pasts, then they discover that a regularity holds between many regular stems and their past forms and create a rule which they apply across the board, including instances in which a memorized irregular form does not come to mind quickly enough; the rule is available to fill the gap, resulting in an overregularization. Consistent with this interpretation, Marcus, et al. (1992) found that children begin to overregularize at the age at which they first start using *regular* forms consistently in the past tense; that, presumably, is the point at which the regular rule has been acquired. As mentioned, the fact that the regular rule is applied even to high-frequency irregular stems, which remain high in frequency in children's input throughout development, shows that children treat the regular rule as having an unlimited range.

Proponents of connectionist models of language have offered two kinds of counterarguments, but both are inadequate. One is that the RM model was a two-layer perceptron, and that threelayer models, whose hidden layer's weights are trained by error-back-propagation, perform much better (see, e.g., Plunkett & Marchman, 1991, 1993; MacWhinney & Leinbach, 1992). However, Sproat

(1992), Prasada & Pinker (1993), and Marcus (1995) have shown that hidden-layer models have the same problems as the original RM model. The other is that the effects of regularity in English come from the fact that regular verbs are in the majority in English, fostering the broadest generalization. German presents the crucial comparison. Marcus, et al. (1995) reviewed the grammar and vocabulary statistics of German in detail, and documented that the participle *-t* and plural *-s* are found in a *minority* of words in the language, compared to irregular alternatives, but nonetheless apply in exactly the "default" circumstances where access to memorized verbs or their sounds fails, including novel, unusual-sounding, and derived words (i.e., the *flied-out* examples have exact analogues in German). The findings were verified in two experiments eliciting ratings of novel German words from German adults. The cross-linguistic comparison suggests that default suffixation is not due to numerous regular words reinforcing a pattern in associative memory, but to a memory-independent, symbol-concatenating mental operation.

In sum, pattern associators handle irregular subclasses reasonably well, but handle the regular class poorly, both in terms of computational ability and psychological fidelity. We suggest that this is a symptom of their relative suitability of this architecture to handle family resemblance and classical categories in general. The reasons, we suggest, are straightforward:

- Classical categories are the product of formal rules.
- Formal rules apply to objects regardless of their content — that is what "formal rule" means.
- Pattern associators soak up patterns of correlation among objects' contents — that is what they are designed to do.
- Therefore, pattern associators are not suited to handling classical categories.

We conclude that the brain contains some kind of non-associative architecture, used in language, and presumably elsewhere.

Epistemological Categories versus Ontological Categories

Rey (1983) has pointed out that even if people can be shown to use prototype (or classical) categories, it doesn't mean that the world contains prototype (or classical) categories — that is, that the lawful generalizations of how the world works, as captured by the best scientific description, make

reference to one kind of category or the other. That raises a question: if there is a psychological distinction between the representation of prototype and classical categories, is it because these representations accurately reflect different kinds of categories in the world? Or does the human system of categorization arise from some limitation or quirk of our neurological apparatus that does necessarily correspond to the lawful groupings in the world?

The question of what kinds of categories are in the mind and what kinds of categories are in the world are clearly related. If the mind evolved to allow us to grasp and make predictions about the world, the mental system that forms conceptual categories should be built around implicit assumptions about the kinds of categories that that the world contains, in the same way that a visual algorithm for recovering structure from motion might presuppose a world with rigid objects and might work best in situations where the assumption is satisfied.

Because the English past tense system shows classical and family resemblance categories, but must have a very different ontology from that underlying concepts of tools, vegetables, animals, and other entities that ordinarily compose conceptual categories, an analysis of the source of classical and family resemblance categories in the past tense system may help us to identify the distinctive conditions in which these two kinds of categories arise.

Where do The Properties of the Regular and Irregular Classes Come From?

The properties of the regular class are simply products of the regular rule. From any speaker's perspective, the class exists "in the world" in the sense that other speakers of the language possess the rule and use it in speaking and understanding. This in turn comes from the basic requirement for parity in any communicative system. Language can only function if the rule system that generates forms is shared by a community of speakers. Thus one person's use of a past tense rule (or any rule) in production presupposes that that same rule is in the head of the listener and will be used to interpret the produced form. Similarly, use of a rule in comprehension presupposes that the speaker used it in programming his speech. So the answer to the question "What class of entities in the world is picked out by a rule-generated class, such as the regular verbs?" is "The class of entities that can be generated by a replica of that rule in other speakers' minds."

For the irregulars, the issue is more complex. Of course, irregulars, like regulars, are usable only because they are shared by other speakers. But unlike the case of regulars, where the rule is so simple and efficient that it naturally fits into a grammar shared by all members of a community, the composition of the irregular class is so seemingly illogical that one must ask how other speakers came to possess it to begin with.

In a previous section we suggested that the family resemblance structure of the irregular past tense subclasses is related to the fact that irregulars must be memorized and human memory has an easier time with family resemblance categories (Rosch and Mervis, 1975). Interestingly, the obvious Darwinian metaphor in which the most easily memorized verbs survive does not apply to the psychology of of the child doing the learning. Note that family resemblance structure is not a property that some individual verbs have and others lack, but a property of *an entire class* of verbs. But unlike the subjects of Rosch & Mervis's experiment, children are not given two classes to learn, one with a random organization, the other with a family resemblance structure, with the latter being better retained in memory. One might suppose that the similarity of a verb to other verbs affects how easy it is for the child to memorize that verb, and that in the aggregate, a family resemblance structure arises. But this, too, does not properly characterize the acquisition of irregular forms. There is relatively little change in the composition of the subclasses between one generation and the next; children end up pretty much learning the same irregulars that their parents learned. Moreover, if the childrens' memory really shaped their irregular classes, we would expect them to arrive at classical categories, not family resemblance categories. For example, a rule that said "all verbs ending in *ing* go to *ang*" would have much higher inter-item similarity than the current English *ing* class, so verbs like *bring* would be even easier to memorize. In fact, given children's ability to regularize the irregular verbs by assimilating them to the regular rule (*bringed*) or generalizing a subregularity (*brang*), children would be in a position to obliterate irregularity altogether if there memory were all that fragile.

A more accurate version of the Darwinian metaphor would point to effects of memory not in the child doing the learning in a given generation but in the children (and adults) of previous generations whose learning shaped the input to the current generation. Even though each generation reproduces the previous generation's irregulars with high accuracy, changes occasionally creep in. These can be characterized as a kind of convergent evolution toward certain attractor states. For example, some lower-frequency irregular verbs may be consistently regularized in a given generation, and this might be more likely for verbs that were are most dissimilar from other irregulars and hence most weakly protected

from forgetting. (Marcus, et al., 1992, documented that irregulars that are more dissimilar from other irregulars are more prone to being overregularized by children.) In the other direction, some regulars might be attracted into an irregular class because of their high similarity to existing irregulars, as is happening with *sneak-snuck* (cf. *stick-stuck, string-strung,* etc.). If some of these occasional forgettings and analogies get fixed in a language community in a contagion-like process (see Cavalli-Svorza and Feldman, 1981) and accumulate across generations, classes of verbs with a family resemblance structure can arise. The past tense forms *quit* and *knelt*, for example, are fairly recent additions to the language, and are presumably irregular because of their similarity to verbs like *hit* and *feel*. This process can be seen even more clearly in the more rapid process of dialect formation in smaller communities, where forms such as *bring-brang, slide-slud,* and *drag-drug* are common (see Mencken, 1936).

