COGNITIVE SCIENCE Vol 22 (2) 1998, pp. 189–228 ISSN 0364-0213
Copyright © 1998 Cognitive Science Society, Inc. All rights of reproduction in any form reserved.

Feature Centrality and Conceptual Coherence

STEVEN A. SLOMAN
Brown University

BRADLEY C. LOVE
Northwestern University

WOO-KYOUNG AHN
Yale University

Conceptual features differ in how mentally tranformable they are. A robin that does not eat is harder to imagine than a robin that does not chirp. We argue that features are immutable to the extent that they are central in a network of dependency relations. The immutability of a feature reflects how much the internal structure of a concept depends on that feature; i.e., how much the feature contributes to the concept's coherence. Complementarily, mutability reflects the aspects in which a concept is flexible. We show that features can be reliably ordered according to their mutability using tasks that require people to conceive of objects missing a feature, and that mutability (conceptual centrality) can be distinguished from category centrality and from diagnosticity and salience. We test a model of mutability based on asymmetric, unlabeled, pairwise dependency relations. With no free parameters, the model provides reasonable fits to data. Qualitative tests of the model show that mutability judgments are unaffected by the type of dependency relation and that dependency structure influences judgments of variability.

INTRODUCTION

The notion of a feature is central to the study of cognition. Models abound which assume that mental representations can be reduced to sets of features and that measures over those features can predict performance. Feature-based models have been applied to the analysis of similarity (Tversky, 1977), metaphor (Ortony, 1993), categorization (e.g., Estes, 1993;

Direct all correspondence to: Steven A. Sloman, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912; E-Mail: Steven_Sloman@brown.edu.

Medin & Schaeffer, 1978), memory (e.g., Murdock, 1993), learning (e.g., Spence, 1936), and induction across categories (Osherson, Smith, Shafir, Gualtierotti, & Biolsi, 1995; Sloman, 1993). A fundamental assumption made by all these models is that concepts can be meaningfully reduced to sets of features; i.e., to constituent aspects, parts, and attributes. These features are treated as independent in the sense that they make separable—usually additive—contributions to the model's output. However, these models need not consider features independent in another sense: The importance of a feature may be a function of its relations to other features. Features may differ in their centrality with respect to a concept and the only viable accounts of centrality appeal to features' roles in networks of knowledge.

A compelling illustration of this hypothesis was provided by Keil (1989) in the domain of categorization (see also Rips, 1989). Keil reported a series of experiments in which children and adults were asked to categorize stimuli having the perceptual features of one category but the internal features of another. Although all participants tended to categorize artifacts according to their perceptual features, older children and adults were more likely to categorize natural kinds according to their internal features. Keil argued that the internal features of natural kinds became weightier in category judgments to the extent that the participants had developed explanatory biological theories. Features were differentially weighted in categorization decisions, and the weighting was somehow a product of people's knowledge about the interrelations between features.

To explain these data and others, theorists like Keil (1989, 1994), Carey (1985), and Murphy (1993; Murphy & Medin, 1985) argue that features are weighted in categorization decisions according to the centrality of the feature in an "intuitive theory" of the category. Unfortunately, a precise model of centrality with respect to an intuitive theory has yet to be articulated. The language of intuitive theories has not been sufficiently well-formulated to support a precise notion of centrality. We will try to offer such a notion. To do so, we will ignore, or at least abstract from, much of the structured knowledge that the intuitive theory view seems to presuppose. We will rely on the simple mechanics of constraint satisfaction. Our model of centrality can thus be understood as a bridge between the intuitive theory view of conceptual structure and the more impoverished, but more clearly articulated, feature-based approach.

Mutability and Dependency Structure

The centrality of a feature represents the degree to which the feature is integral to the mental representation of an object, the degree to which it lends conceptual coherence. We will therefore measure the degree of coherence associated with a feature by asking people how easily they can transform their mental representation of an object by eliminating the feature, or by replacing the feature with a different value, without changing other aspects of the object's representation. We call such judgments measures of "mutability" (cf. Kahneman & Miller, 1986), because they reflect how much a feature of a concept of an object can be mentally transformed.

This paper has two objectives. First, we will argue that the mutability of conceptual features can be represented as a single, multiple-valued dimension. We will show that the features of a concept can be reliably ordered with respect to the degree to which people are

willing to transform the feature while retaining the integrity of a representation; i.e., that a number of conceptual tasks, all of which require people to transform conceptual features, produce similar orderings. Following Medin and Shoben (1988), these tasks have in common that they ask people to consider an object that is missing a feature but is otherwise intact (e.g., a real chair without a seat).

Note that the mutability of a feature is concept-relative. For example, roundness is a mutable feature of oranges. Even if all oranges were round, our notion of orange would be substantially unaltered if we imagined one that was not. But roundness is an immutable feature of wheels. If a wheel is not round, then it has to be completely reconceived to retain its (mental) status as a wheel.

The second objective of this paper is to show that immutability can be modeled as the centrality of a feature in a network of pairwise dependency relations. In particular, we test the hypothesis that features are immutable to the extent that other features depend on them. The idea is that a feature is immutable if changing it would force other aspects of the object concept to change, in the same sense that the bottom book in a stack of books is central because removing it would shift the other books. Applying this idea to Keil's (1989) paradigm, the suggestion is that older children and adults gave internal features more weight in categorization decisions because internal features tend to be more immutable than perceptual ones. Removing the internal features of an animal destroys the integrity of the representation of the animal as an instance of the animal kind more than removing the perceptual features does. Perceptual features depend on internal ones more than internal features depend on perceptual ones, so that substituting the internal parts of a horse (say) with those of a cow plays havoc with the dependency structure that binds the components of our concepts of horse.

Our appeal to dependency structure explains why the mutability of a feature is concept-relative. The mutability of a feature for a concept depends on its relation to other features within that concept. Few features of oranges depend on roundness but many features of wheels do.

Our dependency hypothesis makes three key assumptions: i. Dependency relations are asymmetric; ii. They are generic; i.e., their type is irrelevant to determining centrality; iii. The centrality of a feature is a function of the extent to which other features depend on it. The force of this last assumption is that we view a feature as immutable to the degree that changing it would affect the status of other features. Because we model centrality structurally, in the sense that it is a measure of the location of a feature in a dependency structure that binds a concept, centrality is deemed an immediate product of the forces that determine conceptual coherence. Specifically, our model calls a feature central in a dependency structure if the processing dynamics operating on that structure give the feature more influence over other features than other features have over it.

Our account of centrality in conceptual structure has a resemblance to Quine's (1951) account of centrality of statements in the determination of truth. Statements are central, according to Quine, to the extent they enjoy immunity from revision. A statement is immune from revision if revising it would force a cascade of revisions of other beliefs. Putnam (1975) argues that features that are central in Quine's sense serve as the most natural

category-markers in a host of contexts. Their immutability and consistency give them the power to support the most stable systems of classification.

What Mutability Is Not

Of course, features differ on dimensions other than mutability, and those other dimensions can also influence the impact of a feature on different tasks. As a means of clarifying our notion of mutability, we distinguish the mutability of a feature in a concept from its category centrality, its diagnosticity, and its salience. These dimensions are all intimately related; nevertheless, they are different both conceptually and empirically. We start with their conceptual differences.

Category Centrality

Concepts and categories are, to a large extent, flip sides of the same coin. Roughly speaking, a concept is an idea that characterizes a set, or category, of objects. We construe the terms "concept" and "category" to refer to two different perspectives that a subject can take on a class of objects, what Tversky and Kahneman (1983) call the "inside" and the "outside" views. The inside view regards the internal structure of a concept, its features and what binds them together. The outside view regards some or all of the instances that are believed to be included in the category referred to by the concept. We take mutability to be determined by centrality in a concept's feature space—in an inside view of the concept. Other measures, like naming and variability judgments, are largely determined by the boundaries of a category's instance space—in an outside view of the concept. Because the inside and outside views of a class of objects are usually compatible, measures of conceptual and category structure usually coincide. As we will see, however, situations can be constructed to make them diverge.

Diagnosticity

The term "diagnosticity" has multiple senses. Sometimes it is used to refer to the informational value of a feature for one category relative to a set of categories. If one's task is to categorize different kinds of apples, color may prove highly diagnostic because it excels at distinguishing some types of apples from others. This sense of diagnosticity is well-captured by the likelihood ratio $P\{C_1|F\}/P\{C_2|F\}$which states that the informational value of a feature F for a category C_1 is high in proportion to the probability of the category given that we know the feature relative to the probability of some other category C_2 given the feature. When we want to know how much evidence F provides for C_1 alone, the likelihood ratio is monotonically related to $P\{C_1|F\}$, or what has been called cue validity (Rosch, 1978). Tversky (1977) demonstrated how a feature's weight for determining similarity can change as a function of its informational value within a set of categories.

Informational value is clearly not identical to mutability because some attributes have high informational value but are mutable. For example, "having bones" has little informational value for robin because many things have bones but are not robins. However, it is an immutable feature of robin; imagining a robin that does not have bones but nevertheless

has all the other characteristic features of a robin is difficult. The first study will verify this claim.

Diagnosticity is also used to refer to the inferential potency of a feature. Some authors (e.g., Franks, 1995) call features diagnostic of a category to the extent that they allow us to infer other features of the category. Inferential potency increases with the degree to which the feature is predictive of other features. For example, an animal's shape usually has high inferential potency because it allows us to make inferences about other of its attributes like speed and ferocity. An example of a feature that has little inferential potency is color, especially of an artifact category like house. The knowledge that a house is blue tells us little about the house that we did not already know. The importance of predictability in categorization has been emphasized by others, including Anderson (1991) and Billman and Heit (1988). In Billman's focused sampling theory, the salience of a feature (the probability that the feature will be sampled or attended to during learning) is increased if it successfully predicts values on other dimensions. Billman and Knutson (1996) showed that people learn categories with intercorrelated features better than those with isolated features because the former allows one to predict values of other features.

Conceptually, inferential potency is distinct from informational value. Some features have low inferential potency but high informational value. Features that are unique to a category have high informational value for the category, but might provide few clues to the category's other properties and so have little inferential potency. The names of unfamiliar people or objects have this character.

The inferential potency of a feature is also distinguishable from its mutability. A feature gains inferential potency by virtue of its statistical relations to other features; in contrast, a feature is mutable by virtue of its dependencies. Statistical relations can correspond to dependency relations, but they need not. Some features, like the invariant properties of an animal's shape, are inferentially potent and immutable. But others are inferentially potent yet mutable. For instance, *has buttons* is statistically correlated with *is colored, has material, has a zipper,* and *has sleeves* (Malt & Smith, 1984) and therefore inferentially potent, but nevertheless highly mutable because little about clothing depends on having buttons.

Salience

Salience refers to the intensity of a feature, the extent to which it presents a high amplitude signal in relation to background noise, in a way that is fairly independent of context. For example, the brightness of a bright light or the redness of a fire engine are salient features. Again, a highly salient feature is not necessarily immutable as in the stripes of a zebra.

SUMMARY AND PLAN OF THE PAPER

In summary, the mutability of a feature in a concept is a measure of conceptual centrality, people's willingness to transform the feature in a representation of an object while retaining the belief that the object is represented by the concept. This dimension of conceptual structure can be distinguished from category centrality, diagnosticity, and salience. Study

1 attempts to do so by examining correlations amongst measures of these dimensions. Next, we propose a model of conceptual centrality. The model is tested in Studies 2 and 3.

So far, our notion of mutability is underspecified. Concepts can be transformed in different ways, and features appear to be more or less immutable in each case. For example, *having feathers* is mutable of bird in the sense that imagining feathers plucked is easy, but relatively immutable in the sense that imagining an adult bird that never grew feathers is hard. Our measures of mutability will attempt to focus on the latter, more enduring, sense. Our effort to uncover the systematic aspects of mutability judgments will involve a process of elimination. We will incrementally exclude interpretations in an effort to hone in on a core meaning. In particular, Studies 4 and 5 will exclude some of the interpretations of mutability afforded by Study 1; in particular, we will attempt to show that mutability measures conceptual structure, not category structure.

