Chapter 10
The Nature of Human Concepts: Evidence from an Unusual Source

10.1 Classical and Family Resemblance Categories

This chapter 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 is a worse example.

John Macnamara was a man of extraordinary wit, wisdom, and warmth. His insights on the relations among cognition, logic, and the world greatly influenced the thinking that went into this chapter. It is an honor to contribute to a volume in his memory.

The order of authors is arbitrary. We thank Ned Block, Paul Bloom, Ray Jackendoff, and Ed Smith for comments. Preparation of this chapter was supported by National Institutes of Health grant HD 18381.

A longer version of this chapter may be found in Communication and Cognition, 29, 307–361.
• 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 that may or may not be members of the category at all. One example is the fossil genus Archaeopteryx, characterized by one paleontologist as a "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 saw in the famous controversy that followed the Reagan administration's proposal that ketchup be classified as a vegetable in meeting nutritional guidelines for school lunch menus.

• They often display a family resemblance structure (Wittgenstein 1953). 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.

10.1.1 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 Smith and Medin 1981; Rosch 1973, 1978, 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 et al. 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 analyzable 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 many psychological tasks. For example, people can verify that prototypical members belong to a category faster and more accurately than they do with 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 the names for other exemplars, and that they 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.

10.1.2 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, Gleitman, and Gleitman 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, Gleitman, and Gleitman discovered, in an independent questionnaire, that their subjects believed as much), 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 corrigeable 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, and that raccoons with stripes painted down their backs are raccoons, not skunks (Rey 1983; Armstrong, Gleitman, and Gleitman 1983; Keil 1989; Gelman, Coley, and Gottfried 1994).

Similar demonstrations with adults have shown that inference is not driven by the similarity criteria that define family resemblance categories (see Murphy 1993; Medin 1989; Kelly 1992; Smith, Langston, and 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 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 the butterfly are “the same animal,” but that the caterpillar and the centipede are not, despite appearances to the contrary.

10.1.3 Possible Resolutions
This conflicting evidence can be resolved in several ways.

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-category view can be found in Rey 1983, Fodor 1981, and Armstrong, Gleitman, and Gleitman 1983; and tentative proposals favoring the core-plus-identification-procedure compromise can be found in Smith and Medin 1981, Armstrong, Gleitman, and Gleitman, 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 vs. 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? 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.

10.2 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/gone, sleep/slept, make/made, ring/rang, bring/brought, sink/sank, 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 (see Pinker and Prince 1988 for a full list). Here are some examples:

- Lax the vowel: bleed, breed, feed, lead, mislead, read, speed, plead, meet, hide, slide, bite, light, shoot
- Lax the vowel, add a -t: lose, deal, feel, kneel, mean, dream, creep, keep, leap, sleep, sweep, weep, leave
- Change the rhyme to -ought: buy, bring, catch, fight, seek, teach, think
• Change /æ/ to /æ:/ or /ʌ/: ring, sing, spring, drink, shrink, sink, stink, swim, begin, cling, fling, sling, sting, string, swing, wring, stick, dig, win, spin, stink, sink, run, hung, strike, sneak.
• Change the vowel to /u/: blow, grow, know, throw, draw, withdraw, fly, slay.

Let us consider some properties of the irregular subclasses.

10.2.1 Properties of the Irregular Subclasses

10.2.1.1 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 on or similar vowel to a 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 verb roots 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, 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 /ey/ to /i:/—take, mistake, foriake, shake—could include any word with an /ey/, 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.

10.2.1.2 Family Resemblance Irregular subclasses display a family resemblance structure (Bybee and Slobin 1982a; Bybee and Modér 1983). Consider the subclass that changes an /æ/ to an /ʌ/. Most of the verbs end with a velar nasal consonant: shrink, sink, stink, cling, fling, sling, sting, string, swing, sink. 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 a <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.

10.2.1.3 Prototypicality Bybee and Modér (1983) point out that for many of the subclasses, one can characterize a prototype, based on the kinds of characteristic phonological properties that define the family resemblance structure. According to Bybee and Modér, the prototype of the ing → ung subclass is

\[
\begin{array}{ll}
\text{S} & \text{C C i} \\
\text{velar} & \text{nasal}
\end{array}
\]

where C stands for a consonant. This prototype is maximally similar to most members of the existing subclass, but more interestingly, it predicts subjects’ generalization of the /æ/ → /ʌ/ change to novel verbs. Bybee and Modér 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 skrim as the past of strike 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, plum, 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 Marcus et al. (1995).