Though this convergent evolution process surely occurs, it cannot explain the entire structure of the irregulars in English. First, it does not capture the historical facts completely. The language never contained arbitrary irregular classes whose members were attracted into or drifted out of prototype classes because of fussy learners, leaving the next generation with a slightly more orderly class than they had found. Rather, as we shall see, the strong subclasses are in evidence from the earliest sources. Second, the account posits a kind of harmony between properties of the memory of one generation and properties of the memory of succeeding generations: the errors of forgetting and assimilation of generation n result in a stimulus set that is easier for generation $n + 1$ to acquire without error, because the memory of both generations is biased towards remembering items that are similar along multiple dimensions to other items. But in doing so it just begs the question of why memorization of categories in any generation should be biased toward partial similarities to begin with. Why does memory work that way? Why not do away with remembering patterns of irregularity altogether and give the next generation a nice regular class?

In the history of English, *divergence* has been the more prominent trend. That confronted learners in each generation with the task of learning classes whose family resemblance structure was not simply caused by the psychology of previous generations of learners. In Old English, there were seven "strong" past tense classes, in which the vowel of the stem was altered, and three "weak" classes, in which a suffix containing *d* was added, which sometimes caused a modification of the stem vowel for phonological reasons. Most of the modern irregulars derived from verbs in the strong classes. The modern regular rule, and most of the irregulars that end in *t* or *d* such as *meant* and *made*, evolved from the weak classes. The Old English strong classes had themselves evolved out of classes that can be traced back to Proto-Germanic, and before that, to Proto-Indo-European. Many scholars believe that the proto-Indo-European classes were defined by regular rules: the number and type of segments following the vowel within the stem determined the kind of change the vowel underwent (Johnson, 1986; Bever and Langendoen, 1963; Prokosch, 1939; Campbell, 1959). By the time of Old English the patterns are more complicated, but they were still more pervasive and productive and tolerated fewer arbitrary exceptions than the alternations in the modern English irregular subclasses. That is, many stems that are now regular but fit the characteristic pattern of an irregular subclass in fact used to undergo the irregular change: *deem/dempt, lean/leant, chide/chid, seem/sempt, believe/beleft, greet/gret, heat/het, bite/bote, slide/slode, abide/abode, fare/fore, help/holp,* and many others. Furthermore, there was a moderate degree of productivity within the classes (Johnson, 1986).

Beginning in the Middle English period, there was an even greater decline in the productivity and systematicity of the past tense subclasses, with the exception of one of the weak suffixing processes. The main causes were the huge influx of new words from Latin and French that needed a general, condition-free past tense operation, and the widespread shifts in vowel pronunciation that obscured regularities in the vowel-change operations. The weak suffixing operation was already being used for verbs derived from nouns in Old English, which did not fit the sound patterns defining the strong classes of verbs, so their extension to borrowed words was natural (see Marcus, et al., 1995 for further discussion).

In sum, there has been a consistent trend in the history of English since the Proto-Indo-European period for for the strong classes, originally defined by phonological properties of their stems, to became lists of items to be learned individually. This had an interesting consequence. Originally, lists would have been relatively homogeneous, owing to their once having been generated by rule-like operations. But then, a variety of unrelated processes, operating on individual items, destroyed the homogeneity of the classes. Here are some examples:

Phonological Change: *Blow, grow, throw, know, draw, fly, slay* all begin with <Consonant-Sonorant> cluster except for *know*. The reason that *know* is exceptional leaps from the page in the way it is spelled. As it was originally pronounced, with an initial *k*, it did fit the pattern; when syllable-initial *kn* mutated to *n* within the sound pattern of the language as a whole, *know* was left stranded as an exception within its subclass.

Morphological Category Collapse: In Old English, past tenses were distinguished by person and number. For example, *sing* had a paradigm which can we simplify as follows:

(5)

	Singular	**Plural**
1st	*sang*	*sung*
2nd	*sung*	*sung*
3rd	*sang*	*sung*

When the number distinctions collapsed, each verb had to pick a form for its past tense as if playing musical chairs. Different verbs made different choices; hence we have *sing/sang/sung* alongside *sling/slung/slung*. The contrast between *freeze/froze* and *cleave/cleft* has a similar cause.

Attrition: Earlier, the class in which *t* changed to *d* had the following members: *bend, lend, send, spend, blend, wend, rend, shend, build, geld, gild, gird* (Bybee & Slobin, 1982b). The class is succinctly characterized as containing a vowel followed by a sonorant followed by *d*. In modern American English, the verbs *geld, gird, gild, wend, rend* and *shend* are now obsolete or obscure. The residue of the class has 5 members, 4 rhyming with *end* and 1 with *ild*. Although logically it still can be characterized as ending in a vowel-sonorant-*d* cluster, the presence of regular verbs ending in *eld* and *ird*, and the highly specific nature of the rhyme with *end*, makes it more natural to represent the class as containing verbs that rhyme with *end* but with one exception.

Idiosyncratic pronunciation shifts: In Old and Middle English, many verbs had variant pronunciations. If past tense forms are simply associated with their stems, rather than being generated from them by a rule, it should be possible for one pronunciation to drift out of the language while its corresponding past tense form survives. For example, *run* could once be pronounced as *rin*. The forms *ran* and *has run* would thus fit perfectly into the pattern *sing/sang/has sung*, *ring/rang/has rung*, and so on. *Rin* is no longer possible in standard English, but *ran/has* run has survived. Similarly, *spit* used to have the pronunciations *spete* and *spitte*. The former yields the past tense form that became *spat*, the latter the form that became *spit*. Both past tense forms survive as fuzzy or marginal exemplars in American English, but of course only *spit* survives as its stem.

We conclude that a class of items that originally is homogeneous on account of its being generated by a rule can acquire a family resemblance structure by divergent evolution once the rule ceases to operate and the effects of unrelated processes acting on individual members accumulates through history. Superimposed on these patterns is a convergent process in which the accumulated effects of the analogizing and forgetting tendencies of previous generations of learners cause partly similar forms to accrete onto an existing class. Thus a learner in a single generation is confronted with family resemblance structures as products of these divergent and convergent historical processes, and these structures can be said to exist in the world independent of his or her psychology.

Implications for Conceptual Categories

We have suggested that classical and family resemblance categories can be found in a surprising realm, English past tense forms, and that within it the two kinds of categories have distinct linguistic properties, psychological representations, underlying computational architectures, and real-world counterparts. Do these discoveries offer insight into the role of classical and family resemblance categories in the domain of conceptual categories like birds and mothers? Perhaps the best way to start would be to consider what conceptual categories are for.

