Mutability and the Determinants of Conceptual Transformability

Bradley C. Love
Department of Cognitive and Linguistic Sciences
Brown University
Providence, RI 02912-1978
love@cog.brown.edu

Steven A. Sloman
Department of Cognitive and Linguistic Sciences
Brown University
Providence, RI 02912-1978
sloman@cog.brown.edu

Abstract

Features differ in their mutability. For example, a robin could still be a robin even if it lacked a red breast; but it would probably not count as one if it lacked bones. One hypothesis to explain this differential transformability is that having bones is more critical to a biological theory than having a red breast is. We reject this hypothesis in favor of a theory of mutability based solely on local dependency links and expressed in the form of an iterative equation. We hypothesize that features are immutable to the extent other features depend on them and offer supporting data.

1. Introduction and background

The study of conceptual use and conceptual transformation has taken two distinct directions. On one hand, some theorists assert that human conceptualization is theory-based, in the sense that concepts cohere by virtue of explanatory relations that hold between concepts and their components (e.g., Carey, 1985; Keil, 1989; Murphy & Medin, 1985; Wellman, 1990). On the other hand, some theorists take what Rips (1990) has termed the Loose view of concepts. These theorists explain performance on categorization, reasoning, and other conceptual tasks using statistical, similarity-based, or associative models of cognitive processing (e.g., Holyoak & Thagard, 1989; Sloman, in press; Tversky, 1977).

On the theory-based view, relations between concepts and their components come in qualitatively different varieties. For instance, the theory-based view assumes multiple forms of dependency relations between the components of concepts. For the concept robin, the dependency between the feature "can fly" and the feature "has wings" is causal. However, for the concept guitar, the feature "makes music" is not causally related to "makes sound", but is rather a specialization of it (cf. Collins & Michalski, 1989). In sum, on the theory-based view, relations are labeled by their semantic role.

In contrast, on the Loose view, the relations binding concepts may vary in their magnitudes but they are all of the same semantic type. On this view, only one type of relation is necessary to bind concepts and the components of a concept. For example, both causal and specialization relations would be classified simply as dependency relations. Our aim is to provide support for this hypothesis. We believe that much of human conceptualization can be explained without appealing to labeled relations. We focus on tasks that involve conceptual transformations of everyday concepts and offer evidence that the ease of transforming a feature can be measured on a unidimensional scale of mutability, a scale that we believe is central to explaining performance on a variety of cognitive tasks. We also test the hypothesis that mutability is determined by a uniform type of dependency relation between the features of a concept. More specifically, we hypothesize that a feature is immutable to the extent that other (immutable) features of the concept depend upon it.

2. The scale of mutability

For any category, we have a notion of what members of that category should be like. For instance, when one thinks about robins, one envisions a creature that eats, builds nests, flies, has wings, a red breast, feathers, and so on. Nevertheless, one can successfully perform conceptual transformations in which one can imagine a robin that does not build nests but is still a robin. Consider the two statements below, each of which describes a robin that is atypical:

(A) The robin does not have a red breast, but is otherwise normal.
(B) The robin does not ever eat, but is otherwise normal.

Clearly, you are less likely to encounter the robin described by Statement (B) than the one described by Statement (A) because (B) describes a more difficult conceptual transformation. Something that does not have a red breast is more easily imagined to be a robin than something that does not eat because eating is more central in our representation of "robinhood" than is a red breast. Features that are central to a representation, like "eats", will be referred to as immutable, while those that are more

easily transformed, like "has a red breast", will be referred to as mutable.

3. Determinants of mutability

3.1 Variability

One possible source of mutability judgments is the perceived variability of features across category members. Features that are almost always present will have low variability and thus be immutable in the sense that exceptions are rare. Features present in about half the members of a category are highly variable and necessarily mutable. Variability however does not provide a sufficient explanation for mutability because the psychological determinants of variability itself and the sets it is measured across are not well-defined. Variability does not even have meaning for cases in which one has only a single experience with the category token. Moreover, we know perceptions of variability are not the only source of mutability judgments because differences in mutability exist even when variability is held constant. Variability and mutability can even oppose one another. Consider the feature "is curved" for the categories banana and boomerang. In banana, the feature has low variability (all bananas are curved), but is mutable (we can easily imagine a banana that is not curved). In boomerang, the feature "is curved" happens to be variable, but nevertheless seems immutable (Medin & Shoben, 1988). This reversal can be accounted for in terms of the dependency relations between the features of each category and the feature "is curved". No other features depend upon "is curved" in bananas, while other features do depend upon it in boomerang.