10.2.1.4 Graded Goodness of Membership Within most of the subclasses, there are some verbs that clearly accept the irregular past tense form, but there are others, usually of low but nonzero frequency, for which the specified past tense form is less than fully acceptable and which are felt to be unusual or stilted. In (1) we contrast some "good examples" of the past tense form with "poor examples" of the same kinds of forms; the intuitions vary from person to person, as is true for nonprototypical exemplars of conceptual categories. The judgments for these and similar forms are documented quantitatively by Ullman (1999).
(1) Good examples
hit, split
bled, led
burnt, bent
dealt, felt, meant
froze, spoke
got, forgot
wrote, drove, rode

Poor examples
spit, forbid
pled, sped
learnt, lent, rent
knelt, dreamt
wove, hove
begot, trod
dove, strove, smote, strode

10.2.1.5 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 verbs are restricted to idioms, clichés, or other specialized usages. For example, the expression forgo the pleasure of, as in You will excise 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 subjects by Ullman 1999). Similarly, That dress really becomes you is a natural English sentence, but When you were thinner, 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, and I don’t know how she bears it with I don’t know how she bore it.

10.2.1.6 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 a paradigm case of categorical, all-or-none operations and might be thought to correspond to 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 an 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.

10.2.2 Properties of the Regular Class
Before we accept the claims of Rumelhart and McClelland and of Lakoff, we must ask, Do all linguistic objects fall into family resemblance categories? Indeed, does the class most associated with English irregular verbs—namely, 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.

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 preempted or blocked. Thus, the fact that go has an irregular past went not only allows us to use went; 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 have seen 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 (cf., e.g., kept for keep). As a result wept does not sound terribly good, though it is not perceived as
being completely ungrammatical either (cf. *kept*). This effect has been
documented by Ullman (1999; see also Pinker 1991 and Pinker and Prince
1994), who asked subjects to rate the naturalness of irregular and regu-
larized past tense forms for verbs whose irregular pasts are somewhat
fuzzy in goodness. The two sets of ratings were negatively correlated.

We now put this reciprocity effect due to blocking and try to
determine whether the regular class has family-resemblance-category
properties independent of those of the irregular subclasses with which it
competes.

10.2.2.1 Independence of the Phonology of the Stem The first salient
property of the regular class is that it has no sensitivity to the phonologi-
cal properties of its stems. As a result, it has no phonologically charac-
terized 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
have different past tense forms: *ring*/*rang* versus *wring*/*wring*, *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

bleed/bled

bend/bent

sleep/slept

sell/sold

freeze/froze

grow/grew

take/took

stink/stunk

ring/rang

jut/jutted

need/needed

mend/mended

seep/seeped

yell/yelled

seize/seized

glow/glowed

fake/faked

blink/linked

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 comple-
mentary 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; the 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, 1994),
Kim et al. (1991), and Kim et al. (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 as 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 phonolo-
gical properties.

(3) He braked the car suddenly. ≠ broke

He flew 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-fleed 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. ≠ grandstanded

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 *Clinton has finally out-Nixoned
Nixon*. Like all verbs derived from nonverbs, it is regular. Since any lin-
guistically possible sound can be someone’s name, any linguistically pos-
sible sound can be a regular verb, allowing there to be regular homophone
for any irregular. For example:

(4) a. Mary out-Sally-Rided Sally Ride.

*Mary out-Sally-Rode Sally Ride.

b. In grim notoriety, Alcatraz out-Sing-Singed Sing-Sing.

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

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

This effect has been demonstrated experimentally in several popula-
tions. 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
give 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 also 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 reg-
ular words of differing similarity to existing English regular words. For
example, pilp is close to one of the prototypes for regular verbs in English,
because its rhymes with slips, flit, trip, nip, sip, sip, slip, 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 produced the prototypical and peripheral
forms with the same probability when they had to produce them.