Study 1: Distinguishing Centrality from Diagnosticity and Salience

The purpose of Study 1 is both to demonstrate that conceptual centrality can be reliably measured and to show that it can be dissociated from measures of diagnosticity and salience. For this purpose, we asked people to make a variety of judgments for a number of features of four different categories. Our mutability measures were intended to gauge the amount of conceptual coherence lent by a feature; i.e., the degree to which the feature's presence in an object contributes to a person's certainty that the object is represented by a concept. We assessed conceptual coherence using several measures of certainty. To assess the contribution to coherence made by a feature, we examined the effect of removing that feature. As detailed below, six other measures were also included, two measures each of category centrality, diagnosticity, and salience. All rating tasks are listed in Table 1, along with an example, underneath the dimension each is intended to measure.

Feature Measurement Tasks

Study 1 focuses on the correlations obtained between the 10 feature measures. Our prediction is that tasks which measure the same underlying featural dimension (conceptual centrality, category centrality, diagnosticity, or salience) will be highly correlated; they will converge on an ordering of features within a concept. Tasks which measure different attributes will not necessarily be correlated.

Mutability Measures

Surprise

The first measure of mutability is the surprisingness of an instance missing the target feature. Instances missing mutable features should be less surprising than instances missing immutable features on the assumption that surprise is related to the difficulty of adapting an object representation to a concept. Adaptation should be easy if the object is missing a mutable feature; but hard if it is missing an immutable feature. We asked people how sur-

TABLE 1
Psychological Dimensions Distinguishing Conceptual Features (in capitals), Each with Corresponding Category-Feature Measures Employed in Study 1 (in italics)

1. CONCEPTUAL CENTRALITY: Mutability, or the degree to which a feature in a concept can be transformed while maintaining the concept's coherence.
   Surprise.
   E.g., How surprised would you be to encounter an apple that did not grow on trees?
   Ease-of-imagining.
   E.g., How easily can you imagine a real apple that does not grow on trees?
   Goodness-of-example.
   E.g., How good an example of an apple would you consider an apple that does not ever grow on trees?
   Similarity-to-an-Ideal.
   E.g., How similar is an apple that doesn't grow on trees to an ideal apple?
2. CATEGORY CENTRALITY: perceived relative frequency of an instance within a category.
   Counterfactual Naming.
   E.g., Would something be called an apple even if it did not ever grow on trees?
   Variability (transformation of category validity, Pr{feature | category}).
   E.g., What percentage of apples grow on trees?
3. DIAGNOSTICITY:

i. Informational Value: evidence provide by a feature for one category relative to a set of categories.
   Cue Validity (Pr{category | feature}).
   E.g., Of all things that grow on trees, what percentage are apples?

ii. Inferential Potency: the extent to which a feature allows other features to be inferred.
   Inferential Potency.
   E.g., What proportion of an apple's features would you predict were present in an object if all you knew was that the object grows on trees?

4. SALIENCE: signal-to-noise ratio.
   Prominence.
   E.g., How prominent in your conception of an apple is that it grows on trees?
   Availability (speeded).
   E.g., An apple grows on trees (yes/no).

prised they would be to encounter a transformed instance. For example, "how surprised would you be to encounter an apple that did not grow on trees?"

Ease-of-Imagining

Feature centrality should also determine people's ability to construct a mental image of an object that is missing the feature of interest. We asked people how easily they could imagine an actual instance of the category without the feature; e.g., "how easily can you imagine a real apple that does not grow on trees?" Imagining an instance missing a mutable feature should be easy because little depends on it, so that it can be easily removed from a representation. Imagining an instance missing an immutable feature should be hard because transforming it requires transforming many other features. After providing responses for each category, participants were asked to list all items for which they had drastically changed their conception of the category (e.g., thinking about a toy robin rather than a real one). These items (about 9 percent) were discarded.

Goodness-of-Example

If mutability reflects the degree of structural coherence provided by a feature, then transformations of mutable features should affect the typicality rating of a category less than transformations of immutable ones. One standard measure of typicality is goodness-of-example. Hence, we asked people to rate the goodness-of-example for a category of an instance missing the critical feature. We asked participants questions with the same general form as "how good an example of an apple would you consider an apple that does not ever grow on trees?"

Similarity-to-an-Ideal

Finally, we asked participants how similar an instance missing a feature is to an ideal instance. We defined an ideal as an instance having all the features of the category, both mutable and immutable. Transforming features should be perceived to violate this ideal conception to the extent that features are immutable. We asked participants, for instance, "how similar is an apple that doesn't grow on trees to an ideal apple?"

Measures of Category Centrality

Counterfactual Naming

As a measure of category centrality, we asked people whether an instance would remain a member of a name category even if it did not have the feature. For example, "Would something be called an apple even if it did not ever grow on trees?" Unlike our other measures, this question used the modal "would." This modal suggests that the question has a conventionally appropriate answer; after all, name categories are primarily matters of social agreement. We maintain that judgments of name appropriateness rely more on beliefs about category membership—on an outside view of a category—than on beliefs about conceptual structure—on an inside view (see Ahn & Sloman, 1997, and the introduction to Study 4 below for more detail). We posit that, when asked whether a category label for an object is appropriate, people have a tendency to consider their experience with instances of the category, and not the degree of match between the object and the concept elicited by the label.

Variability

Features can be more or less stable across category instances; i.e., they are differentially variable. Note that variability represents, to a large extent, the same empirical structure as mutability. Judgments of mutability reflect the degree to which a feature can be mentally transformed; judgments of variability reflect estimates of actual feature transformations across remembered category instances. To examine the relation between judgments of mutability and variability, we asked subjects to estimate the percentage of category members displaying a feature (e.g., "what percentage of apples grow on trees?"). This measure is often called category validity. Because we are treating all features as binary, we were

able to convert those judgments to variability judgments by transforming them using the binomial variability measure [X/100*(1 – X/100)].[1]

Measures of Diagnosticity

Cue Validity

To measure informational value, we used the most common measure of diagnosticity, cue validity. This is the probability of a category given knowledge of the feature. We asked participants to estimate the proportion of all instances with the target feature that were members of the specified category. For example, "of all things that grow on trees, what percentage are apples?"

Inferential Potency

As a measure of participants' estimates of the extent to which a feature allows other features to be inferred, we asked them questions like "what proportion of an apple's features would you predict were present in an object if all you knew was that the object grows on trees?"

Measures of Salience

Prominence

One indication that a feature is salient is that it is prominent in a representation; it jumps out at the conceiver. Hence, one of our measures of salience was how prominent the feature seemed to participants when they thought about the category. We told them "we are interested in what aspects of various objects seem most prominent or noticeable when you think about the object. How much do certain characteristics or features 'jump out' at you when you consider an object?" and asked them, for example, "how prominent in your conception of an apple is that it grows on trees?"

Availability

A second indication of a feature's salience is that people have easy access to it. When asked in a speeded task whether a category displays a feature, they should be more likely to respond "yes" if the feature is salient of the category. We asked them to respond "yes" or "no" to statements like "Apples grow on trees."

Materials

All 10 ratings were collected for features of two artifact categories, *chair* and *acoustic guitar*, and two natural kinds, *apple* and *robin*. The features were taken from an independent source, Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976b), and appear in Table 2. Rosch et al. report 9 features for each category except *robin*, for which they report 14. Following Rosch et al., we treat all features as binary.

TABLE 2
Features from Most Immutable to Most Mutable According to Factor Scores.

CHAIR:	ROBIN:
• has a seat.	• has a beak.
• you can sit on it.	• has wings.
• has legs.	• eats.
• can hold people.	• has feathers.
• has a back.	• moves.
• has four legs.	• has two legs.
• has arms.	• can fly.
• is made of wood.	• has a red breast.
• is comfortable.	• chirps.
	• eats worms.
	• is small.
	• builds nests.
	• is living.
	• lays eggs.
ACOUSTIC GUITAR:	**APPLE:**
• has a neck.	• grows on trees.
• makes sound.	• has a core.
• you strum it.	• has seeds.
• makes music.	• has skin.
• has a hole.	• you eat it.
• has strings.	• is round.
• has tuning keys.	• is juicy.
• made of wood.	• has a stem.
• used by music groups.	• is sweet.

Note. Factor scores were derived using the regression method from a principal components analysis followed by an oblique rotation.

Participants

Each rating was made by a different group of 20 Brown University undergraduates who were paid for their participation with the following exceptions: the same group provided prominence and cue validity judgments; a group of 30 provided availability judgments; and similarity-to-an-ideal judgments were provided by 20 Northwestern University undergraduates who received course credit for their participation. Measures of the same attribute, such as the four measures of mutability, were obtained from different groups to prevent spurious correlations arising from individual differences.

Procedure

Each category was tested separately. For each measure, a series of questions was asked, one for each feature, about an instance of the category. The form of the questions was identical to the corresponding examples in Table 1. Mutability, counterfactual naming, and inferential potency judgments differed from the other judgments in that the questions presupposed knowledge of the features an instance of the category would have. For example, the ease of imagining "a real apple that does not grow on trees" depends on what other fea-

tures a real apple is ascribed. Therefore, before making these judgments, participants were told what the features of each category were, and were asked to consider only "real" category instances; not, for example, toy or stuffed ones. The other measures were collected without first showing participants all the category's features because knowledge of those features should not have affected their judgments.

All ratings were made on a 0–1.0 scale except similarity-to-an-ideal which used a 0–10 scale, cue validity and category validity which used 0–100 percentage estimates, and availability which used 0 or 1 (no or yes) responses.

All ratings but the availability judgments were collected by questionnaire in small groups. Participants were shown a sheet of instructions describing the task and providing two example judgments with justification. Next, participants provided a rating for each feature within each category. The same random order of features was used for each participant and each category appeared in each serial position an equal number of times. Participants proceeded at their own pace. Participants were encouraged to ask questions at any time and were told that they would not have to justify their responses.

The availability judgments were collected individually at computer terminals. The task was speeded and reaction times were collected although will not be reported here. Each participant read statements from all four categories, such as "A robin has wings," and typed 'P' to respond yes and 'Q' no. Mixed in with statements were blatantly false ones, like "A robin is made of wood." False statements were constructed using features from the other tested categories. Each participant was tested on the 41 true statements from all four categories and 41 false statements in 1 of 10 random orders.

Results and Discussion

For each feature, the mean rating across participants for each category-feature was calculated. In the case of availability, this refers to the mean proportion of subjects who said "yes." Correlations across all 41 category features are reported in Table 3 for each pair of tasks. Correlation matrices for each of the four categories appear in Appendix A.

TABLE 3
Pearson Correlations between 10 Category-Feature Rating Tasks

TASKS	sprz	e-of-i	gd-ex	sim	name	var	cue valid	infer poten	prom	
surprise	1									
ease of imag	–0.88	1								
goodness of ex	–0.86	0.87	1							
sim to ideal	–0.79	0.79	0.84	1						
naming	–0.87	0.84	0.91	0.79	1					
variability	–0.75	0.66	0.71	0.75	0.68	1				
cue validity	–0.018	–0.039	–0.15	–0.052	–0.24	0.065	1			
infer potency	0.27	–0.23	–0.34	–0.14	–0.35	–0.18	0.68	1		
prominence	0.20	–0.31	–0.51	–0.39	–0.44	–0.14	0.21	0.068	1	
availability	0.16	–0.093	–0.075	–0.092	–0.049	–0.053	0.057	0.33	–0.11	

Note. Column labels refer to the same tasks, in the same order, as the row labels.