The Function of Conceptual Categories: Inference of Unobserved Properties

No two objects are exactly alike. So why do we use conceptual categories? Why don't we treat every object as the unique individual that it is? And why do we form the categories we do? Why lump together salmon, minnow, and sharks, as opposed to sharks, leaves, and spaghetti? These are elementary questions, but possible answers to them have not informed research on conceptual categories as much as would be desirable. Often it is suggested that people need categories to reduce memory or processing load, but given that the cortex has on the order of a trillion synapses and that long term memory is often characterized as "infinite," the suggestion carries little force. Furthermore for many categories (e.g., months, baseball teams, one's friends) both the category and every individual member of it are stored in memory. Rey (1983) provides a list of the main functions that concepts are supposed to perform, including stability of concepts at different times in a given individual or at the same time for different individuals, the ability of a concept to serve as the basis for a word's meaning, the basis for things to belong to categories in the world, and the basis for people to know which things belong to which categories in the world. But none of the functions had anything to do with why we form conceptual categories at all, or why some categories in the world are natural bases for concepts and others unnatural.

Bobick (1987), Shepard (1987), and Anderson (1990) have attempted to reverse-engineer human conceptual categories, seeking principles motivating the choice of particular representations for concepts in terms of their function in people's dealings with the world. They have independently proposed that categories are useful because they allow us to infer objects' unobserved properties from their observed properties (see also Rosch, 1978, and Quine, 1969.) Though we being cannot know everything about an object, we can observe some things; the observed properties allow us to assign the object to a category, and the structure of the category then allows us to infer the values of the objects' unobserved properties. Categories at different levels of a hierarchy (e.g., cocker spaniels, dogs, mammals, vertebrates, animals, living things) are useful because they allow a variety of tradeoffs between the ease of categorization and the power of the licensed inference. For low-level, specific categories, one has to know a lot about the object to know that it belongs in the category, but one can then infer many unobserved aspects of the nature of the object. For high-level, general categories, one need know only a few properties of an object to know it belongs to the category, but one can infer only a few of its unobserved properties once it is thus categorized.

To be concrete: knowing that Peter is a cottontail, we can predict that he grows, breathes, moves, was suckled, inhabits open country or woodland clearings, spreads tularemia, and can contract myxomatosis. If we knew only that that he was a mammal, the list would include only growing, breathing, moving, and being suckled. If we knew only that he was an animal, it would shrink to growing, breathing, and moving. On the other hand, it's much harder to tag Peter as a cottontail than as a mammal or an animal. To tag him as a mammal we need only notice that he is furry and moving, but to tag him as a cottontail we have to notice that he is long-eared, shorttailed, long-hind-legged, and has white on the underside of his tail. To identify *very* specific categories we have to examine so many properties that there would be few left to predict. Most of our everyday categories are somewhere in the middle: "rabbit," not mammal or cottontail; "car," not vehicle or Ford Tempo; "chair," not furniture or Barcalounger. They represent a compromise between how hard it is to identify the category and how much good the category does. These compromises correspond to Rosch's (1978) notion of the "basic level" of a category.

We can get away with inductive leaps based on categories only because the world works in certain ways. Objects are not randomly distributed through the multidimensional space of properties that humans are interested in; they cluster in regions of cooccurring properties that Bobick calls "natural modes" and Shepard calls "consequential regions." These modes are the result of the laws of form and function that govern the processes that create and preserve objects. For example, the laws of geometry dictate that objects formed out of multiple parts have concavities at the part boundaries. The laws of physics dictate that objects denser than water will be found on lake bottoms rather than lake surfaces. Laws of physics and biology dictate that objects that move quickly through fluid media have streamlined shapes, and bigger objects tend to have have thicker legs. Knowing some of the coordinates of an object in property space, the existence of natural modes allows us to infer (at least probabilistically) some of its unknown coordinates.

Classical Categories: Inferences within Idealized Lawful Systems

All this raises the question of what kinds of regularities in the world generate natural modes that humans can exploit by forming concepts. In the most general sense, regularities in the world are the result of scientific and mathematical laws (e.g., of physics, geometry, physiology). Laws can be captured in formal systems, given a suitable idealization of the world. By "formal system" we mean a symbol manipulation scheme, consisting of a set of propositions and a set of inference rules that apply to the propositions by virtue of their form alone, so that any knowledge not explicitly stated in the propositions cannot affect the inferences made within it. Formal systems, we suggest, are the contexts in which classical categories are defined. Therefore, under whatever idealization of the world a set of scientific or mathematical laws applies, the world contains classical categories. For example, when the texture, material, thickness, and microscopically ragged edges of real-world objects are provisionally ignored, some can be idealized as plane geometry figures. Under this idealization, objects with two equal sides can be assigned to the category "isosceles triangle". Once the object is assigned to that category, one can make the inference that it also has two equal angles, among other things. Frictionless planes, ideal gases, randomly interbreeding local populations, and uniform communities of undistractable speaker-hearers are other idealizations under which regularities in the behavior of objects can be captured in formal systems. A

smart organism could use formal systems as idealizations of the world to infer unknown properties from known ones. In the psychology of categorization, no less than in the history of science, idealization or selective *ignoring* of salient correlational structure is crucial to apprehending causal laws.

We suggest, then, that wherever classical categories are to be found in human cognition, they will be part of a mentally represented formal system allowing nontrivial deductions to be made. Given the function of concepts, why else would one bother to assign an object to a classical category? What is unnatural, then, about traditional experiments in concept formation, such as those of Hull (1920), Bruner, Goodnow, & Austin (1956), and Hunt (1962), in which subjects learn categories like "red square with two borders," is not that the categories have sharp boundaries or necessary and sufficient conditions, but that the categories are not part of a system allowing interesting inferences to be drawn — they are unnatural because they are literally useless.

Though one tends to think of formal systems as the province of systematic education in modern societies, there are a variety of kinds of formal systems capturing inference-supporting regularities that could be accessible to people, including those in preindustrial and preagricultural societies. For example, bodies of folk science need not resemble their counterparts in modern scientific systems, but they can reproduce some of their visible predictions with alternative means. Mathematical intuitions too are incorporated into many other systems of common knowledge. Here are some examples:

- Arithmetic, with classical categories like "a set of 3 objects," supporting inferences like "cannot be divided into two equal parts," independent of the properties of objects that can be grouped into threes.

- Geometry, with classical categories like "circle," supporting inferences like "all points equidistant from the center" or "circumference is a constant multiple of diameter," regardless of whether previously encountered circles are sections of tree trunks or drawings in sand.

- Logic, with classical categories like "disjunctive proposition," supporting inferences like "is true if its second part is true" or "is false if the negations of both its parts are true".

- Folk biology, with classical categories like "toad of kind x," which support inferences like "extract of mouth gland when boiled and dried is poisonous," regardless of its similarities to nonpoisonous toads or its dissimilarities to other poisonous toads.

- Folk physiology, with the famous all-or-none category "pregnant," supporting the inferences "female," "nonvirgin," and "future mother," regardless of weight or body shape.