10.2.2.2 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, freezingness, or restricted
syntactic contexts. In earlier work (Pinker and Prince 1988), we 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 fleeced, fleered, and anastomosed to be perfectly
good as the past tense forms of those verbs. These observations have been
confirmed experimentally by Ullman (1999); subjects' ratings of regular
pasts correlated 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 correlated less strongly with their stem
ratings but significantly with past frequency, partialling out stem rating.

Unlike what happens with 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 subjects 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.

These phenomena show why the apparent gradedness of acceptability
for regular forms like pleaded or weeped can be localized to the graded-
ness 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.

10.2.2.3 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),
and 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.

10.2.2.4 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 condi-
tions for membership can be stated simply: a verb, unless it has an
irregular root.

10.3 Psychological Implications

We have shown that by standard criteria the irregular subclasses are pro-
totype or family resemblance categories, and the regular class is a classical
category. This conclusion has several implications.
10.3.1 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 English-speaking child learns it and begins to use it productively in the third year of life (Marcus et al. 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 3 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 (fled 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 gross-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 joyride, first attaining popularity in the 1920s, was usually given the past tense form joyrided, as we would predict given its derivation from the noun a joyride. Prescriptive grammarians unsuccessfully tried to encourage joyrode in its place. Similarly, Kim et al. (1994) showed that children display the effect despite the fact that most have rarely or never heard regularized past tense forms for irregular-sounding verbs in the speech of adults.

10.3.2 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 causal 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” makes no sense. 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 clas-
chical all-or-none operation such as axing 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 -ingink, 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 laxing (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 (dog/did). 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.

10.3.3 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 can be determined from the meanings of the parts and the way in which they are combined.

Family resemblance categories, in contrast, are generalization 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 f<em>orgo</em> or the idiomatic meanings of stand or become.

A related account could help explain the genesis of irregular verbs' family resemblance structures. Rosch and Mervis (1975) found that people find lists of letter strings that display family resemblance structures easier to remember than ones 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 implications: classical and prototype categories are suited to different kinds of computational architectures, and the mental mechanisms giving rise to the two kinds of categories are suited to representing different kinds of entities in the world.

10.3.4 Computational Architecture

The acquisition of English past tense morphology has been implemented in a widely discussed computer simulation model by Rumelhart and McClelland (1986; see also Pinker and Prince 1988, 1994; Lachter and Bever 1988; Sproat 1992; Prasada and Pinker 1993; Marcus et al. 1992, 1995). The Rumelhart-McClelland (RM) model makes use of a pattern associator architecture, which is paradigmatic of the parallel distributed processing (PDP) or connectionist approach to cognitive science (Rumelhart and McClelland 1986; McClelland and Rumelhart 1986; Pinker and Mehler 1988). 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 and 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 on the basis of 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 and 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 and 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. After training on a set of regular and irregular verbs, the model 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*, *chung* for *clung*, and *wept* for *weep*. In addition, it showed a tendency to extend some of the subregular alternations to regular verbs on the basis of 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 and Prince 1988, Sproat 1992).

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. In earlier work (Pinker and Prince 1988), we 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 *brawn* for *brown*. It would often blend the /t/ and /d/ variants of the regular past tense form, producing *stepped* for *step* or *typted* for *type*. Sometimes the blends were quite odd, such as *membled* for *mailled* or *toureder* for *tour*.

Furthermore, we 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. Our 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 *smaetfspurice*, *tribfcrelth*, *smeefj* *leeflaog*, and *fifgfreezedl*.

Finally, the model is inconsistent with certain kinds of 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 *brokead*) 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 and 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 that 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 three-layer models, whose hidden layer’s weights are trained by error-back-propagation, perform much better (see, e.g., Plunkett and Marchman 1991, 1993; MacWhinney and Leinbach 1991). However, Sproat (1992), Prasada and 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

fail, including novel, unusual-sounding, and derived words (i.e., the fled out examples have exact analogues in German). The findings were verified in two experiments eliciting ratings of novel German words from German adults. The crosslinguistic comparison suggests that default suffixation arises not because numerous regular words reinforce a pattern in associative memory but from 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 in terms of psychological fidelity. We suggest that this is a symptom of the relative suitability of this architecture to handle family resemblance and classical categories in genera. 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 nonassociative architecture, used in language and presumably elsewhere.