One pattern in the correlation table is particularly prominent. The correlations amongst the first 6 tasks have markedly high absolute magnitudes. All of these tasks are related to centrality, either conceptual centrality (mutability) or category centrality, and all correlations with the mutability tasks are in the direction predicted by a common underlying mutability scale.[2]

The four mutability tasks and counterfactual naming are similar in a sense because they all ask people to imagine instances without a feature, so some correlation amongst them should have been expected. The tasks are not identical of course; neither are they paraphrases of one another. Surprise, imagining, similarity, goodness-of-example, and naming are not obviously related on their own terms, although commonalities have often been demonstrated amongst the latter three (reviewed in Goldstone, 1994), but so have differences (e.g., Carey, 1985; Gelman & Markman, 1986; Keil, 1989; Rips, 1989). All five tasks, as well as the variability task, do seem to draw on a common resource, the degree to which a concept seems normal. This is consistent with our position: People can quickly and easily evaluate the extent to which a feature of an object is transformable. The tasks became similar merely by asking people to consider an object with a feature removed.

The four mutability tasks and counterfactual naming are also similar in that they were all preceded by lists of objects' features. However, this common procedural element cannot explain their high correlations because they were also correlated with variability judgments, which were not preceded by a list of features, and they were not highly correlated with inferential potency judgments, which were preceded by a list of features.

The correlations obtained in Study 1 did not distinguish conceptual centrality from category centrality. Tasks that required an inside view were highly correlated with tasks that required an outside view. Studies 4 and 5 below focus on this issue.

The two measures of diagnosticity, cue validity and inferential potency, correlated with each other (0.68), but not with any other task. Cue validity judgments did not correlate with other tasks in part because they were consistently low, the highest average judgment for any feature was 34% (the percentage of things that have a red breast that are robins). No feature was judged to pick out one and only one category (even "has a seat" is not unique to chairs). Solitary features tend to give little evidence in favor of any single, specific category which contributes to their inability to predict mutability (or salience). However, the low correlations cannot be dismissed as an artifact of range attenuation because cue validity did correlate with inferential potency and because inferential potency also did not correlate with other tasks, even though its range was not attenuated (normalized for scale, inferential potency had a larger range than 5 of the other tasks).

A surprising finding is the low correlation between prominence and availability (–0.11). In part, this was because features were verified of the category a majority of the time so availability (feature verification) was not a sensitive measure. Hence, mutability must measure some conceptual property beyond mere knowledge about the presence or absence of features.[3]

Factor Analysis

To make the correlation matrix of Table 3 easier to interpret, we used factor analysis to reduce the dimensionality of the task space. Factor analysis fits a model positing a set of factors composed of linear combinations of the original variables with the aim of generating a close approximation to the correlation matrix using an appreciably smaller number of factors than original variables. In this case, the original variables are the different tasks. The weight afforded each task on a factor is called its "loading" on that factor. Factor analysis consists of an extraction phase, "underlying" factors are extracted from the correlation matrix, and a rotation phase, the resultant factors are rotated to increase their interpretability by making the loadings closer to 0 or 1. A representative solution across all categories using principal components analysis followed by an oblique rotation (the "oblimin" method described by Harman, 1967) appears in Table 4. The loading solution was robust for different extraction and rotation methods. The solution presented used a single correlation matrix for all categories. Similar results were obtained using separate matrices for each category.

The first factor to emerge (accounting for 53% of the variance in the correlation matrix) was a factor in which all the tasks related to centrality loaded highly (absolute value above .85) and other tasks—cue validity, inferential potency, prominence, and availability—hardly loaded at all (absolute value below .30).

Cue validity and inferential potency loaded highest on the second factor, a factor reflecting feature diagnosticity. Their loadings are 0.94 and 0.86, respectively. The second factor accounted for another 17% of the variance. The third factor, accounting for 12% of the variance, was salience. Availability loaded highest on it (0.76) and prominence is the only other measure to make a noticeable appearance (its loading was –0.68). The negative loading for prominence reflects its small negative correlation with availability. All other factors had eigenvalues less than 1.0.

The intent of presenting this analysis is to increase the comprehensibility of Table 3's correlation matrix by providing some additional descriptive statistics. The results can be

TABLE 4
Factor Loadings of 10 Conceptual Tasks on the First 3 Factors from a Principal Components Analysis Followed by Oblique Rotation

	Factor 1	Factor 2	Factor 3
surprise	–0.96	–0.055	0.14
ease of imagining	0.92	0.017	0.0067
goodness of example	0.92	–0.13	0.14
similarity to ideal	0.91	0.051	0.079
counterfactual naming	0.88	–0.19	0.13
variability	0.86	0.16	–0.12
cue validity	0.15	0.94	–0.13
inferential potency	–0.12	0.86	0.28
prominence	–0.29	0.26	–0.68
availability	–0.13	0.25	0.76

Note. Factor 1 appears to reflect centrality, Factor 2 diagnosticity, and Factor 3 salience.

summarized as showing that the primary factor describing the particular set of measures that we used is centrality, the second is diagnosticity, and the third is salience.

Study 2: Reducing Mutability to Dependency

Study 1 provided evidence that centrality is psychologically real in the sense that the ordering of features by centrality converges across a number of different cognitive tasks and is different than the orderings produced by diagnosticity and salience. Study 2 now explores the determinants of centrality. We pointed out above that the assumption is often made that a feature is immutable to the extent that it is central in an intuitive domain theory (e.g., Keil, 1989). This notion has not yielded to precise specification, however, because the notion of theory is not well-specified. More success might be achieved in modeling centrality by disregarding whatever content is embodied by the notion of a theory and focusing instead on the formal structure of the relevant concept. Such a change in focus has at least two advantages. First, it allows us to ignore a critical but poorly understood quality of a theory, that it maintain global coherence. Theories, in the scientific sense, strive for global consistency between their internal aspects—their postulates, laws, principles, rules, types and tokens—and between themselves and the data that they purport to explain. Theories also strive for completeness; i.e., to explain as much of the data as possible. But maintaining global coherence is hard (indeed, it is often computationally intractable; e.g., Boolos & Jeffrey, 1980) and, for concepts, not well-defined. What would it mean for our concept of *lawn* or *friend* to be globally coherent when their extensions are not clear?

We can reduce the demands on our model by employing a measure of centrality based on local relations between features, a measure that allows the possibility that two parts of a concept will be incoherent (as some people's superstitions concerning ladders do not cohere with their beliefs about the material essence of ladders). Thagard (1989) offers a view of conceptual coherence based on local and not global relations. He suggests that concepts cohere by virtue of their multiple explanatory links. Extrapolating this view, features are central to the extent that they serve to explain the existence of other features. For example, bones are central to our concept of *bird* because having bones helps to explain the structure and operation of bird parts and functions. This view of centrality is local in the sense that it is based on local explanatory links, each of which binds a pair of features. Of course, to the extent that one believes an intuitive theory is merely a concatenation of explanatory links, then the local coherence requirement is not new.

The task of modeling centrality becomes further simplified by assuming even less structure. Centrality is easy to define if we make our explanatory relations content free by assuming they come in only one kind (Thagard, 1989). We call these unlabeled relations "dependencies" because the term denotes an asymmetric relation but is otherwise general (it covers causal, categorical, indeed any kind of directional relation between features). By assuming a single kind of relation, we can model centrality using a simple associative network. Our hypothesis is that centrality can be reduced to one kind of pairwise dependency relation between features, not that all cognitive phenomena can be reduced to unlabeled relations. Tasks, like analogy, which require analysis of a concept into its constituent fea-

tures and consideration of the specific relations those features play, will require consideration of labeled relations.

We hypothesize that a feature is central to the extent that other (central) features depend on it, and that it will be judged immutable in proportion to its centrality. Preliminary evidence in support of these hypotheses is reported in Love (1996). He shows that priming people with a feature that is dependent upon a second feature increases judgments of the immutability of the second feature, but presenting the second feature has no effect on judgments of the first. For example, in the context of birds, the ability to fly depends on having wings whereas having wings does not depend on the ability to fly. Presenting *can fly* increased the immutability of *has wings*, but presenting *has wings* had no effect on the immutability of *can fly*. Increasing the availability of a feature that depends on a second feature increased the immutability of the second feature.

We further test our hypotheses by implementing them in a simple, iterative, linear equation. Specifically, let D be a matrix to be empirically determined that represents the pairwise dependencies between all features of a concept. A particular cell of D, d_{ij}, is a positive number representing the degree to which feature j depends on feature i. D should also be indexed by the relevant concept, but this will always be clear from context and so will be omitted. We express our centrality hypothesis as

$$c_{i,t+1} = \Sigma_j \, d_{ij} \, c_{j,t} \tag{1}$$

where $c_{i,t}$ is the centrality of feature i at time t. According to the equation, the centrality of feature i is determined at each time step by summing across every other feature's degree of dependence upon feature i multiplied by that feature's centrality. In terms of mutability, if a highly immutable feature depends upon feature i, feature i becomes more immutable than if a mutable feature were instead to depend on it. Furthermore, the feature would be much more immutable if a feature depended upon it that other features, in turn, depended upon. The iterative nature of Equation (1) is intended to accommodate such non-local effects.

To implement the model, immutability ratings must be set to some initial arbitrary value, a value that is almost always inconsequential for the final state because Equation (1) is linear. In the implementations below, all $c_{i,0}$ were set to 0.5. The model iterates until it converges. Mathematically, the model is a repetitive matrix multiplication and is known to converge to a solution in a small number of steps (Wilkinson, 1965); specifically, it converges to a family of vectors in the direction of the eigenvector of the dependency matrix with the largest eigenvalue. The model converges when it is attracted to a state in which satisfactory immutability assignments are made for all features simultaneously.

What are Dependency Relations?

Dependency relations exist between features. Of course, features can be defined at a variety of levels of abstraction. In a formal sense, features are isomorphic to classes (the feature of being a multi-millionaire athlete picks out the same people as are encompassed by the class of multi-millionaire athletes). Classes obviously come at various levels of abstraction and, due to the isomorphism, features must too. The convention in the study of categorization is to assume a basic-level of categorization (see Murphy & Lassaline, 1997, for a

review). We make the corresponding assumption for feature parsing: a basic-level for identifying psychologically relevant features, a level which achieves generality while avoiding vagueness. Our dependency relations lie between these "basic" features (see Figure 1 for examples).

We take dependency relations to be very general representational elements. Every directional, semantic relation between features can be treated as a generic dependency relation. Presumably, like all relations, they can be formed in either of two complementary ways, through learning of featural covariation or through a process of explanation. The relative contribution of these two sources of knowledge is as yet unknown. To the extent that dependency relations do reflect people's explanations of objects and events, we might appeal to intuitive theories to account for mutability judgments. Such theories would define central features as those on which explanations for our common interactions with objects and events depend, explanations whose force can be represented as an associative strength.