In addition, the world of humans contains other humans, and there is reason to expect mentally-represented formal systems to arise that govern the conduct of humans with one another. Given the fuzziness and experience-dependent individual variation inherent to family resemblance categories, it is not surprising that conflicts of interest between individuals will often be resolved by reasoning within systems that have a classical structure, allowing all-ornone decisions whose basis can be agreed to by all parties. There is a rationale to assigning drinking privileges to people after their twenty-first birthday, arbitrary thought that is, rather than attempting to ascertain the emotional maturity of each individual when he or she asks for a drink. Furthermore, Freyd (1983) and Smolensky (1988) have suggested that certain kinds of socially transmitted knowledge are likely to assume the form of discrete symbol systems because of constraints on the channels of communication with which they must be communicated between individuals and transmitted between generations. It is not hard to identify formal systems involved in social interactions that define classical categories:

- Kinship, with classical categories like "grandmother of X," supporting inferences like "may be the mother of X's uncle or aunt" or "is the daughter of one of X's great-grandparents," regardless of hair color or propensity to bake muffins.
- Sociopolitical structure, with classical categories like "president" or "chief," supporting inferences like "decisions on entering wars are carried out," regardless of physical strength, height, sex, and so on.
- Law, with classical categories like "felon," supporting inferences like "cannot hold public office," regardless of presence or absence of a sinister appearance, social class, and so on.
- Language, with the category "verb," supporting the inference "has a past tense form suffixed with *d* unless it has an irregular root," regardless of its phonological properties.

It is unlikely to be a coincidence that humans uniquely and nearly universally have language, counting systems, folk science, kinship systems,

music, and law. As we have seen, classical categories deriving from formal systems require a neural architecture that is capable of ignoring the statistical microstructure of the properties of the exemplars of a category that an individual has encountered. One can speculate that the development of a non-associative neural architecture suitable to formal systems was a critical event in the evolution of human intelligence.

Family Resemblance Categories: Inferences within Historically-Related Similarity Clusters

In a previous section we showed that learners of English are presented with a family resemblance structure and must cope with it if they are to speak the same language as their parents. Are there cases where learners of conceptual categories are similarly forced to cope with a family resemblance structure in nature if they are to be able to make inferences about it? Many people have noted similarities between linguistic and biological evolution (see, e.g. Cavalli-Svorza & Feldman, 1981), and there is a particularly compelling analogy in the formation of family resemblance categories in the evolution of biological taxa.

It is generally believed that a novel species evolves from a small interbreeding population occupying a local, hence relatively homogeneous stable environment. Through natural selection, the organisms become adapted to the local environment, with the adaptive traits spreading through the population via sexual reproduction. As a result the population assumes a morphology that is relatively uniform — since selection acts to reduce variation (Sober, 1984; Ridley, 1986) — and predictable in part from engineering considerations to the extent that the organism's niche and selection pressures can be identified (Hutchinson, 1959; Williams, 1966; Dawkins, 1986).

Subsequent geographic dispersal can cause the members of the ancestral population to form reproductively isolated subgroups. They are no longer homogenized by interbreeding, and no longer subject to the same set of selection pressures imposed by a local environment. In the first generation following dispersal, the species is still homogeneous. Then, a set of distinct processes destroys the homogeneity of class: genetic drift, local geographic and climatic changes imposing new selection pressures, adaptive radiations following entry into empty environments, and local extinctions. As a result, the descendants of the ancestral species form a family resemblance category — the category of "birds," for example. Robins, penguins, and ostriches share many features (e.g.

feathers) because of their common ancestry from a single population adapted to flying, while differing because of independent processes applying to different members of that population through history.

This suggests that as in the case of irregular past tense subclasses, the family resemblance structure of many biological taxa comes from the world, not just the minds of those learning about them. Note that such family resemblance structures are not always identical with classically-defined categories, and may be indispensable even in the best scientific theories. Many traditional biological taxa are somewhat arbitrary, serving as useful summaries of similar kinds of organisms. There are, to be sure, some biological categories that are well defined, including species (a population of interbreeding organisms sharing a common genepool), and monophyletic groups or clades (all the descendants of a common ancestor also belonging to the category). But many important biological taxa are neither. For example, fish comprise thousands of species, including coelocanths and trout. But the most recent common ancestor of coelocanths and trout is also an ancestor of mammals. Therefore no branch of the genealogical tree of organisms corresponds to all and only fish; trout and coelocanths are grouped together and distinguished from mammals by virtue of their many shared properties. To some biologists this is reason to deny the scientific significance of the category altogether, but most probably agree with the sentiment captured by Gould when he writes: "A coelocanth looks like a fish, tastes like a fish, acts like a fish, and therefore — in some legitimate sense beyond hidebound tradition — is a a fish" (Gould, 1983; p. 363). In other words, biologists often recognize a category that is characterized as a cluster of co-occurring properties. Indeed some taxonomists have tried to characterize taxa using clustering algorithms that use criteria similar to those thought to lead to the formation of prototype conceptual categories in humans (see Ridley, 1986; Bobick, 1987).

Thus we have seen two examples of family resemblance categories that exist in the world, and that have the same genesis: a law-governed process creating a relatively homogeneous class, followed by a cessation of the influence of the process and the operation of independent historical causes that heterogenize the class, though not to such an extent that the inter-member similarities are obliterated entirely. Since objects can escape the direct influence of laws while retaining some of their effects, a smart organism cannot count on always being able to capture the world's regularities in formal systems. For example, no observer knowing only the United States Constitution would be able to explain why presidents are always wealthy white Christian males. Similarly,

presumably no observer, not even a scientist equipped with a knowledge of physiology and ecology, would be able to explain why penguins have feathers, like robins, rather than fur, like seals. Instead, it will often be best simply to record the interpredictive contingencies among objects' properties to infer unknown properties from known ones. Thus a smart observer can record the contingencies among feathers, wings, egg-laying, beaks, and so on, to note that the world contains a set of objects in which these properties cluster, and to use the presence of one subset of properties to infer the likely presence of others.

Just as irregular subclasses were shaped both by divergent and convergent historical processes, in the domain of conceptual categories there is a convergent process that can cause objects to cluster around natural modes even if the objects are not linked as descendants of a more homogeneous ancestral population. For example, there is no genealogical account of chairs that parallels the ones we give for languages or species. The similarities among chairs are caused solely by a convergent process, in which a set of properties repeatedly arises because it is particularly stable and adaptive in a given kind of environment, several historically unrelated groups of organisms evolve to attain that set. Examples include nonhomologous organs such as the eyes of mammals and of cephalopods, the wings of bats and of birds, and polyphyletic groups such as cactuslike plants (which have evolved succulent leaves, spines, and corrugated stems as adaptations to desert climates in several parts of the world). As in the case of divergent evolution discussed above, there is a mixture of shared and distinct properties that are respectively caused by law-governed adaptation and historical accident, though here the influences are temporally reversed. For example, although vertebrate and cephalopod eyes are strikingly similar, in vertebrates the photoreceptors point away from the light source and incoming light has to pass through the optic nerve fibers, whereas in cephalopods the photoreceptors point toward the light in a more "sensible" arrangement. The difference is thought to have arisen from the different evolutionary starting points defined by the ancestors to the two groups, presumably relating to differences in the embryological processes that lay down optic and neural tissue. Artifacts such as chairs develop via a similar process; for a chair to be useful, it must have a shape and material that is suited to the function of being stable and accessible (Winston, Binford, Katz, & Lowry, 1983), but it is also influenced by myriad historical factors such as style, available materials, and ease of manufacture with contemporary technology. Social stereotypes, arising from the many historical accidents that cause certain kinds of people to assume certain roles, are another example.