10.3.5 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 representations 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 not necessarily correspond to the lawful groupings in the world?

The questions 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 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.

10.3.6 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 or her 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.

Earlier 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 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 and 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 children’s 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), they would be in a position to obliterate irregularity altogether if their 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 forms. 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 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/stringed, etc.). If some of these occasional forgettings and analogies get fixed in a language community in a contagion-like process (see Cavalli-Sforza and Feldman.
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 phenomenon 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-**drag** 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 will illustrate, the irregular 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 toward remembering items that are similar along multiple dimensions to other items. But in doing so it 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 generation of learners. The Old English strong classes, ancestors of most of today’s irregular verbs, evolved out of classes that can be traced back 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](#); Prokosch [1939](#); [Campbell 1959](#)). By the time of Old English, the patterns were 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**/>belief, **greet**/>gret, **heat**/>heat, **bite**/>bute, **slide**/>slode, **abide**/>labode, **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 strong past tense subclasses. 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 the strong classes, originally defined by phonological properties of their stems, to become 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 three examples:

**Phonological change** The verbs **blow**, **grow**, **throw**, **know**, **draw**, **fly**, **slay** all begin with a **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 that we can simplify as follows:

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>sang</td>
<td>sung</td>
</tr>
<tr>
<td>2nd</td>
<td>sung</td>
<td>sung</td>
</tr>
<tr>
<td>3rd</td>
<td>sang</td>
<td>sung</td>
</tr>
</tbody>
</table>

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 **sing**/>slung/>slung. The contrast between **freeze**/>froze and **cleave**/>cleft has a similar cause.
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.

Bobick (1987), Shepard (1987), and Anderson (1990) have attempted to reverse-engineer human conceptual categories 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; Quine 1969). Though we 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 object’s 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 trade-offs between the ease of categorization and the power of the licensed inference. For low-level, specific categories, one must 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 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, short-tailed, and long-hind-legged, and that he 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.

10.4 Implications for Conceptual Categories

Do the discoveries about classical and family resemblance categories in past tense forms offer insight into the role of classical and family resemblance categories in the domain of conceptual categories like birds and mothers? To answer this question, we must begin by considering what conceptual categories are for.

10.4.1 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 objects 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? Sometimes 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
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 co-occurring 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 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 thicker legs. If we know 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.

10.4.2 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 proposition by virtue of their form alone, so that any knowledge not explicitly stated in the proposition 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 the 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-heaters 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 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 Bruner, Goodnow, and Austin (1956), in which subjects learn categories like “red square with two borders,” is 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 three 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-or-none 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 though 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 “carries out decisions on entering wars,” 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 discussed, 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 nonassociative neural architecture suitable to formal systems was a critical event in the evolution of human intelligence.

10.4.3 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-Sforza and 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 (Huichinson 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 cases 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 they 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 gene pool) 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 coelacanths and trout. But the most recent common ancestor of coelacanths 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 coelacanth 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 coelacanth looks like a fish, tastes like a fish, acts like a fish, and therefore—in some legitimate sense beyond hidebound tradition—is a fish” (1983, 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 with the help of 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 noted 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 intermember 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 interpretative 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 by both divergent and convergent historical processes, so 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, and several historically unrelated entities evolve to attain that set. Examples from biology 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 holding people up, 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 correlation to be less than perfect—which is to say, almost everywhere.

10.4.4 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 Armstrong, Gleitman, and
Gleitman (1983) found to be a 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 dessert,” 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
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, Gleitman, and
Gleitman (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 pecking 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 through 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 the behavior of Armstrong, Gleitman, and Gleitman’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’s
(1975) and Kripke’s (1972) 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 neces-

...
cept 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, Coley, and Gottfried 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, Langston, and Nisbett (1992) have emphasized the importance of intuitive rule-like theories in the organization of people's conceptual categories.

In sum, natural kind terms like cet or gold are linked both to stereotypes, or family resemblance categories acquired by observing the correlational structure in sets of similar familiar objects, and to abstract essences, or hidden traits within an intuitive theory that unite an object's 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 one has 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 reasoning 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.

10.5 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 make 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 differ only 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.

Note

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

References