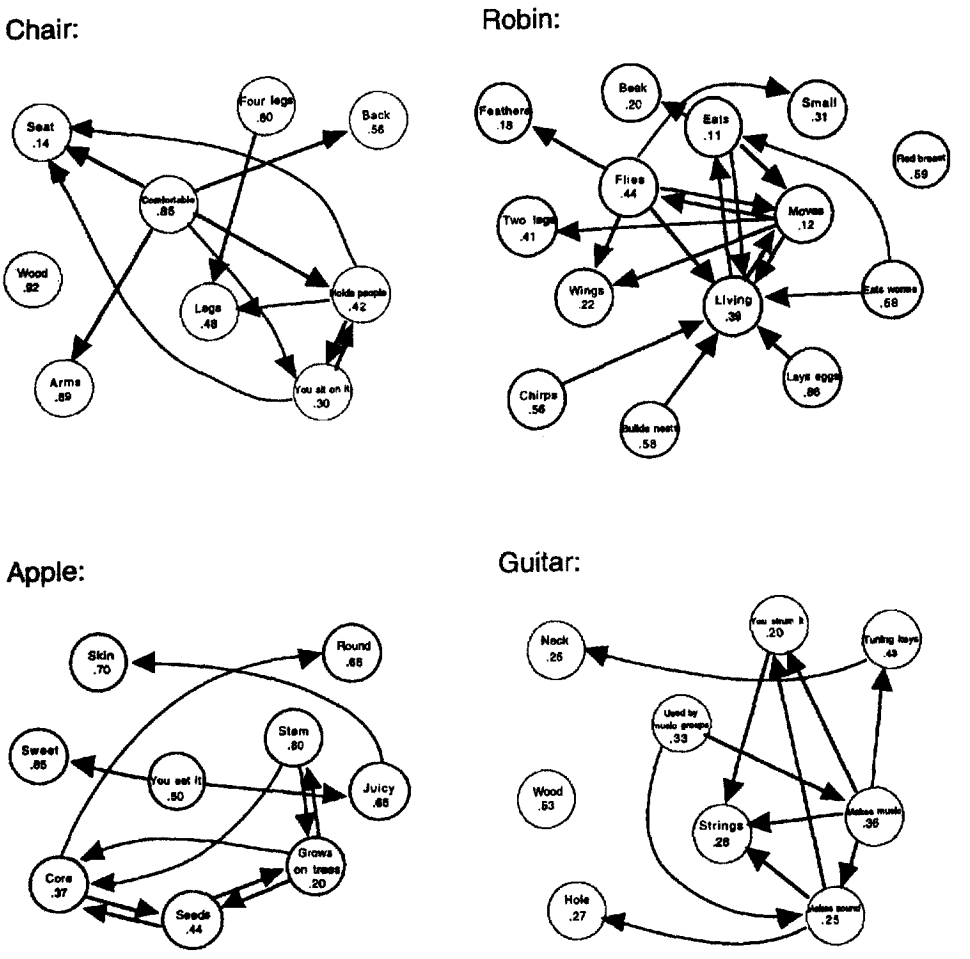
The strategy adopted by this essay is to cover as much empirical ground as possible with minimal theoretical edifice. We attempt to analyze a key attribute of conceptual structure using only an iterative, linear combination of asymmetric dependency strengths. The use of such simple parallel computations is probably limited to fast, associative processing. Conclusions that require slower, more deliberative and analytic, rule-based processing depend on specific attributes of the relevant relations, like their label, level of abstractness (Clement & Gentner, 1991), and projectibility (Goodman, 1955).[4] In other words, we propose that early access to concepts ignores the content of relations; only slower processing makes use of it. One cognitive activity that involves such slow processing is the *explicit* application of (naive or sophisticated) scientific theories to categorization. Although people sometimes have access to explanatory principles that make categories coherent (e.g., Murphy & Medin, 1985; Rips, 1989), our model suggests that those principles do not directly govern judgments of mutability. Our claim is that, when used to make quick judgments not requiring justification, the cognitive system processes multiple generic dependency relations in parallel.

Measuring Dependency Relations

To test the model, we use it to derive centrality values (the vector C). C should be highly correlated with our participants' immutability judgments. To derive C using Equation (1), estimates of matrix D are required. To obtain such estimates, a group of 20 Brown University students were shown, simultaneously, all the features of a particular category from the previous study. Each feature was inscribed in a circle and participants were asked to draw arrows from each feature to every other feature that they believed the feature depended on, creating a graph like those shown in Figure 1. Three different colored markers were used to indicate the strength of the dependency. The weakest links were assigned the value 1, medium links 2, and the strongest links 3. Instructions were clarified using a graph of the category "12," with mathematical features like "can be divided by 6."

Results and Discussion

Figure 1 displays each category along with its features and estimated dependency links. To maintain the readability of the graphs, only the strongest dependencies have been drawn. A link is shown if its mean judged dependency strength was above some criterion. Criteria were set to maintain an average of 1.25 links per feature, although all dependency information was used in our simulations. Figure 1 also shows the mean ease-of-imagining judgment for each feature, inscribed in its respective node (higher values represent higher mutabilities). The graphs illustrate that, as predicted, features with few other features depending upon them tended to have been judged mutable while features with many features depending upon them tended to be judged immutable.



Note. The arrows point from a feature to one that it depends upon. Mean ease-of-imagining judgements are also shown beside each category-feature

Figure 1. Category Graphs

TABLE 5
Rank correlations between modeled dependency (Equation 1) and 10 judgment tasks for each category and across categories

Feature-rating Tasks	Dimension	Chair	Guitar	Apple	Robin	Across Categories
surprise	Mutability	0.95	0.11	0.55	0.30	0.61
ease of imagining	Mutability	−0.92	−0.62	−0.60	−0.59	−0.72
goodness of example	Mutability	−0.95	−0.63	−0.38	−0.15	−0.65
similarity to ideal	Mutability	−0.98	−0.40	−0.57	−0.65	−0.77
counterfact. naming	Category centrality	−0.93	−0.57	−0.62	0.042	−0.63
variability	Category centrality	−0.95	−0.56	−0.82	−0.38	−0.76
cue validity	Diagnostic.	0.37	0.017	0.75	−0.67	0.14
inferential potency	Diagnostic.	0.55	0.23	0.73	−0.47	0.31
prominence	Salience	0.64	0.57	−0.49	−0.011	0.22
availability	Salience	0.76	0.0089	−0.18	0.021	0.21

Note. The average correlation across categories is the mean Fisher's Z transformation of each category's correlation, converted back into units of correlation.

Centrality predictions were derived by iterating Equation (1) until it converged using mean dependency judgments as estimates of the d_{ij}. Table 5 presents Spearman rank correlations across features between each of the 10 feature rating tasks, averaged over participants, and the dependency model of Equation (1). Rank correlations are shown for each category individually and for the mean across categories. Mean correlations were obtained using Fisher's Z-transformation to reduce the effects of the skewness of correlations' sampling distributions. Correlations with the surprise task are positive because surprise varies with centrality; correlations with the other three measures of mutability are negative because they vary inversely with centrality.

The correlations between the mutability measures and the model are in the expected direction in every case. The magnitude of these correlations varies considerably across categories. *Chair* shows correlations whose absolute value is above .9 on every measure; correlations for the other categories are appreciably lower and less consistent. We believe that the mean correlations (absolute values of .61, .72, .65, and .77 for the four measures, respectively) indicate a reasonably close fit for the model given that they were obtained without any free parameters: the model's predictions are calculated without reference to the mutability data, they consider only the dependency data. This suggests the plausibility of Equation (1) and our hypothesis that a feature's immutability varies with the extent to which other immutable features depend upon it.[5]

The category centrality measures, counterfactual naming and variability, also tended to correlate with centrality (negatively, as would be expected). Like Study 1, this study does not indicate a dissociation between conceptual and category centrality. The only exception is *robin* on the counterfactual naming measure whose correlation is near 0. Note that of our 4 categories, *robin* is the only clearly subordinate one. Study 4 tests the hypothesis that conceptual and category centrality are more likely to dissociate when categories are more specific.

Finally, no consistent relation obtained between Equation (1) and the measures of diagnosticity or salience. Some of the individual category correlations for these measures have fairly high magnitude. However, some are positive and others negative, with the result that the average across categories is low. Mutability ratings are the only ones that were always in one direction, a direction that we predicted in each case.

To help interpret the correlations between mutability measures and Equation (1), we compared them to four other models: one representing the centrality of a feature as the sum of dependencies on it; another was based on Equation (1) with an added nonlinearity and a free parameter; another had a term representing the critical feature's dependence on other features; and, finally, we tested a model representing centrality as a feature's total connectivity, without regard for the direction of dependencies. The only model to fit mutability judgments closer than Equation (1) was its nonlinear counterpart requiring a free parameter, and it did only marginally better. The models and their fits are reported in Appendix B.

These analyses provide evidence in support of the hypothesis that features are immutable to the extent they are central in a concept's dependency structure. In particular, the data support the predicted asymmetry: Features are central to the extent that other features depend on them.

Study 3: Varying Dependency Type

The strongest claim embodied by our centrality hypothesis is that mutability is insensitive to the type of relation binding features. Equation (1) presumes only one kind of relation, a generic asymmetric dependency that varies in strength. In other words, we have supposed that relations can be treated as a single type for the sake of combining them to determine mutability. We test this assumption using artificial categories because varying the type of relations embodied by familiar concepts entails dubious claims about how features depend on others.

Study 3a

Using fictitious disease categories, Ahn and Lassaline (in preparation) have varied the dependency structure of symptoms. Figure 2 displays 3 symptoms associated with the disease Yorva. The symptoms are not independent; rather, symptom S causes symptom V which, in turn, causes symptom L. Ahn and Lassaline taught people about such diseases and then asked them to judge the likelihood that someone had disease Yorva if either they displayed symptoms S and V, but not L, or symptoms V and L, but not S. (No figure was presented to the participants in this study.) The results showed that people assigned higher likelihood judgments in the case that L (the "resultant" symptom) was missing than in the case that S (the "causal" symptom) was missing. They judged causes to be less mutable than effects when evaluating disease likelihood.

Our hypothesis is that this result does not occur because of the causal relations per se, but rather because of the asymmetric dependency structure of the diseases. We predict that similar results can be achieved using other kinds of non-causal dependency relations. To

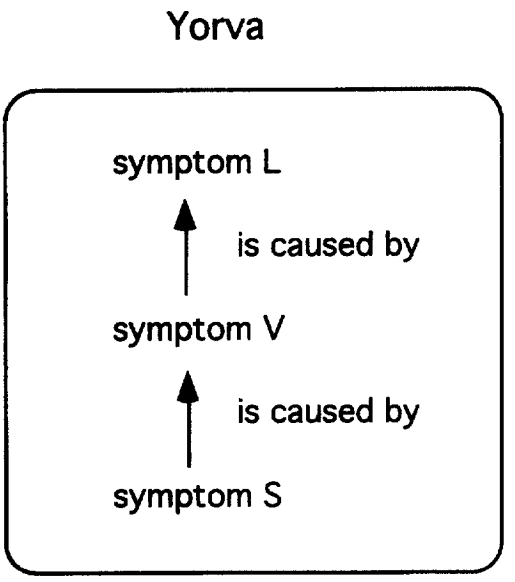


Figure 2. Example of Symptom Dependency Structure

test this prediction, we attempted to replicate Ahn and Lassaline's (in preparation) result using temporal and generic dependency relations as well as causal ones.

Method

Participants

Twenty-five undergraduates at the University of Louisville participated in partial fulfillment of course requirements.

Materials and Design

Four sets of problems were developed. Each set described a hypothetical disease with three characteristic symptoms (e.g, "Scientists have found that symptoms **A**, **W**, and **K** are associated with Disease **Xeno**"). Different alphabets and disease names were used for the four sets. In the control condition, no relation was specified among the three features. A figure was provided with the disease name on the top and three symptoms listed in a box as in Figure 2 except that no arrows were shown.

In the Causal condition, participants learned that one symptom causes the second symptom which causes the third symptom (e.g., "Scientists have found that symptoms **S**, **V**, and **L** are associated with Disease **Yorva**. In addition, the scientists have found that symptom **L** is caused by symptom **V**, and symptom **V** is caused by symptom **S**.") Along with the verbal description, they received Figure 2.