In sum, mutability and variability are related inasmuch as both types of judgments are sensitive, directy or indirectly, to the extent to which a feature actually does vary across instances. We predict therefore that the two judgments will be correlated. However, the judgments are not the same; mutability is a property of conceptual structure and variability is an extensional property of frequency distributions. We therefore expect the two judgments to sometimes diverge.

3.2 Dependency

Our centrality hypothesis states that those features that have many other features depending upon them will be immutable, while those features that do not have other features depending upon them will be mutable. Transforming a representation by varying an immutable feature will be difficult because it will be disruptive. Other features that depend upon the immutable feature will also change and this can have ramifications for the entire representation. Performing conceptual transformations across mutable features is relatively easy because little hinges on these features; they are relatively peripheral.

We express this hypothesis using the following iterative equation:

$$C_{i,t+1} = \sum A_{ij} C_{j,t} \qquad (1)$$

where $C_{i,t}$ is the immutability of feature i at time t and A_{ij} is the dependency link from feature j to feature i (the dependence of feature j upon feature i). According to the equation, the immutability of feature i is determined at each time step by summing across the immutability of every other feature multiplied by that feature's degree of dependence upon feature i. In other words, if a highly immutable feature depends upon feature i, feature i becomes more immutable than if a mutable feature were instead to depend upon it. A feature cannot become central to a representation merely because a peripheral feature depends upon it. The feature would be much more central if a feature depended upon it that many other features, in turn, depended upon. If feature X depends upon feature Y, and feature Y depends upon feature Z, then feature X also depends upon feature Z. All other things being equal, feature Z would be less immutable if feature X did not depend upon feature Y. These non-local effects are accommodated by the iterative nature of Equation (1).

To implement the model, immutability ratings must be set to some initial arbitrary value. The model iterates until it converges. Mathematically, the model is a repetitive matrix multiplication and is known to converge to a stable solution in a small number of steps (Wilkinson, 1965). The solution is 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.

Equation (1) describes our attempt to reduce mutability to pairwise, unlabeled dependency relations. These relations can be conceived of as associative strengths. Although we cannot justify assigning them a probabilistic interpretation, the value of A_{ij} may turn out to be a nondecreasing function of Pr{feature j | feature i} - Pr{feature j}. We do not present a model of the origin of the dependencies. We assume that they have multiple sources, including the detection of feature covariations and causal explanations of category structure.

4. Testing the model

A series of three studies was performed to explore the relations between mutability, dependency, and variability. We test two predictions: i. Mutability judgments can be fit by Equation (1) using empirically obtained dependency judgments; ii. Mutability judgments are correlated with

judgments of variability. Each study involved questionnaires which were filled out by 20 Brown University undergraduates who were paid for their participation.

4.1 Study 1: Assessing the mutability of the features of a category

In this study, mutability ratings were collected for the features of the categories Guitar, Apple, Chair, and Robin. The features used for these categories were taken from Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976). Rosch et. al. used a three-step procedure to collect features. First, subjects were given 90 seconds to list features for a category. Second, responses were tallied and features listed by less than one third of the subjects were discarded. Next, seven judges deleted features that they believed were not true of all category members and added previously listed features that they believed were true of all category members. At the end of this process, the categories Chair, Guitar, and Apple each had 9 features, while the category Robin had 14 features. In Study 1, subjects gave mutability judgments for these features.

Before making their mutability judgments, subjects were told what the features of each category were, and were asked not to deviate drastically from this conception of the category. They were then asked to answer questions like, "How easily can you imagine a real apple that is not round?" Subjects responded with a number from 0 to 1 that reflected the ease of the transformation. At the end of each section, subjects were asked to list all items for which they had drastically changed their perception of the category (e.g. the subject considered a toy robin instead of a real one). These items (about 9 percent) were discarded.

4.2 Study 2: Measuring dependency relations

Subjects were shown, simultaneously, all the features of a particular category from the previous study. Each feature was inscribed in a circle and subjects were asked to draw arrows from each feature to each other feature they judged the feature dependent upon, 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".

4.3 Study 3: Assessing the variability of the features of a category

The features and categories from the previous studies were used. Subjects were asked questions such as "What percentage of robins have a red breast?" and they responded with a number between 0 and 100. Variability was calculated by transforming percentage estimates using the binomial variability measure X/100*(1 - X/100) for each feature.