We might expect family resemblance categories to be formed whenever there is a correlational structure in the properties that people attend to among sets of objects they care about, and that the world will contain opportunities for such clusters to form wherever there are laws that cause properties to be visibly correlated and historical contingencies that cause the correlations to be less than perfect — which is to say, almost everywhere.

Schematic Summary

These ideas may be summed up by sketching four schematic worlds. Each contains objects that have three attributes of interest to an organism, and each can assume a small set of values, so an object can be represented as, say, ABC. The organism needs to know as many of the values as possible but can observe only subsets of them.

A random world. In the world schematized in (6), there is nothing to gain by forming categories; every attribute occurs equally often with every other attribute, and the maximum predictive power is achieved by consulting base rate probabilities across all objects.

(6)
ABC
ABF
AEC
AEF
DBC
DBF
DEC
DEF

An ideal law-governed world. In the second world, classical categories are useful. In world (7), an observer can categorize any object as being in one of two classes, "X" and "Y," on the basis of, say, the first attribute, and can predict the values of the other two attributes with certainty.

(7)
ABC
ABC
ABC
ABC

DEF
DEF
DEF
DEF

Such a world can come about if there are laws operating, in particular, X → A B C and Y → D E F, and an observer who internalized the laws can forget the actual groups of objects and consult the laws themselves.

A world evolved by divergent processes. In the third world, prototype categories are useful, (8). Knowing that an object has the attribute A one can assign in to the first category and then predict with .67 confidence that it will have a B (as opposed to the .375 base rate for B's in general).

(8)
ABC
ABR
PBC
AQC
DEF
DER
PEF
DQR

Such a world can come about if the classical world in (7) was subject to divergent evolutionary forces in which some of the attributes of each object within a category were subject to unsystematic replacement with other values through history:

(9)
ABC
AB(C → I)
(A → P)BC
A(B → Q)C
DEF
DE(F → R)
(D → P)EF
D(E → Q)F

A world evolved by convergent processes. Finally, another world in which prototype categories would be useful is 10.

(10)

ABC
ABC
ABC
ABF
ABC
DBC
AEC
DBF
AEF
DEF
DEC
DEF
DEF
DEF

It could have evolved out of the random world of (6) by a convergent evolutionary process if certain pairs of values were more stable than others and served as attractor states. For example if the selective pressures in (11) were at work, world (10) could have developed through the historical sequence shown in (12).

(11)

AB > AE
BC > BF
AC > DC
DE > DB
EF > EC
DF > AF

(12)

ABC → ABC
ABF → ABC, DBF
AEC → ABC, AEF
AEF → ABF, DEF
DBC → ABC, DEC
DBF → DBC, DEF
DEC → AEC, DEF
DEF → DEF

Interactions Between Classical and Family Resemblance Categories

The referents of many words, such as *bird* and *grandmother*, appear to have properties of both classical and family resemblance categories. How are these two systems to be reconciled? The distinction between cores used for reasoning and stereotypes used for identification was of no help in the case of English past tense forms, and the distinction does not do much better when applied to conceptual categories. Many classical categories have no family-resemblance identification procedure associated with them, for example, the number "-3". Many family resemblance categories have no classical category serving as a core that they identify, such as "seafood" or Wittgenstein's famous example, "game". Furthermore some classical categories can be identified by simple, easily-computable, all-or-none tests. For example, odd numbers can be quickly identified by tests such as "divide by 2 and check for remainder" or "see if last digit is 1, 3, 5, 7, or 9"; in fact, the features of the associated family resemblance class, such as "has many odd digits" (which Armstron, et al., 1983, found to be feature that led subjects to judge that a given number was a better example of the "odd" class), are not even probabilistically diagnostic. On the other side, family resemblance classes can support nonperceptual reasoning, sometimes quite reliably, such as "presidents are well-off," "vegetables are not served for desert," or "tools have metal in them". We are not denying that categories may have "cores" in the sense that some kinds of knowledge are given priority in over others when they conflict, but it does not seem that this distinction can be equated either with quick-identification versus reasoning or with classical versus family resemblance categories (Armstrong, et al., 1983, and Rey, 1983, mention some of these problems.)

A more likely reconciliation is that people have parallel mental systems, one that records the correlational structure among sets of similar objects, and another that sets up systems of idealized laws. Often a category within one system will be linked to a counterpart within the other. In general we might expect family resemblance categories to be more accessible to observers than classical categories. Most objects in the world are cluttered by the effects of the myriad historical processes that led to their creation and preservation, obscuring underlying laws. In the lucky cases when people are able to see these laws peeking through the clutter and try to capture them in idealized systems, the elements of these systems may be seen to apply to many of the objects belonging to the family resemblance clusters that were independently formed though simple observation of the correlational structure displayed by frequently-encountered exemplars. In such cases, languages appear to assign the same verbal label to both. This is what leads to the ambiguity of *A penguin is a perfectly good bird*, one of whose readings is true, the other false. It also is what leads to such paradoxes as Armstrong, et al's subjects who could assert both that odd numbers form an all-or-none category tolerating no intermediate degrees of membership, and that 13 is a better example of it than 23.

The fact that these systems are distinct is at the heart of Putnam (1975) and Kripke (1972)'s well-known argument that natural kind terms are not defined by a set of conditions that pick out the members of the category in the world. Thus even though we think of "animal" as a necessary part of the definition of *cat*, if we were to discover that cats were in fact robots controlled from Mars, we would not conclude that *cat* no longer referred to the entities formerly called cats, or that it did refer to catlike entities on some other planet that really were animals. Rather, the label *cat* is rigidly assigned to a set of objects in the world. According to Putnam, people have a "stereotype" of such objects that helps them at tentative identification, but will defer to an expert in establishing category membership more definitively.