In the Temporal condition, participants learned that one symptom follows the second which follows the third symptom. In order to ensure participants understood that the temporal relation was not causal, the following instructions were provided:

> Scientists have found that symptoms **F**, **P**, and **J** are associated with disease **Betox**. In addition, the scientists have found that a symptom **follows** another symptom although there is no causal relationship between the two symptoms. Here's what we mean by the 'follow' relationship. When you go to a restaurant, for example, you first sit at the table and this follows ordering food. However, this does not mean that your sitting at the table caused you to order food. It's just that one event followed the other. In the case of the above disease Betox, the scientists have found the following relationship. Although symptom P does not cause symptom J, symptom J follows symptom P. That is, patients with Betox disease display symptom J after they display symptom P. Also, although symptom F does not cause symptom P, symptom P follows symptom F. That is, patients with Betox disease display symptom P after they display symptom F.

Next, participants in the Temporal condition received a figure similar to the one in Figure 2 with the labels, "is caused by," replaced with "follows."

In the Dependency condition, participants received similar background information as in the Temporal condition except that the relation specified was a general "depends on" relation introduced with the following example: "We are more likely to see a mustache when we see a mouth than when we do not see a mouth. We clearly know that the presence of a mouth does not cause the presence of a mustache. However, the presence of a mustache depends on the presence of a mouth." Then, participants received a figure similar to Figure 2 with the labels "is caused by," replaced with "depends on."

In each condition, participants received two problems describing a patient with one missing symptom. The two problems differed in the position of the symptom in the dependency structure although the control condition had no dependency structure, so this distinction was not made. One problem described a patient missing a feature that was more central (the other features depended on it). For example, in the Temporal condition, a patient was missing a symptom that occurs first. This problem will be called Central Missing. The other problem (Peripheral Missing) described a patient missing a feature that was least central (no other features depended on it). For example, in the Temporal condition, the missing symptom was the one that occurs last.

Procedure

Participants were instructed that they would receive some background knowledge about symptoms associated with new diseases that scientists had recently observed. They were told that for some of the problems they would also be given some information about how these symptoms were related, and their task was "to judge how easy it is for you to imagine a person with the disease who displays only a certain set of symptoms." They were told to respond on a 9-point scale with 9 being "very easy to imagine" and 1 being "very difficult to imagine."

All participants received all four sets of problems. The order of the four sets was randomized for each participant. Of the four sets, two sets presented the Central Missing prob-

TABLE 6
Mean Ease-of-Imagining Judgments for Category Instances Missing Either a Central or a Peripheral Feature Defined by 4 Types Of Relations

Type of Relation	Central Missing	Peripheral Missing	Difference
Causal	3.28	5.92	2.64
Dependence	2.88	5.60	2.72
Temporal	3.52	5.56	2.04
None	4.52	5.28	0.76
Mean	3.55	5.59	

Note. Data from Study 3a.

lem first and the others presented the Peripheral Missing problem first. This order was counterbalanced.

Results and Discussion

Mean ease-of-imagining judgments appear in Table 6. Along with replicating Ahn and Lassaline's (in preparation) centrality effect with causal relations, we obtained the effect with dependence and temporal relations, as predicted. In all three cases, removing a central feature reduced ease-of-imagining judgments more than removing a peripheral one. Moreover, in the control condition with no relations, no effect of centrality obtained. An analysis of variance shows a significant effect of centrality, $F(1, 24) = 22.92$, $MSe = 9.08$, $p < .001$, because the ease-of-imagining judgments were lower on average when the central feature (3.55) rather than the peripheral feature (5.59) was missing. Type of relation had no main effect, $p > .50$, although it did interact with feature centrality, $F(3, 72) = 2.82$, $MSe = 3.64$, $p < .05$, because the difference between central and peripheral features was greater when there were relations than when there were not, in the control condition. More importantly, contrasts between central feature missing and peripheral feature missing conditions were significant for the causal, dependence, and temporal conditions, $t(24) = 4.89$, 5.04, and 3.78, respectively, $p < .05$ in all cases, but not for the no relation condition, $t(24) = 1.41$, n.s.

Table 6 shows that the effect of centrality was almost identical for the causal and dependence relations but somewhat less for the temporal relation. However, this difference was not significant. An analysis of variance was conducted without the control condition. As predicted, type of relation did not interact with feature centrality, nor was there a main effect of type of relation, both $Fs < 1$. The only reliable effect was the main effect of centrality, $F(1, 24) = 23.53$, $MSe = 9.69$, $p < .001$. In conclusion, Study 3a confirms that the effect of centrality on judgments of mutability does not depend on the type of relation determining centrality.

Study 3b

The lack of effect of relation type in Study 3a could have been due to the presence of figures (like that of Figure 2) in the descriptions of disease categories. Participants may have relied heavily on the arrows in the figures to comprehend and retain categories' dependency struc-

tures, thereby neglecting the verbal descriptions. This neglect could have been responsible for the null effect of relation type we observed, because that variable was manipulated using verbal descriptions. In an attempt to show that our null effect was not due to the presence of figures, we replicated Study 3a without presenting figures to participants.

Method

The method of Study 3b was identical to Study 3a except that a different group of 21 University of Louisville students was tested and disease descriptions were presented without accompanying figures.

Results and Discussion

Mean ease-of-imagining judgments from Study 3b appear in Table 7. The pattern of results is virtually identical to that of Study 3a. Again, on average, excluding the central feature from the symptom list (4.40) affected ease-of-imagining judgments more than excluding the peripheral feature (6.20), regardless of the type of relation, $F(1, 20) = 18.35$, $MSe = 7.40$, $p < .001$. In addition, there was a significant interaction effect, $F(3, 60) = 5.70$, $MSe = 3.11$, $p < .01$. Again, contrasts between central feature missing and peripheral feature missing conditions were significant for the causal, dependence, and temporal conditions, $t(20) = 5.07, 2.54$, and 5.34, respectively, $p<.05$ in all cases, but not for the no relation condition, $t < 1$. Unlike Study 3a, the effect of type of relation was significant, $F(3,60) = 5.38$, $MSe = 4.26$, $p<.01$. Contrasts show that this was due to the difference between the causal (4.71) and the no relation conditions (6.30), $t(20) = 2.50$, $p<.05$, and the difference between the dependence (4.78) and the no relation conditions, $t(20) = 2.39$, $p<.05$. These differences obtained because, as predicted, ease-of-imagining judgments were relatively high in the no relation condition.

An analysis of variance was conducted without the control condition to test for differences among the conditions with relations. Again, the interaction between type of relation and feature centrality was not significant, $F(2, 40) = 2.50$, $MSe = 2.98$, $p = .10$, and the main effect of type of relation was only marginally significant, $F(2,40) = 2.58$; $MSe = 6.06$; $p = .09$. The only reliable effect was the main effect of centrality, $F(1, 20) = 17.11$, $MSe = 10.16$, $p < .001$. Thus, Study 3b replicated Study 3a without using figures in that removing

TABLE 7
Mean Ease-of-Imagining Judgments for Category Instances Missing Either a Central or a Peripheral Feature Defined by 4 Types of Relations

Type of Relation	Central Missing	Peripheral Missing	Difference
Causal	3.33	6.10	2.76
Dependence	4.10	5.48	1.38
Temporal	3.95	6.86	2.90
None	6.24	6.38	0.14
Mean	4.40	6.20	

Note. Data from Study 3b.

a central feature reduced ease-of-imagining judgments more than removing a peripheral feature no matter what type of relation bound features.

The intent of Studies 3a and 3b was to demonstrate a null hypothesis, no effect of relation type. Proving the null hypothesis is well-known to be highly problematic. That is one reason that we went about replicating the effect in Study 3b. In an experiment that we will not report here, we have replicated it once again. Note that we found no effect of relation type using a design that was powerful enough to show an effect of feature centrality.

DISTINGUISHING CONCEPTUAL AND CATEGORY CENTRALITY

Study 4: Mutability Versus Name Centrality[6]

A long history of argument in philosophy (Wittgenstein, 1953) and cognitive science (e.g., Malt, 1994; Rosch, 1978) argues that few natural concepts have defining features, at least outside the domain of mathematics (for a review, see Lakoff, 1987). Alternative views have, however, encountered difficulty explaining why people tend to call bats "mammals" and ostriches "birds" without appealing to defining features. A major motivation for the current work is to replace the dichotomous defining/non-defining distinction by a dimension of conceptual centrality that permits gradations. We believe that features differ by degree in the amount of structural support they provide for a concept.

But the notion of defining feature lurks. Imagine a blue beret. Now transform that blue beret so that it is yellow. This is easy. Blueness is a mutable property of our concept of blue berets because nothing but the beret's color is violated when the feature is changed. However, a blue beret that is yellow is not really a "blue beret" at all. Blueness is an essential property of "blue berets." So some features are essential with regard to delimiting the extension of a category and deciding whether a label is applicable without being immutable. In this sense, the centrality of a feature in a category can be distinguished from the feature's conceptual centrality.

Despite this distinction between conceptual and category centrality, the two dimensions are highly correlated. The features most responsible for binding the internal structure of a concept are generally also the ones that serve to delimit the corresponding category. However, we have pointed out one class of exceptions to this rule: features that distinguish a specific category from other similar categories, as "blue" distinguishes *blue beret* from other berets. The two types of centrality are dissociable because, when naming, sometimes we wish to discriminate an object from others that share immutable properties with it. In particular, features are more likely to be critical for naming but nevertheless mutable when discourse concerns specific—as opposed to abstract—categories. This is because, at lower levels of abstraction, categories that share a superordinate also tend to share dependency structure. For example, one kind of sedan has much the same dependency structure as another kind of sedan. Ford sedans differ in important ways from Nissan sedans, but the differences are not in their most immutable properties. Both have engines, seats, steering wheels, etc. The features that are critical for determining whether an object is a Ford as opposed to a Nissan include some that are central to the category without being conceptu-

ally central. These include the label that appears on the object's name plate, the details of the object's shape, its gas mileage, color, etc. Specific categories include features that dissociate name centrality from mutability because, by virtue of sharing their dependency structure with other categories, they will be strongly associated with features that are distinct from those of other categories and are not conceptually central. Therefore, the hypothesis of this study is that the name centrality and mutability of a feature will differ more at lower than at intermediate levels of category abstraction. We varied the level of abstraction of a variety of categories and took measurements of name centrality and mutability of selected features at each level.

Note that the features used in this experiment were selected to dissociate mutability and name centrality. Hence, the experiment is not intended to provide evidence about any general relation between features at different levels of abstraction. It is merely intended to demonstrate the validity of the distinction between name and conceptual centrality.

Method

Participants

Seventeen undergraduates at Yale University participated in this experiment. Five participated as partial fulfillment of a course requirement. Twelve received $7.00 for their participation in this and other experiments that followed it.

Materials

Nine taxonomies with two levels each were selected from Rosch et al. (1976a). The full list of the 18 categories is provided in Table 8. Among these categories, nine were specific or what Rosch et al. called "subordinate-level". Categories one level higher than each of these nine constituted the intermediate-level categories in the experiment. These correspond to what Rosch et al. called "basic-level" categories. Five taxonomies involved artifacts and four involved biological kinds.

For each category at each level, one feature was selected from the set of attributes compiled by Rosch et al. (1976b) who asked undergraduates to write down attributes for each

TABLE 8
Categories and Features Used in Study 4

Specific Category and Its Feature	Intermediate-Level Category and Its Feature
grand piano (is large)	piano (makes music)
green seedless grapes (is green)	grapes (is juicy)
Phillips head screwdriver (has a cross-shaped tip)	screwdriver (has a handle)
Levis (is blue)	pants (have legs)
living room chair (has a cushion)	chair (has a seat)
sports car (is small)	car (has an engine)
birch tree (has white bark)	tree (has a trunk)
seabass (is large)	fish (swims)
cardinal (is red)	bird (has wings)

category. Attributes not commonly listed were eliminated. The following criteria were used to select features from this set: The feature associated with a specific category had to distinguish it from other specific categories sharing a superordinate. For instance, "being large" distinguishes a grand piano from other pianos, and therefore was selected as a feature for *grand piano*. The feature associated with an intermediate-level category had to be one that generally holds in all of its subordinates. For instance, "makes music" holds for both upright and grand pianos, and therefore was selected as a feature for *piano*. The features used in the experiment appear in Table 8 in parentheses next to their corresponding category.