4.4 Results and discussion

Figure 1 displays the mean dependency link values and mutability judgments (on a scale of 0 to 1) given by subjects. To maintain the readability of the graphs, only the strongest dependencies have been drawn (an average of 1.25 links per feature), although all dependency information was used in our simulations.

From these graphs, one can see that features with few other features depending upon them tend to be mutable while features with many features depending upon them tend to be immutable, as predicted. Table 1 presents Spearman rank correlations between mutability judgments and three dependency models across item means. The first model simply sums the incoming dependencies to compute immutability. This model's performance confirms our intuition that a feature's immtability varies with the number of other features that depend upon it. The second model is Equation (1) itself. Clearly, Equation (1) outperforms the incoming connections model, demonstrating that a feature's immutability is not only a function of a feature's incoming connections, but also a function of a feature's place in an overall dependency structure. This result confirms the need for the iterative aspect of Equation (1). In the third model, the modified model, a nonlinearity was added to Equation (1) to optimize the fit to mutability judgments. In this model, the result of each iteration was normalized to fall in the range 0 to 1, then raised to a power in the range 0 to 1, chosen to maximize the resulting correlation. This model gives slightly better predictions of immutability than Equation (1), but its advantage is modest relative to its greater complexity.

Table 1: Rank correlations of three models of dependency with mutability judgments for four categories, data from Studies 1 and 2.

Category	Sum of dependencies	Equation (1)	Optimized Equation (1)
Chair	-.86	-.92	-.92
Guitar	-.43	-.62	-.72
Apple	-.28	-.60	-.60
Robin	-.75	-.59	-.74

All of the correlations for the basic model (Equation (1)) and optimized model in Table 1 are statistically greater than 0 using a significance level of 0.05 except for three: those for the basic model of Guitar and both models of Apple. However, all the correlations are significant at the 0.10 level. We have also tried other variations on Equation (1), none of which consistently perform as well