Schwartz (1979) points out that such intuitions about natural kind terms are driven by a belief that their members have an "underlying trait" in common. People act as if they believe in the existence of such a trait even if they are prepared to accept that their current belief of the nature of the trait is incorrect, even if they have no idea of the nature of the trait, indeed even if *no one* knows the nature of the trait. Schwartz's analysis suggests that people's intuitions are influenced by a metatheory, a kind of essentialism, that asserts that the varying forms that an object can assume are causally related in terms of their relation to a hidden trait or essence. These essences are clearly not family resemblance categories since it is possible, indeed typical, that they are not associated with *any* properties of the relevant objects at all, let alone a cluster of frequently co-occurring properties (that is, if Putnam and Kripke are correct, there is no property associated with the concept underlying "cat" or "gold" that cannot be relinquished by a person while he or she still believes the concept to apply to that category of objects.) Rather, the hidden essences must be represented as abstract symbols within internally represented formal systems, defining slots for particular traits provided by folk or formal science, and allowing inferences to be made about heredity, growth, physical structure, change, and behavior. Subsequent research by Keil (1989) and Gelman (Gelman, et al., 1994) has gathered evidence for essentialist thinking in preschool children and adults in nonliterate cultures. More generally, Medin (1989), Murphy (1993), Rips (1989), and Smith, Nisbett, and Langston (1992) have emphasized the importance of intuitive rule-like theories in the organization of people's conceptual categories.

In sum, natural kind terms like *cat* or *gold* are linked both to stereotypes, or family resemblance categories acquired by observing the correlational structure in sets of similar familiar objects, and abstract essences or hidden traits within an intuitive theory which unite an objects' varying appearances and provide the infrastructure for bits of folk science and institutionalized science. (The Putnam-Kripke puzzles arise from thought-experiments in which these systems are separated.)

The human tendency to induce categories from clusters of similar objects they have encountered, to construct formal systems of rules applying to ideal objects, and to link entities of the two kinds with each other is probably the root of many apparent paradoxes in the study of concepts and often within the conceptual systems themselves. For example, exactly this duality can be found in the legal system in the distinction between reasoning by constitutionality and by precedent. Legal questions are commonly resolved by appealing to precedents, with more similar prior decisions carrying more weight. However when the constitutionality of a current decision is at issue, only a restricted set of principles is relevant, and similarity to earlier cases must be ignored.

First-Order and Higher-Order Systems

We have been highlighting the distinction between formal systems and correlational structures, but that does not mean one will commonly see either system in isolation. We have already noted that the two kinds of systems can cross-reference the same objects. Here we suggest that more complex kinds of interaction are also possible, and that the nature of these interactions can explain why the distinction may not appear clearcut in many cases.

Vanishing classical categories. Sometimes close scrutiny reveals a family resemblance structure even for categories that would seem to be classical through and through. Lakoff (1987) points out that even the set of "mothers" has a family resemblance structure. He is not even referring to stereotypes about children, church, and kitchen, but to the literal characterization of the concept "mother" itself. There are, he points out, surrogate mothers, adoptive mothers, eggdonor mothers, and foster mothers. Similarly, the notion of a "species" has unclear cases. There are populations dispersed over a wide area in which animal A might be able to mate with neighbor B, B can mate with neighbor C, but A cannot mate with C. Lakoff suggests that classical categories never exist in the world; they are artifacts of an outmoded Aristotelian mode of thinking.

Our account suggests a different analysis. Classical categories are implicitly defined by formal systems, but formal systems work only as *idealizations*. Real objects are never idealizations, by definition. Each is a nexus at which many influences converge. Therefore objects as they are found in the world should not directly fall into classical categories. Only as idealized within a single formal system do they function as members of classical categories. Often, *several* formal systems pick out the same class of individuals. Let us call them "first-order" systems. For example, the concept "mother" is used within several formal systems. In genetics, it corresponds to the contributor of half of an offspring's genes including those on the X chromosome in mammals. In evolutionary theory it corresponds to the producer of the larger of the two kind of gametes in sexual reproduction. In the theory of reproductive physiology it is the site of prenatal growth and birth. In the ethology of mammals it is the adult that provides the greater minimal parental investment. In law it might be the spouse of the father, or the female guardian. In genealogy it is the immediate female ancestor.

This multiple ambiguity is usually invisible because the different notions of mother are generally the same individuals. These coinciding roles are put into register by other systems, which we call "second-order" systems. For example, theories within sociobiology predict that under certain general conditions the birth mother will also be the nurturance mother because of the effects of investing resources in individuals likely to share one's genes. Physiological considerations explain why it is the contributor of the larger of the two gametes that is likely to be the site of prenatal growth. The social anthropology of a society might dictate that the spouse of the father is the one who nurtures the offspring after birth.

These theories about theories, or second-order systems, systematically link entities within subordinate, first-order formal systems. And under the conditions in which the *second*-order systems apply, the category "mother" is classical in all of its usages. However, when the idealized conditions of the second-order system are *not* satisfied because of historical or other accidental contingencies, the different notions of "mother" are severed, and the entire superordinate set of mothers defines a family resemblance category. For example, with changes in reproductive technology, the contributor of half an offspring's genes may no longer be the birth mother. With changes in social systems allowing adoption, the birth mother may no longer be the mother legally responsible for nurturing and protecting the child. A family resemblance category is thereby formed containing members such as stepmother, surrogate mother, donor mother, and adoptive mother, for the same reason that family resemblance categories can arise in general: a law in formal system (in this case, the second-order system) no longer applies at a point in history; unrelated sets of local processes partially destroy category homogeneity. The difference between mothers and other family resemblance categories is that we have a formal system (the second-order system) applying and then ceasing to apply over roles within other (first-order) formal systems, rather that applying and then ceasing to apply over simple properties. Thus even within the full, family-resemblance set of mothers in modem western society, there exists a classical category of "genetic mother" defined within the first-order formal system of genetics, a classical category "legal mother" within the legal system, a classical category "birth mother" within embryology, and so on.

A similar account can be provided for "species," which biologists recognize to be a multifaceted concept (Ridley, 1986; Maynard Smith, 1986; Dawkins, 1985; Mayr, 1982; Gould, 1983). The notion of "species" has a at least the following interpretations: phenetic sense as a cluster of morphological traits, an ecological sense as an occupier of a niche, a genealogical sense as the descendents of a common ancestor, a reproductive-physiological sense as a set of organisms that can produce fertile offspring, and a genetic sense as a common set of genetic structures. The different notions within these first-order theories are linked by various second-order theories. The theory of natural selection explains why a "phenetic species" is also a "genealogical species". Chromosomal genetics explains why populations with a common genetic structure are those that can produce fertile offspring. Population genetics explains why a cluster of morphological traits becomes uniform within a population of interbreeders. Physics and physiology explain why certain morphological traits are associated with certain ecological niches. However, some of these second-order theories apply only under the idealization of an interbreeding population in a stable local environment; when the idealization no longer can apply because of geographic barriers, dispersal, asexual reproduction, artificial selection, and so on, the second-order linkages fail. This gives rise to family resemblance categories such as the members of a "ring species" or a domesticated species, and other cases where the standard notion of species does not easily apply. However there still may first-order theories in which some notion of a species is a useful category to make generalizations over, for example, when a population of widely-spread organisms retains a unique derived morphological trait even if not all members can interbreed.

Complexities in the Past Tense System. The ability of classical and family resemblance categories to arise as second-order phenomena brings us full circle to the past tense system and can explain the *dis*analogies between the regular/irregular distinction and the classical/prototype distinction. Skeptics could point out a number of counterexamples to our general claim: examples in which the regular system appears to have prototype effects. But in these cases the counterexamples can be shown to come from second-order systems rather than the first-order system of morphology that we have been describing.