Two sets of questions were developed using these category-features, one for name centrality and the other for mutability judgments. Name centrality questions took the form "Suppose an object is in all ways like *X* except it does not have feature *Y*. How appropriate would it be to call this object *X*?" where *X* was one of the categories in Table 8 and *Y* was its corresponding feature. Participants answered each question using a 9-point scale where 9 indicated "very appropriate" and 1 indicated "very inappropriate." Mutability questions were in the form "Imagine an *X* that has all the usual characteristics and properties of *X*. Now, change this image of *X* so that it is in all ways like *X* except it does not have feature *Y*. Rate the ease of this transformation." Again participants answered using a 9-point scale, this time 9 indicated "very difficult" and 1 indicated "very easy."

Procedure

For the name centrality judgments, participants were instructed "to evaluate the appropriateness of a label" and were told, "Sometimes, objects seem to require certain attributes to warrant a specific label. It would be inappropriate, for example, to call a man a 'bachelor' if he were known to have a spouse. We'd like to know what properties you think are necessary to apply a label to an object. Your task is to rate how appropriate a label is for an object that is missing a specified property." Example questions were shown and discussed.

For the mutability judgments, participants were instructed "to evaluate the ease with which you can transform an image of an object." They were further told, "We'll ask you to imagine an ideal object, and then change some specified part or aspect of it. For example, imagine a door, then transform it in your mind into a door without a doorknob. We would like to know how easily you can complete this transformation. Your task, then, is to rate the ease of the transformation required to get from the original to the mutated form." Examples were discussed.

Each question was displayed on a computer screen one at a time. Participants entered their responses on the keyboard and could correct any mistake before they proceeded to the next question. The entire experiment was self-paced. Eight participants received the name centrality judgment questions first, followed by the mutability judgment questions. Nine received the mutability judgment questions first, followed by the name centrality judgments. Within each type of question, the order of the 18 questions was randomized for each participant.

Results

The order in which the participants received the two types of questions did not affect the responses. Therefore, all data were collapsed across the two orders. The name centrality and the mutability scales used are in the opposite direction in that 1 in the name centrality scale (i.e., very inappropriate to call it *X* without feature *Y*) means the feature is very central whereas 1 in the mutability scale (i.e., very easy to transform feature *Y* in *X*) means the feature is peripheral. To ease presentation, the name centrality scale was reversed so that the two scales had the same directionality.

As shown in Table 9, the name centrality ratings were close for the two category levels, indicating that the features selected in this experiment did not vary much in name centrality. However, the mutability ratings show that the features unique to the specific level were easier to mentally transform than those associated with the intermediate level. A two-way repeated-measures analysis of variance was conducted on type of task and category level. As predicted, there was a reliable interaction between the two factors, $F(1,15) = 4.68$, $MSe = 0.83$, $p < .05$. In addition, regardless of the type of task, the features selected for the intermediate-level categories were judged to be reliably more central (mean of 5.10) than those selected for the specific categories (mean of 4.11), $F(1,15) = 35.96$, $MSe = .42$, $p < .01$. This main effect of category level is due to the low mutability ratings at the specific level. Type of task also had a significant effect. The ratings on the name centrality task were reliably higher (mean of 5.66) than those on the mutability task (mean of 3.54), $F(1,15) = 42.86$, $MSe = 1.68$, $p < .01$.

The pattern of results was consistent across items. For 8 of 9 categories, the mutability judgment for the feature at the specific level was lower than at the intermediate level (the exception was "car"). In contrast, the name centrality judgment was lower at the specific than at the intermediate level for only 4 categories.

Discussion

In contrast to Studies 1 and 2, Study 4 shows a dissociation between mutability and name centrality. Changing the specificity of the object category increased the mutability of features without affecting their name centrality. Because we carefully selected the features at both levels of abstraction in this experiment, we cannot draw any general conclusions about the difference between specific and intermediate-level categories. Nevertheless, we have shown that centrality is not a homogeneous phenomenon; it has at least two aspects,

TABLE 9
Mean Mutability and Name Centrality Ratings from Study 4 for Specific and Intermediate-Level Categories and Features

	Mutability Judgment	Name Centrality Judgment
Specific	2.81	5.42
Intermediate-level	4.28	5.90

Note. On both scales, 1 means less central and 9 means more central.

conceptual and naming. One possibility is that a feature is central for naming in proportion to its relative frequency; specifically, its category validity, the probability of the feature given the category. In contrast, we propose that a feature is conceptually central (immutable) to the extent that the feature is depended on by others.

One implication of this study is that feature centrality is relative to the function being served by the feature. Concepts have multiple facets. The importance of a feature depends not only on the identity of the feature and its relation to other conceptual features, but also on the goal of the agent using the concept. In particular, using a concept to name an object requires different information about the internal structure of the concept than does transforming the concept in the service of a conceptual task like imagination.

Study 5: Mutability versus Variability

Study 4 demonstrated a dissociation between judgments of mutability and of counterfactual naming. To the extent that naming judgments reflect beliefs about category and not conceptual structure as essentialists like Putnam (1975) would have us believe, this result represents a dissociation between conceptual and category centrality. Study 5 pursues this distinction.

In the introduction, we pointed out that mutability and variability are flip sides of the same coin. Mutability refers to how much the internal structure of a concept allows a feature to transform, variability to the likelihood of transformation over a set of instances. Therefore, the relation between mutability and variability judgments should give us some insight into whether people focus on the internal structure of a concept or its extension when thinking about transformability.

Because of their common reference, judgments of mutability and variability should be highly correlated, as they were in Study 1, and Equation (1) should show reasonable fits to variability judgments, as it did in Study 2. However, the two variables are not the same. Features can be mutable and yet tend not to vary because of circumstance. A banana could be straight even if no straight bananas exist (Medin & Shoben, 1988). In America, men have rarely worn skirts although, conceivably, fashions could change dramatically. Because mutability and variability are different, the question arises as to how they influence one another. Most pertinent here, do mutability judgments measure assessments of the internal structure of a concept, as our model supposes, or do they measure assessments of the actual frequency of category instances missing the critical feature; i.e., do people substitute judgments of variability for mutability? To address this question, Study 5 attempts to show, not only that mutability judgments vary as expected when features' dependencies are manipulated, but that such a manipulation can reverse judgments of frequency; i.e., that frequency judgments are sensitive to dependency structure.

Spalding and Ross (1994) report suggestive data. Using an artificial category learning task, they had participants rank features for their importance in category membership and also by their frequency. They found that features rated more important tended to be judged more frequent than less important features, even though they were not. Most of our features are present in more than 50% of category instances. In such a situation, high frequency fea-

tures are less variable than low frequency ones, so Spalding and Ross's result is equivalent to finding that mutable features were judged more variable.

To test this question more directly, Study 5 pits mutability against variability and looks at judgments of both. Participants were shown features whose mutability was manipulated by varying their dependencies (the feature either depended on another, had no dependencies, or another feature depended on it). On the hypothesis that the mutability of a feature is inversely related to the degree to which other features depend on it, we predict that the feature depended on will be judged more immutable than the feature with no relations. On the hypothesis that mutability influences variability judgments, frequency judgments should also be greater for immutable than mutable features, even though actual frequency is in the opposite direction.

Method

Participants were given a description of an indigenous people of the Malay Peninsula. Each description referred to three features, *has a scar, has a tattoo*, and *has a cheek piercing*. Each feature played one of three roles: It bore either no relation to any other feature, it was depended on by another feature, or it depended on another feature. Twenty-nine Nothwestern University students were paid for their participation. Five students were tested in each of five of the six possible combinations of feature/role assignments. Due to an error, only four students were tested in the sixth combination.

The following is an illustration of what participants saw in one condition:

> The Semang are an Oceanic people of the Malay Peninsula who speak an Austro-Asiatic language. The Semang live in autonomous local bands led by an elder male. Each member of a band is affiliated with a particular family or clan. The different clans peacefully coexist together. Young men belonging to the dominant clan usually get a cheek piercing. Not every young male is allowed to have a cheek piercing. Only young men of the dominant clan that are old enough are allowed to get a cheek piercing.
>
> The Semang are traditional nomadic hunters and do not practice agriculture. Accordingly, the Semang place a great emphasis on hunting. Young Semang males must pass through a series of "coming of age" ceremonies before becoming a hunter.
>
> The first step is getting a ceremonial tattoo (using dyes from plant extracts). The tattoo has religious significance. After receiving the tattoo, a young male is eligible to receive a scar on his upper left arm (marking the young man's allegiance to the leader of the band). Young Semang men who do not first get the tattoo cannot get the scar. Getting the tattoo is the first pivotal step in a young male's rite of passage.
>
> Young men can become hunters and not be a member of the dominant clan. The Semang do not exhibit inter-Clan prejudices. Young men of all clans can become hunters (after they first receive a tattoo and then a scar).

In this example, *cheek piercing* is the no-relation feature, *has a scar* is the depending-on feature, and *has a tattoo* is the depended-on feature, because receiving a scar depends on having received a tattoo. Participants could refer to this description at any time during the experiment. Participants' task was to view data on a computer screen and draw conclusions about the Semang's culture based on brief descriptions of 24 Semang adolescent males (all the adolescent males of a local band). They were encouraged to look for interre-

lations among the information they saw. Each description consisted of the man's name and indicated whether each of the above features was present or absent. Two other features, uncorrelated with any of the critical features, were included in the profiles to increase the complexity of the task ("likes to eat mangos" and "has a ponytail"). The order of the features was random for each description. Subjects were free to study each profile for as long as they wished. They pressed the space bar to continue to the next profile.

The feature in the role of sharing no relation with any other feature was most frequent with a base rate of 2/3, followed by the depended-on feature with a base rate of 0.625. The base rate of the depending-on feature was 0.5, but the probability of this feature rose to 0.8 in the presence of the depended-on feature. The feature sharing no relation was uncorrelated with either of the other features. The introductory story about the Semang and the statistical relation between the depending-on and depended-on features were designed to converge to suggest a dependency between the two features.

After viewing each of the 24 profiles, participants made a series of ratings. Mutability was operationalized as similarity-to-ideal. Participants rated how similar a male missing each feature would be to an ideal male that possessed all the features. They also judged the frequency of each feature. For example, they were asked for the percentage of Semang young men who have a tattoo. Fifteen of the participants (approximately half of each of the six feature/role assignment groups) rated the frequency of each feature and then made a mutability judgment for each feature, while the other 14 participants did these tasks in the reverse order. Both groups then judged which features shared dependency relations by making pairwise judgments on a scale from 0 to 5, with 5 indicating a strong dependency.

Results and Discussion

As predicted, frequency judgments and similarity-to-ideal (mutability) judgments were negatively correlated. As shown in Table 10, features that were rated more mutable tended to be judged less frequent. A 3 (type of dependency relation) × 2 (order of tasks) analysis of variance showed a main effect of dependency relation for both similarity-to-ideal, $F(2,54) = 4.19$; $MSe = 3.63$; $p < 0.05$, and for frequency judgments, $F(2,54) = 7.93$; $MSe = 167.05$; $p < 0.001$. Specifically, the feature depended on was judged more immutable than the feature with no relations, $t(28) = 2.83$; $SE = 0.51$; $p<.01$, and was also judged more frequent than the no relations feature, $t(28) = 3.46$; $SE = 3.87$; $p<.01$, even though it was actually less frequent. The feature depended on was not judged significantly more immutable than the feature depending on it, $t(28) = 1.42$; $SE = .53$; n.s., nor significantly more frequent, $t(28) = 1.60$; $SE = 3.15$; n.s., although differences were in the predicted direction. Participants may have failed to always preserve the direction of dependency. In general, frequency judgments appear biased by the same dependency relations that determine mutability judgments.