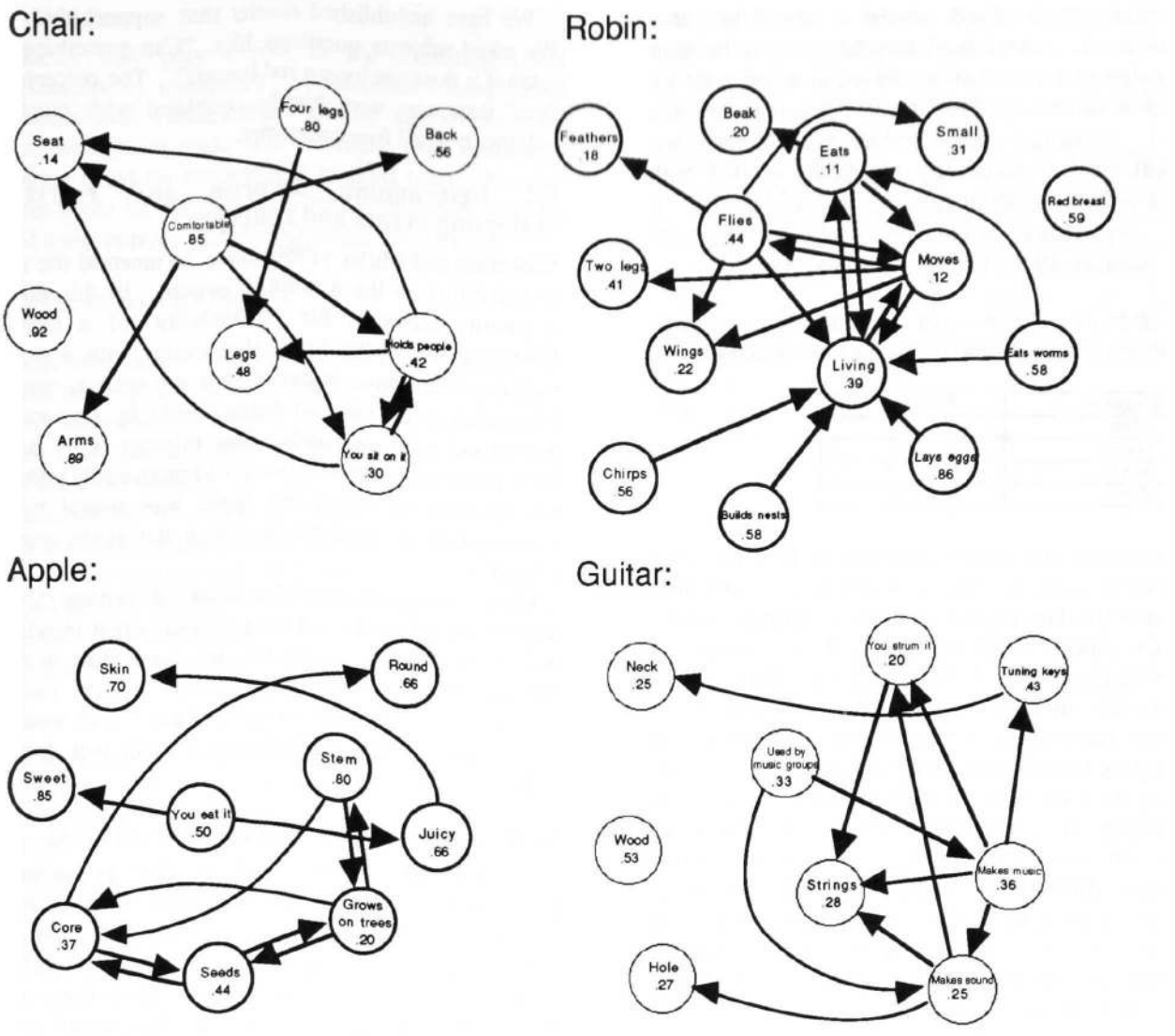

Figure 1: Category graphs. The arrows point from a feature to one that it depends upon. Mean mutability judgments are also shown for each category-feature.

as that model. Therefore, because of its combination of simplicity and empirical adequacy, we conclude that Equation (1) is the appropriate model of mutability and that we succeeded in predicting mutability judgments using unlabeled dependency relations.

Our method of measuring mutability spawned an unexpected factor limiting the performance of our model . Extremely immutable features, like "is living" for robin, are so immutable that they tend to cause subjects to consider a different category. Subjects are unable to imagine a real robin that lays eggs, eats worms, and flies but is not living and therefore instead imagine a toy or decomposing robin. In the context of the new category, the feature is no longer judged immutable although its dependency relations predict that it should be. Mutability judgments for such features had bimodal distributions, suggesting that some subjects experienced difficulty performing the transformation and that others did not perform the task we asked of them. Despite our efforts to eliminate such judgments from analysis (see section 4.1), we were not always able to because subjects were not always aware of their error. Three features led to this problem: one each from the categories Robin, Guitar, and Apple. The rank correlations between mutability judgments and Equation (1) improve if we eliminate these features from analysis to -.74, -.69, and -.66 for the categories Robin, Guitar, and Apple, respectively.

Table 2 presents Spearman rank correlations between mutability and variability judgments across item means. As expected, features judged variable also tended to be judged mutable ($p < .05$ in all four cases). Relatively high correlations should be expected for variability because, in those cases in which variability does vary

across features, it is closely related to mutability at a conceptual level. Indeed, their correlations may be high because judgments of mutability served as a surrogate for judgments of variability. They are also closely related at a task level. Variability judgments are made at the same ontological level as mutability judgments in that both consist of judgments about isolated category-features. In contrast, dependency judgments considered pairwise relations amongst all the features of a category.

Table 2: Rank correlations of mutability (Study 1) and variability (Study 3) judgments for four categories.

Category	Correlation
Chair	.98
Guitar	.69
Apple	.74
Robin	.53

In conclusion, the results demonstrate that unlabeled dependency relations are effective in predicting mutability. No aspect of the data suggests that the performance of the dependency model could be improved by considering labeled relations. The feature graphs of Figure 1 do display further structure. For instance, in the Apple graph, two subnetworks of features can be discerned, one concerning the reproductive aspects of apples and the other containing the food related features of apples. However, this structure is discernible without attributions of causality or any other label to dependency links. Furthermore, although this structure is undoubtedly useful for certain cognitive tasks (such as, probably, analogical reasoning), we have no reason to believe that it contributes to determining the transformability of a feature. Admittedly, our conclusion would be more compelling if we had directly contrasted our model's results with those obtained with a labeled relations model. Unfortunately, the theory-based view remains too ill-specified to provide such a model.