First, the regular system is not a unified class but has three variants: *t, d, ed* (e.g. *walked, jogged, patted*). However, this is because the output of the regular rule, a stem affixed with *d*, is fed into a separate system, phonology, which adjusts the phonetic shape of the past tense form in ways that are general across the sound pattern of English, and independent of the past tense form itself (in any case these phonological adjustments are themselves perfectly rule-governed and hence define three classical categories; see Pinker & Prince, 1988). Second, as we noted earlier, some regular forms (e.g. *sneaked, dreamed* are indeed graded in acceptability. However, this is because regularly affixed forms are fed into the (second-order) morphological paradigm system, in which regular and irregular forms must compete for a single paradigm slot and the gradedness of the irregular root can give rise to graded intuitions concerning the final output form of the product of the regular rule.

A disanalogy in the opposite direction comes from the irregular class. In fact, they are not just arbitrary statistical associations among phonological properties of the stem and past tense form, as a standard family resemblance category is and as the Rumelhart-McClelland model treats them. Rather, the associations are between a restricted set of properties of the stem (principally the number of syllables, rhyme, and alliteration), not just any phonological properties, and a constrained set of morphological processes, such as copying, changes of vowel quality, and affixation, not just any mapping between stem forms and past forms. Thus it is at least possible that the family resemblance structure of the irregulars may itself be a second-order associative structure defined over first-order, possibly classical mini-rules, not a first-order system defined over primitive phonological features. This has implications for the use of pattern associators to account for the representation of family resemblance categories. Even under our charitable suggestion that they may be useful for this task, the entities that are associated with one another are not always primitive features associated with simple binary units, but might themselves have to be complex rule-like structures or pointers to them.

Conclusion

It may be surprising to see so many parallels drawn between two phenomena that seem to be in such different domains. We are not claiming that past tense forms and conceptual categories are alike in all essential respects or that they are generated by a single cognitive system. But often widespread similarities in remote domains makes the case for *some* common underlying principles compelling. English past tense forms come in two versions that are identical in function and at first glance only differ in size and degree of uniformity. On closer examination they turn out to represent two distinct systems that correspond point for point with classical and family resemblance categories, respectively. Moreover the two systems are linked with distinct psychological faculties, developmental courses, real world causes, and computational architectures. A fundamental distinction must lie at the heart of this duality. Specifically, we suggest, human concepts can correspond to classical categories or to family resemblance categories. Classical categories are defined by formal rules and allow us to make inferences within idealized law-governed systems. Family resemblance categories are defined by correlations among features in sets of similar memorized exemplars, and allow us to make inferences about the observable products of history.

Notes

* The order of authors is arbitrary. We thank Ned Block, Paul Bloom, Ray Jackendoff, and Ed Smith for comments. This paper was prepared while the first author was a visitor at the MRC Cognitive Development Unit, University of London, and was supported by NIH grant HD 13831.

1. Note that the term means only "a pattern of resemblance such as one sees in a family"; it does not imply literal genealogical links.

2. There are also problems with the model's treatment of these phenomena; see Pinker and Prince (1988); Lachter and Bever (1988), and Sproat (1992).

References

ANDERSON, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

ARMSTRONG, S.L., GLEITMAN, L.R. & GLEITMAN, H. (1983) What some concepts might not be. *Cognition, 13,* 263-308.

ARONOFF, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

BEVER, T. & Langendoen, T. (1963) (a) The formal justification and linguistic role of variables in phonology. (b) The description of the Indo-European E/O ablaut. (c) The E/O ablaut in Old English. *RLE Quarterly Progress Report* (summer). Cambridge, MA: MIT Research Laboratory of Electronics.

BOBICK, A. (1987). *Natural object categorization.* Unpublished doctoral dissertation, Department of Brain and Cognitive Sciences, MIT.

BROWN, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

BROWN, R. (1977). Word from the language acquisition front. Invited address at the 48th Annual Meeting of the Eastern Psychological Association, Boston, April.

BRUNER, J.S., GOODNOW, J. & AUSTIN, G. (1956). *A study of thinking*. New York: Wiley.

BYBEE, J.L. (1985). *Morphology*. Philadelphia: Benjamins.

BYBEE, J.L. (1991). Natural morphology: The organization of paradigms and language acquisition. In: T. Huebner and C. Ferguson (Eds.) *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Amsterdam: Benjamins. 67-92.

BYBEE, J.L. & MODER, C.L. (1983). Morphological classes as natural categories. *Language, 59,* 251-270.

BYBEE, J.L. & SLOBIN, D.I. (1982a). Rules and schemes in the development and use of the English past tense. *Language, 58,* 265-289.

BYBEE, J.L. & SLOBIN, D.I. (1982b). Why small children cannot change language on their own: Suggestions from the English past tense. In A. Ahlqvist (Ed.), *Papers from the 5th Internation Conference on Historical Linguistics. (Current Issues in Linguistic Theory Vol. 21, Amsterdam Studies in the theory and history of linguistic science IV.)* Philadelphia/ Amsterdam: John Benjamins.

CAMPBELL, A. (1959). *Old English Grammar.* Oxford: Oxford University Press.

CAVALLI-SFORZA, L.L. Feldman. M.W. (1981). *Cultural transmission and evolution: A quantitative approach.* Princeton, NJ: Princeton University Press.

CURME, G. (1935). *A Grammar of the English Language II.* Boston: Barnes & Noble.

DAWKINS, R. (1986) *The blind watchmaker.* New York: Norton.

ERVIN, S. (1964). Imitation and structural change in children's language. In E. Lenneberg (Ed.), *New directions in the study of language.* Cambridge, MA: MIT Press.

FODOR, J.A. (1981). The present status of the innateness controversy. In J.A. Fodor, *Representations*. Cambridge, Mass.: MIT Press.

FODOR, J.A., GARRETT, M.F., WALKER, E.C.T., and PARKES, C.H. (1980). Against definitions. *Cognition* 8: 263-267.

FREYD, J.J. (1983). Shareability: The social psychology of epistemology. *Cognitive Science, 7,* 191-210.

GELMAN, S.A. & MARKMAN, E. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532-1540.

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. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture.* New York: Cambridge University Press.

GORDON, P. (1985). Level-ordering in lexical development. *Cognition, 21,* 73-93.

GOULD, S.J. (1983). What, if anything, is a zebra? In S.J. Gould, *Hen's teeth and horses toes.* New York: Norton.

HOOPER, J.B. (1976). *Introduction to natural generative phonology.* New York: Academic Press.

HULL, C.L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs, 28,* Whole No. 213.

HUNT, E. (1962). *Concept learning: An information processing problem.* New York: Wiley.

HUTCHINSON, G.E. (1959). Homage to Santa Rosalia, or whey are there so many kinds of animals. American Naturalist, 93, 145-159.

JESPERSEN, O. (1942). *A modern English grammar on historical principles, VI.* Reprinted 1961: London: George Allen & Unwin Ltd.