The effects of task order were not significant, $F < 1$ for both similarity-to-ideal and frequency. However, the interaction between frequency judgments and order was, $F(2,54) = 3.92$, $p < .05$. The effect of mutability on frequency judgments was stronger when subjects first rated mutability than when the task order was reversed, although the order of means was identical. One explanation for this is that participants' memories for the profiles

TABLE 10
Mean Mutability and Name Centrality Ratings from Study 4 for Specific and Intermediate-Level Categories and Features

	Relational Role of Feature		
Judgement Task	No Relation	Depending-On	Depended-On
Similarity-to-Ideal (mutability)	6.0	5.3	4.5
Frequency	43.9	52.3	57.3

Note. Similarity-to-ideal judgments range from 0 (most immutable) to 10 (most mutable). Frequency judgments range from 0 to 100.

degraded while they made mutability ratings when frequency judgments were second, so they were forced to rely more on dependency information when making frequency judgments than they did when frequency was judged first. Perhaps participants were able to keep track of frequencies in the short term, but eventually came to rely on mutability. Another possibility is that rating mutability first made participants more aware of the features' internal relations to one another, which increased the likelihood of using dependencies to estimate frequency. No such interaction obtained for the mutability ratings, $F < 1$; mutability was not influenced by prior frequency judgments.

As a check of the dependency manipulation, pairwise dependency ratings were collected. As expected, the dependency of the depending feature on the depended-on feature was stronger than any other dependency ($p < .05$, Bonferroni adjusted t-tests).

This experiment demonstrates experimentally that the mutability of a feature can be decreased by causing other features to depend on it and that frequency—and therefore variability—judgments can be biased by those dependencies. Variability represents how the instances in a category differ. A concept's dependency structure is not directly relevant to such judgments, although it does provide a clue about how instances might be expected to differ. Our demonstration that dependency structure affects variability judgments implies that people attend to dependency structure, even when it is not relevant. They also use it when making mutability judgments, where it is relevant.

GENERAL DISCUSSION

Summary

The current results support two hypotheses. First, mutability captures a systematic aspect of conceptual structure. This follows from the convergence of our mutability measures, suggesting that features can be reliably ordered according to their transformability, and the divergence of mutability from other measures. Mutability did not correlate with measures of diagnosticity or salience (Study 1) and it dissociated from a measure of naming centrality (Study 4). Our wager is that our four measures of mutability (surprise, ease-of-imagining, goodness-of-example, and similarity-to-an-ideal) will prove hard to distinguish from each other empirically.

Second, the relative success of our dependency model lends support to the hypothesis that features are mutable to the extent that other features do not depend upon them. The model fits data quantitatively reasonably well without any free parameters, better than comparable models, and almost as well as a more sophisticated model with a degree of freedom. We tested a qualitative property of the model, namely, its assumption that mutability judgments are insensitive to the type of relation between features. We varied whether symptoms were related by causal, temporal, or generic dependency relations. No effect of type of relation was observed.

The model assumes that mutability judgments are derived from thinking about the internal structure of a concept, and not by enumerating instances of its corresponding category. Evidence for this assumption came in the form of a dissociation between counterfactual naming and mutability in Study 4. Study 5 showed that dependency structure influences not only judgments of mutability, but judgments of variability too. So dependency structure influences judgments of conceptual structure (mutability) and category structure (variability), producing a bias in the latter case.

Implications

Above, we argued that immutability is a primary determinant of feature weighting in object categorization. A feature of an object is central to its category type to the extent that the object's functions and appearance depend on that feature. The analysis of a variety of other conceptual tasks could benefit from an understanding of mutability:

Metaphor

Mutability facilitates the interpretation of nominative metaphors. The metaphor *my desk is a junkyard* is understood as referring to the tidiness of my desk and not (say) how flat its surface is. Metaphorical statements such as this one map aligned characteristics of the source onto the target (Gentner & Wolff, in press; Ortony, 1993) and mutability constrains the choice of target features that can be successfully mapped onto. Only mutable features can be mapped onto because immutable features resist transformation. Tidiness is a mutable property of desks because it is not central. Other features of desks, like having drawers, legs, and a top do not depend on a desk's tidiness. Therefore, the metaphor transforms the estimate of my desk's tidiness, and not some other feature, from those of the typical desk to those of a typical junkyard.

Conceptual Combination

By a similar reasoning, mutability plays a role in determining how people interpret novel conceptual combinations like *giraffe car*. A pragmatically viable interpretation can be resolved while making only minimal changes to the concept of the head (car) by mapping a property of the modifier (giraffe) onto a mutable feature of the head. Thus, a reasonable interpretation is a particularly tall car because tallness is mutable of cars. A more immutable property of cars is that they run on fuel. Hence, people would not ordinarily resolve the combination to mean a car that eats leaves. Of course, mutability is only one of the vari-

ables that can help to predict feature relevance in conceptual combination. Another critical constraint is that the dependency structure of the head must be able to support the mapped feature. The noun-noun combination *frog car* is more likely to be interpreted as a green car than as a car that hops because the feature *can hop* depends on other features that cars do not have (Love, 1996). Hence, knowing the mutability of head features is not enough to predict feature mapping, one must also consider how well the specific dependency structures of the modifier and head fit with respect to the mapped property.

Inductive Inference

All else being equal, an immutable feature is more likely than a mutable one to be projected from one object to another. For example, suppose you have just learned that your computer's central processor depends for its operation on built-in memory registers. We suspect that this fact would substantially increase your belief that the central processor of the next computer you come across will also require memory registers. We also suspect that if you have just learned that your computer has a built-in cache, you will be less certain whether to generalize that property to the next computer you see. You already know that a large number of a computer's properties depend on its central processor, so if the central processor depends on memory registers, memory registers must be central and therefore immutable features of your computer. A natural inference is that they are also immutable of other computers. But the news that a computer has a feature, like a cache, without any indication that anything depends on it, seems to provide less sanction for the inductive projection of that feature.

Explanation Generation

Mutability also helps to determine feature relevance when people are constructing explanations. Explanations tend to focus on mutable features (Kahneman & Miller, 1986). For example, appealing to an immutable feature (like *to move*) to explain why some birds have webbed feet is less satisfying than appealing to a mutable property (like *to swim*). Because it is more mutable, *can swim* has a greater chance of not holding, which makes it useful for distinguishing those birds that have webbed feet from those that do not. Of course, a good explanation depends on attributes unrelated to mutability, like the most significant difference between an object and its contextually relevant contrast set. Our claim is only that explainers will be biased to appeal to mutable differences.

Event Understanding

Kahneman and Miller (1986) and Kahneman and Varey (1990) point out that knowledge about mutability is at the heart of our understanding of events. Responses to outcomes are influenced by other outcomes that could have occurred. The reaction to the outcome of a game or contest, for instance, depends on how close the loser came to winning. Whether or not we notice a person's behavior depends directly on our expectations of that person and of people in general. Events are not perceived in isolation but within a background of counterfactuals—events that might have, but did not, occur. This background can determine

how surprised we are by an outcome (e.g., a court decision is not surprising if we cannot imagine it having turned out differently), how we assign blame (e.g., a person is guilty if they should have behaved differently), and when we experience regret (e.g., we regret not having behaved in some other fashion).

The analysis of event mutability could be enhanced by consideration of dependency structure. The temporality effect (Miller & Gunesagaram, 1990; Byrne, Culhane, & Tasso, 1995) is the finding that people consider the second of two independent events in sequence to be more mutable. For example, we have asked people to consider the following problem:

> Imagine two individuals (Jones and Cooper) who are offered the following very attractive proposition. Each individual is given one yellow ball and one green ball, and asked to choose one of them. If the two individuals pick balls with the same color, then each individual wins $1000. However, if they pick balls of different colors, neither individual wins anything. During this game, they are not allowed to see each other's choices. Jones goes first. Jones picks a yellow ball because Jones prefers yellow. Cooper goes next and picks a green ball because Cooper prefers green. Thus, the outcome is that neither individual wins anything.
>
> What could have been different that would have allowed Jones and Cooper to win $1000? Please circle one of the following two options:
>
> Jones could have picked a green ball. Cooper could have picked a yellow ball.

Replicating the standard result, we find that most people choose Cooper, the second actor in the problem, over Jones, the first. Byrne et al. (1995) have argued that the second of two sequential but otherwise independent events is more mutable because the first serves as an initial anchor in the construction of a mental model of the sequence. Whether or not this is the case, we find it less than adequate as an explanation of the effect, because it fails to tell us why the first and not the second event serves as a conceptual anchor. We suggest that the difference is in the encoded dependency structure. The relevance of the second event depends on the first in a way in which the relevance of the first does not depend on the second. The initial interpretation of the second event is governed by the outcome of the first event, but the initial interpretation of the first event cannot depend on the second because the second is, initially, unknown. Because more depends on the first event than the second, the second is more mutable.

The current analysis of object representation shares with the analysis of counterfactual events the assumption that the world is interpreted not just in terms of how it is, but also in terms of how it could be. People perceive objects not as static entities but in terms of their dispositions to change. This is clearly true of objects which are intrinsically dynamic, like clouds, tumors, trees, and hurricanes (cf. Leyton, 1988). Consequently, the analyses of objects and events share a concern with the determinants of mutability. However, they also display a critical difference. Kahneman and Varey (1990) champion a distinction between dispositions, causes which serve as the background for an event, and propensities, causes which determine developments during the event itself. The bases of counterfactuals are propensities; a counterfactual outcome is one that was almost produced by the causal sequence intrinsic to an event. However, the concern of this paper is the dispositions of objects, the kinds of mutations that background knowledge supports, not with the propen-

sities that are introduced by the specific context in which an object is encountered. Indeed, our empirical work tries to minimize such contextual considerations. Ultimately, a theory of object concepts will require a representation of objects in their ecological contexts, but our current theoretical goals are more limited.

CONCLUSIONS

Dependency structures can be hierarchically arranged in the sense that parts and aspects of concepts can have their own dependency structures. For instance, in the apple graph of Figure 1, two subnetworks of features can be discerned, one concerning the reproductive aspects of apples and the other containing the food-related features of apples. Each of these aspects of apples enforces its own coherence constraints which contribute to the overall coherence of the apple concept (cf. Rumelhart, Smolensky, McClelland, & Hinton, 1986).

As proponents of the view that conceptual coherence emerges from the interaction of multiple local (pairwise) dependency relations, we have no obvious way of distinguishing those relations that are intrinsic to a concept from those relations that are known but are not part of a concept's internal structure. For example, the relation between a ram's horn and the acoustical properties of a musical instrument do not seem intrinsic to the concept of ram and would not participate, we expect, in the determination of the mutability of the features of the concept corresponding to ram. This form of the binding problem is well-known and not easily solved by any learning device. Our belief, along with Thagard (1989), is that the dynamical properties of the kind of constraint satisfaction systems discussed by Hopfield (1982) and reviewed by Rumelhart and McClelland (1986) afford a useful perspective on this issue. If a concept is best conceived as an attractor in a high dimensional state space—as the solutions to Equation (1) are—then the set of relevant relations for a concept are jointly determined by the constraints imposed by all the relations in parallel.

The success of our abstract network representation provides more evidence for the utility of conceiving of concepts as attractors in a large-dimensional state space. One advantage of such a conception is that it illustrates how concepts can be both flexible and have structure. A concept emerges as multiple constraints are simultaneously satisfied. We have posited that the key constraints are pairwise dependencies between features. Concepts are flexible because little depends on some features, so they are easily transformed. Concepts, nevertheless, have structure because much depends on other features; they are relatively immutable.