5. The role of mutability in other conceptual tasks

We believe that mutability can serve as an explanatory device in a variety of cognitive tasks.

5.1 Categorization

Mutability plays a role in determining the relative importance of features in judgments of category membership. A token that matches a category representation in all but a mutable dimension should be a better candidate for category membership than a token that differs in an immutable dimension (Medin & Shoben, 1988). For example, we expect robins without red breasts to be categorized as robins with higher probability than robins that do not eat.

We have unpublished results that support this view. We asked subjects questions like, "Can something be a robin if it does not have a red breast?" The percentage of "yes" responses were highly correlated with mutability judgments in all four categories.

5.2 Determining surprise and regret in evaluating events and concepts

Kahneman and Miller (1986) have documented the effects of mutability in the domain of events. In this domain, mutability refers to the "undoability" of a situation. Kahneman and Miller found that events with a negative outcome elicit more regret if they are seen as mutable. For instance, missing an airline flight by five minutes was judged more regrettable than missing it by half an hour, presumably because one could more easily transform the situation in which the flight was missed by five minutes into a situation in which the flight was not missed.

Mutability is also a useful indicator of surprise. Greater surprise should be elicited from subjects upon viewing an object varying in an immutable dimension than an object varying in a mutable dimension. For instance, encountering a robin that does not have wings would be more surprising than encountering a robin that does not chirp.

5.3 Explanation generation and evaluation

Mutability may be a factor in the generation of explanations. An appropriate explanation for what makes a good computer would not center upon highly immutable features like "is a three dimensional object" or "performs calculations", but would instead center upon features that are more mutable like "has a very fast clock speed" or "has a large cache".

Explanations that focus on immutable features will be unsatisfactory. Because oxygen is an immutable feature of the atmosphere, an explanation that a building burned down because there was oxygen in the atmosphere seems inadequate (Kahneman and Miller, 1986).

5.4. Problem solving

In reasoning tasks where a forward or backward inference must be made to determine how to move from one problem state to another, mutability may indicate the features of the problem space that are manipulable. For instance, if an autonomous agent must manipulate objects in its environment to achieve some goal state configuration of objects, a good strategy might be to first focus on solutions involving easily transformable objects (objects not affixed to the ground, light objects, objects without other objects on top of them, etc.).

5.5 Metaphor

Mutability may play a role in the construction and interpretation of metaphors. Sometimes, metaphorical statements map characteristics of the source onto the target. Mutability could help determine what features of the target could be successfully mapped onto. Mutable features could be mapped onto, while immutable features would resist reinterpretation. Consider this example.

(C) The surgeon is a butcher.

Our representation of surgeon contains the features "has medical training" and "cuts with great precision and care". Our representation of butcher contains neither of these features. Alignable features of butcher could be mapped onto the representation of surgeon. The feature "has medical training" is a fairly immutable feature of surgeon and resists conceptual transformation, while the feature "cut with great precision and care" is mutable and is mapped onto by butcher. The resulting representation of surgeon is one in which the surgeon has medical training, but is not highly skilled at operating.

6. Conclusion

More work needs to be done in analyzing the determinants of mutability, the nature of dependency relations, their organization at all category levels, and their role in cognitive processes. Our hope is to contribute to the understanding of how the interdependencies between the elements that compose our concepts govern how our concepts cohere and transform. At present, that understanding remains mutable.

Acknowledgement

This work was supported by a grant from Brown University to Steven Sloman.

References

Carey, S. (1985). *Conceptual Change in Childhood.* Cambridge: MIT Press.

Collins, A. & Michalski, R. (1989). The logic of plausible reasoning: a core theory. *Cognitive Science, 13*, 1-50.

Holyoak, K. J. & Thagard, P. R. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Kahneman, D. & Miller, D. T. (1986). Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review, 93*, 136-153.

Keil, F. C. (1989). *Semantic and conceptual development: An ontological perspective.* Cambridge, MA: Harvard University Press.

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

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

Rips, L. J. (1990). Reasoning. Annual Review of Psychology, 41, 321-353.

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

Sloman, S. A. (in press). The empirical case for two systems of reasoning. *Psychological Bulletin.*

Tversky, A. (1977). Features of Similarity. *Psychological Review, 84*, 327-352.

Wellman, H. M. (1990). *The child's theory of mind.* Cambridge: MIT Press.

Wilkinson, J. H. (1965). *The algebraic eigenvalue problem.* Oxford: Clarendon Press.