JOHNSON, K. (1986). Fragmentation of strong verb ablaut in Old English. *Ohio State University Working Papers in Linguistics 34:* 108-122.

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

KELLY, M.H. (1992). Darwin and psychological theories of classification. *Evolution and Cognition, 2,* 79-97.

KIM, J.J., MARCUS, G.F., PINKER, S., HOLLANDER, M., & COPPOLA, M. (1994) Sensitivity of children's inflection to morphological structure. *Journal of Child Language, 21,* 173-209.

KIM, J.J., PINKER, S., PRINCE, A. & PRASADA, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science, 15,* 173-218.

KIPARSKY, P. (1982a). From cyclical to lexical phonology. In H. van der Hulst, & N. Smith (Eds.), *The structure of phonological representations.* Dordrecht, Netherlands: Foris.

KIPARSKY, P. (1982b). Lexical phonology and morphology. In I.S. Yang (Ed.), *Linguistics in the morning calm.* Seoul: Hansin, pp. 3-91.

KONNER, M. (1982). *The tangled wing.* New York: Harper and Row.

KRIPKE, S. (1972). Naming and necessity. In D. Davidson and G. Harman (Eds.), *Semantics of natural language.* Dordrecht: Reidel.

KUCZAJ, S.A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16,* 589-600.

KUCZAJ, S.A. (1978). Children's judgments of grammatical and ungrammatical irregular past tense verbs. *Child Development, 49,* 319-326.

KUCZAJ, S.A. (1981). More on children's initial failure to relate specific acquisitions. *Journal of Child Language, 8,* 485-487.

LACHTER, J. & BEVER, T.G. (1988). The relation between linguistic structure and associative theories of language learning — A constructive critique of some connectionist learning models. *Cognition 28:* 195.

LAKOFF, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

MACWHINNEY, B. & LEINBACH, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40,* 121-157.

MARCUS, G.F. (1995). The acquisition of inflection in children and multilayered connectionist networks. *Cognition, 56,* 271-279.

MARCUS, G.F., BRINKMANN, U., CLAHSEN, H., WIESE, R. & PINKER, S. (1995). German inflection: The exception that proves the rule. *Cognitive*

Psychology, 29, 189-256.

MARCUS, G., PINKER, S., ULLMAN, M., HOLLANDER, M., ROSEN, T.J. & XU, F. (1992). *Overregularization in language acquisition.* Monographs of the Society for Research in Child Development, 57 (4, Serial No. 228).

MAYNARD SMITH, J. (1986). *The problems of biology.* Oxford: Oxford University Press.

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

MCCLELLAND, J.L. & RUMELHART, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159-188.

MCCLELLAND, J.L., RUMELHART, D.E. & The PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: Bradford Books/MIT Press.

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

MENCKEN, H. (1936). *The American Language.* New York: Knopf.

MURPHY, G.L. (1993). A rational theory of concepts. In G.H. Bower (Ed.) *The psychology of learning and motivation. Vol. 29.* New York: Academic Press.

OSHERSON, D.N. & SMITH, E.E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition 15,* 35-58.

PINKER, S. (1984). *Language learnability and language development.* Cambridge, Mass.: Harvard University Press.

PINKER, S. Rules of language. *Science, 253,* 530-535.

PINKER, S. & MEHLER, J. (Eds.) (1988). *Connections and symbols.* Cambridge, MA: MIT Press/Bradford Books.

PINKER, S. & PRINCE, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S.D. Lima, R.L., Corrigan, & G.K. Iverson (Eds.), *The reality of linguistic rules.* Philadelphia: John Benjamins.

PINKER, S. and PRINCE, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* 28: 73-193.

PLUNKETT, K. & MARCHMAN, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38,* 43- 102.

PLUNKETT, K. & MARCHMAN, V. (1993). From rote learning to system building. *Cognition, 48,* 21-69.

PRASADA, S. & PINKER, S. (1993). Generalizations of regular and irregular morphology. *Language and Cognitive Processes, 8,* 1-56.

PROKOSCH, E. (1939). *A comparative Germanic grammar.* Philadelphia: Linguistic Society of America.

PUTNAM, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, mind, and knowledge.* Minneapolis: University of Minnesota Press.

QUINE, W.V.O. (1969). Natural kinds. In W.V.O. Quine, *Natural kinds and other essays.* New York: Columbia University Press.

REY, G. (1983). Concepts and stereotypes. *Cognition 15,* 237-262.

RIDLEY, M. (1986). *The problems of evolution.* Oxford: Oxford University Press.

RIPS, L.J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* New York: Cambridge University Press.

ROSCH, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

ROSCH, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

ROSCH, E. (1988). Coherences and categorization: A historical view. In F. Kessel (Ed.), *The development of language and of language researchers: Papers presented to Roger Brown.* Hillsdale, NJ: Erlbaum.

ROSCH, E. & MERVIS, C.B. (1975). Family resemblances: Studies in the internal representation of categories. *Cognitive Psychology, 7,* 573-605.

RUMELHART, D.E. & MCCLELLAND, J.L. (1986b). On learning the past tenses of English verbs. In J.L. McClelland, D.E. Rumelhart & The PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: Bradford Books/MIT Press.

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

SHEPARD, R.N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323.

SLOBIN, D.I. (1971). On the learning of morphological rules: A reply to Palermo and Eberhart. In D.I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium.* New York: Academic Press.

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

SMITH, E.E., LANGSTON, C. & NISBETT, R. (1992). The case for rules in reasoning. *Cognitive Science, 16,* 1-40.

SMITH, E.E., MEDIN, D.L. and RIPS, L.J. (1984). A psychological approach to concepts: Comments on Rey's "Concepts and Stereotypes." *Cognition 17,* 265-274.

SOBER, E. (1984). *The nature of selection.* Cambridge, MA: MIT Press.

SPROAT, R. (1992). *Morphology and computation.* Cambridge, MA: MIT Press.

STROMSWOLD, K.J. (1990). Learnability and the acquisition of auxiliaries. Doctoral Dissertation, Department of Brain and Cognitive Sciences, MIT.

ULLMAN, M. (1993). The computation and neural localization of inflectional morphology. Unpublished doctoral dissertation, Dept. of Brain & Cognitive Sciences, MIT.

WHITTLESEA, B.W.A. (in press). Selective attention, variable processing, and distributed representation: Preserving particular experiences of general structures. In R.G. Morris (Ed.), *Parallel Distributed Processing: Implications for psychology and neuroscience.* New York: Oxford University Press.

WILLIAMS, G.C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought.* Princeton: Princeton University Press.

WINSTON, P.H., Binford, T.O., Katz, B. & Lowry, M. (1983). Learning physical descriptions from functional definitions, examples, and precedents. MIT Artificial Intelligence Laboratory Memo 679.

WITTGENSTEIN, L. (1953). *Philosophical investigations.* New York: Macmillan.

XU, F. & PINKER, S. (1995). Weird past tense forms. *Journal of Child Language, 22,* 531-556.