Acknowledgments: This work was supported by a Richard B. Salomon Faculty research Award and by a grant from Brown University to Steven Sloman. Some of the results from Study 2 were reported at the Seventeenth Annual Conference of the Cognitive Science Society and the results from Study 4 at the Nineteenth Annual Conference. We would like to thank Nick Haslam, Doug Medin, Dorrit Billman, and two anonymous reveiwers for their comments on prior drafts.

APPENDIX A

Pearson Correlations Between 10 Category-Feature Rating Tasks for each of 4 Categories

TASKS	sprz	e-of-i	gd-ex	sim	name	cat valid	cue valid	infer poten	prom
CHAIR									
surprise	1								
ease of imag	−0.94	1							
goodness of ex	−0.96	0.98	1						
sim to ideal	−0.97	0.92	0.96	1					
naming	−0.94	0.97	0.98	0.93	1				
categ. validity	−0.92	0.93	0.96	0.93	0.96	1			
cue validity	0.59	−0.74	−0.71	−0.55	−0.8	−0.65	1		
infer potency	0.63	−0.75	−0.75	−0.6	−0.82	−0.67	0.97	1	
prominence	0.76	−0.66	−0.75	−0.7	−0.72	−0.80	0.39	0.49	1
availability	0.75	−0.71	−0.65	−0.68	−0.64	−0.52	0.42	0.40	0.38
GUITAR									
surprise	1								
ease of imag	−0.71	1							
goodness of ex	−0.64	0.77	1						
sim to ideal	−0.44	0.4	0.64	1					
naming	−0.78	0.82	0.87	0.72	1				
categ. validity	−0.19	0.23	0.41	0.76	0.41	1			
cue validity	−0.15	−0.05	0.21	0.52	0.25	0.44	1		
infer potency	−0.13	−0.12	0.057	0.49	0.14	0.36	0.97	1	
prominence	−0.046	−0.25	−0.61	−0.33	−0.33	−0.22	−0.5	−0.36	1
availability	−0.24	−0.18	−0.084	0.43	0.016	0.44	0.55	0.65	0.023
APPLE									
surprise	1								
ease of imag	−0.95	1							
goodness of ex	−0.95	0.83	1						
sim to ideal	−0.91	0.78	0.95	1					
naming	−0.99	0.94	0.96	0.94	1				
categ. validity	−0.82	0.74	0.74	0.86	0.82	1			
cue validity	0.24	−0.37	−0.043	−0.18	−0.26	−0.49	1		
infer potency	0.5	−0.63	−0.26	−0.32	−0.46	−0.65	0.77	1	
prominence	0.025	0.097	−0.21	−0.053	−0.061	0.27	−0.57	−0.69	1
availability	−0.74	0.71	0.68	0.60	0.71	0.36	0.012	−0.17	0.00
ROBIN									
surprise	1								
ease of imag	−0.76	1							
goodness of ex	−0.79	0.76	1						
sim to ideal	−0.70	0.79	0.77	1					
naming-0.73	0.64	0.82	0.57	1					
categ. validity	−0.72	0.53	0.65	0.69	0.44	1			
cue validity	−0.091	0.33	−0.032	0.15	−0.26	0.12	1		
infer potency	0.096	0.34	−0.21	0.026	−0.25	−0.092	0.56	1	
prominence	0.21	−0.42	−0.54	−0.35	−0.63	−0.11	0.25	0.28	1
availability	−0.039	0.40	0.22	−0.084	0.17	−0.12	0.2	0.54	−0.32

APPENDIX B

Four alternatives to Equation (1) were examined. Rather than reporting correlations between the models and every measure, we focus on the dimension of primary interest, mutability, and will use a single measure as our proxy for mutability. The factor analysis reported in Table 4 illustrates that only one measure seems to be a "pure" measure of mutability in the sense that it loads highly on Factor 1, the mutability factor, but has near-0 loadings on the other factors, namely ease-of-imagining. Therefore, tests below report correlations for ease-of-imagining only.

The first alternative model represents a feature's immutability as the simple sum of other features' dependencies on it:

$$c_i = \Sigma_j d_{ij}.$$

Correlations between this model and ease-of-imagining judgments were −.86, −.43, −.28, and −.75, for the categories *chair, guitar, apple* , and *robin,* respectively. Three of these correlations are lower than the corresponding correlations for Equation (1), −0.92, −0.62, −0.60, and −0.59, although the differences are not statistically significant. The relative superiority of Equation (1) over the sum of dependencies model confirms the need for the iterative aspect of Equation (1).

To construct the second alternative model, we added a nonlinearity to Equation (1) to optimize the fit to mutability judgments. In this model, the result of each iteration was normalized using a linear transformation which placed the lowest centrality value at 0 and the highest at 1, then raised each value to a power in the range 0 to 1, a power chosen to maximize the resulting correlation. This model gave only slightly better predictions of immutability than Equation (1), −.92, −.72, −.60, and −.74 for the four categories, respectively, despite its greater complexity and its free parameter.

We also tried a model which considered not only the extent to which other immutable features depended on the focal feature, but also the extent to which the focal feature depended on other mutable features, by adding a new term $\Sigma_k d_{ki} (1 - c_{k,t})$. Logically, a feature's dependence on other mutable features should increase its mutability, so this term should be subtracted to predict immutability:

$$c_{i,t+1} = \alpha\Sigma_j d_{ij} c_{j,t} - (1 - \alpha)\Sigma_k d_{ki} (1 - c_{k,t})$$

where α is a number between 0 and 1. If α is close to 0, then the centrality of feature i is proportional to what feature i depends on; if α is close to 1, then the model approaches Equation (1), feature *i*'s centrality is proportional to what depends on it. We varied α from 0 to 1 in increments of .01. The correlations with ease-of-imagining were highest for all categories when α was close to 1 (in all cases, greater than 0.8), suggesting that the second term adds little of psychological relevance to the model.

We also evaluated a model which considered the total connectivity of each feature; one that sums over the two directions of dependency:

$$c_{i,t+1} = \Sigma_j d_{ij} c_{j,t} + \Sigma_k d_{ki} c_{k,t}$$

This model was again not as good as Equation (1). Its rank correlations with ease-of-imagining were −.88, −.23, −.59, −.15, all lower than the corresponding Equation (1) magnitudes.

NOTES

1. The category validity question might admit of an alternative interpretation. We intended subjects to read it as asking them to estimate the frequency of the feature at some point during the lifetime of objects in the category. But those who weren't careful could have read it as asking them to estimate the frequency of the feature for all objects in the category at a specified point in time ("what percentage of apples now in existence are growing on trees?"). The data suggest that this second reading was rare: The measure turned out to be highly correlated with other centrality measures.
2. Correlations involving surprise are negative because the surprisingness of an instance without a feature should decrease with the mutability of the feature whereas the other mutability measures should increase.
3. Availability was the least reliable measure. Its split-half correlation was only 0.76; the split-half correlations for other measures were all greater than 0.80. The pattern of correlations in Table 4 remains unchanged by a correction for reliability. The effect of such a correction is to disproportionately increase the largest correlations. Hence, the results are not due to range attenuation or any other source of unreliability.
4. A detailed analysis of the distinction between associative and rule-based processing can be found in Sloman (1996). Evidence that associative (similarity-based) categorization is faster than rule-based categorization can be found in Allen and Brooks (1991) and Smith and Kemler (1984). Evidence that it is less deliberative can be found in Smith and Sloman (1994).
5. Our method of measuring mutability spawned an unexpected factor limiting the performance of the model. Extremely immutable features, like "is living" for robin, are so immutable that they tended to cause participants to consider a different category. Participants seemed unable to imagine a real robin that lays eggs, eats worms, and flies but is not living and therefore instead imagined a toy or decomposing robin. In the context of this different category, the feature was no longer judged immutable although its dependency relations predict that it should have been. Ease-of-imagining judgments for such features had bimodal distributions, suggesting that some participants experienced difficulty performing the transformation and that others did not perform the task asked of them. Despite our efforts to eliminate such judgments from analysis (see methods), we were not always able to because participants were not always aware of their error. Three features led to this problem: one each from the categories guitar, apple, and robin. The rank correlations between ease-of-imagining and the basic model improve if we eliminate these features from analysis to −.69, −.66, and −.74 for the three categories, respectively.
6. A fuller exposition of this argument appears in Ahn and Sloman (1997).

REFERENCES

Ahn, W. K. & Lassaline, M. (in preparation). Causal structure in categorization: A test of the causal-status hypothesis (Part 1).

Ahn, W. K. & Sloman, S. A. (1997). Distinguishing name centrality from conceptual centrality. Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, Stanford, CA, 1–6.

Allen, S. W. & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Billman, D., & Heit, E. (1988). Observational learning from internal feedback: A simulation of an adaptive learning method. *Cognitive Science, 12,* 587–625.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 458–475.

Boolos, G. & Jeffrey, R. (1980). *Computability and logic,* 2nd ed. New York: Cambridge University Press.

Byrne, R. M. J., Culhane, R., & Tasso, A. (1995). The temporality effect in thinking about what might have been. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 385–390.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: Plenum.

Clement, C. A. & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15,* 89–132.

Estes, W. K. (1993). Models of categorization and category learning. In G. V. Nakamura, D. L. Medin, & R. Taraban (Eds.), *Categorization by humans and machines. The psychology of learning and motivation: Advances in research and theory,* Vol. 29. (pp. 15–56): Academic Press: San Diego.

Franks, B. (1995). Sense generation: A "quasi-classical" approach to concepts and concept combination. *Cognitive Science, 19,* 441–505.

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

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

Gentner, D. & Wolff, P. (in press). Metaphor and knowledge change. In A Kasher & Y. Shen (Eds.), *Cognitive aspects of metaphor: Structure, comprehension, and use.*

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge: Harvard University Press.

Harman, H. H. (1967). *Modern factor analysis,* 2nd ed. Chicago: University of Chicago Press.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, USA, 81, 6871–6875.

Kahneman, D. & Miller, D. T. (1986). Norm theory: comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kahneman, D. & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology, 59,* 1101–1110.

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

Leyton, M. (1988). A process-grammar for shape. *Artificial Intelligence, 34,* 213–247.

Love, B. C. (1996). Mutability, conceptual transformation, and context. Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, San Diego, CA. Hillsdale: Erlbaum.

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

Malt, B. C. & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23,* 250–269.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (1986) (Eds.). *Parallel Distributed Processing,* Vol. 2. Cambridge: MIT Press.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). New York: Cambridge University Press.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20,* 158–190.

Miller, D. T. & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology, 59,* 1111–1118.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183–203.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts, and Categories.* Hove: Psychology Press.

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

Ortony, A. (1993). The role of similarity in similes and metaphors. In A. Ortony (Ed.) *Metaphor and thought,* 2nd ed. New York: Cambridge University Press.

Osherson, D., Smith, E. E., Shafir, E., Gualtierotti, A. & Biolsi, K. (1995). A source of Bayesian priors. *Cognitive Science, 19,* 377–405.

Putnam, H. (1975). The meaning of 'meaning.' In K. Gunderson (Ed.) *Language, mind and knowledge, Minnesota studies in the philosophy of science, VII,* Minneapolis: University of Minnesota Press.

Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review, 60,* 20–43.

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976a). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976b). Basic objects in natural categories. Working Paper #43. The Language Behavior Research Laboratory, UC Berkeley.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986) (Eds.). *Parallel Distributed Processing,* Vol. 1. Cambridge: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, and the PDP Research Group (Eds.). *Parallel distributed processing.* Cambridge: MIT Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280.

Smith, E. E. & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22,* 377–386.

Smith, J. D., & Kemler, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General, 113,* 137–159.

Spalding, T. L. & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1251–1263.

Spence, K. W. (1936). The nature of discrimination learning in animals. *Psychological Review, 43,* 427–449.

Thagard, P. R. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–502.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

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