# Mutability, Conceptual Transformation, and Context

**Bradley C. Love**
Department of Psychology
Northwestern University
2029 Sheridan Road
Evanston, IL 60208
loveb@nwu.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. I have hypothesized (Love & Sloman, 1995) that features are immutable to the extent other features depend on them. We can view a feature's mutability as a measure of transformational difficulty. In deriving new concepts, we often transform existing concepts (e.g. we can go from thinking about a robin to thinking about a robin without a red breast). The difficulty of this transformation, as measured by reaction time, increases with the immutability of the feature transformed. Conceptual transformations are strongly affected by context, but in a principled manner, also explained by feature dependency structure. A detailed account of context's effect on mutability is given, as well as corroborating data. I conclude by addressing how mutability-dependency theory can be applied to the study of similarity, categorization, conceptual combination, and metaphor.

## 1 Introduction: the importance of relations

Cognitive scientists have begun to gain an appreciation that concepts (in the psychological sense) are more than independent sets of features. Any account of concept representation must address the relations that exist among features. Relations help explain why some features are more central to a representation, while others are easily transformable. For instance, relations among features explain why it is difficult to imagine a normal robin without a heart, while imagining a robin without a red breast is more plausible. Having a red breast is a mutable feature of robins, while having a heart is an immutable feature of robins.

There have been varying accounts of why some features are relatively immutable, while others are mutable. On the theory-based view (e.g., Carey, 1985; Keil, 1989), the importance of the heart can be explained by appealing to a biological theory of how a robin functions. Such a theory would deem the heart central to our notion of what it means to be a normal robin, based on the web of relations in which the heart is embedded (e.g. "the heart *pumps* blood", "blood *carries* oxygen", "the brain *needs* oxygen", etc). The relations among features are labeled by the type of relation they represent (e.g. *carries, pumps, needs*). On this view (Murphy & Medin, 1985; Wellman, 1990), the concept robin coheres by virtue of the explanatory relations that hold between its components (and perhaps those of other concepts).

In contrast, the feature "has a red breast" does not play as critical of a role in the overall explanatory coherence of the concept robin, making the feature more mutable. That is, it is easy to imagine a robin not having a red breast (perhaps the robin has a brown breast). Not having a red breast does not have serious ramifications for a theory of what it means to be a robin.

The story becomes more complex when we consider that the mutability of a feature can vary with context. For instance, in certain contexts, the feature "has a red breast" can become more immutable. If one is reminded or alerted to the mating purposes of having a red breast, the feature will become more immutable. Effectively, the context of mating highlights features with relations in common with the feature "has a red breast", making "has a red breast" more immutable. The effects of context on categorization and similarity ratings are well documented (Medin et. al., 1993). Context can facilitate the interpretation of noun-noun compounds, analogies, and nominative metaphors (Gerrig & Murphy, 1992; Gick & Holyoak, 1980; Gildea and Glucksberg, 1983).

## 2 The dependency stance: an implemented theory

The theory-based view can explain why certain features are more critical or immutable, but the explanation has an ad hoc flavor and seems overly complex. It is unclear how a theory-based model could be implemented that predicts which features of a concept are mutable and which are immutable. It is difficult to see how qualitative statements like "plays a critical role in the overall explanatory coherence of the concept " can be made formal and yield quantitative predictions. The problem becomes more acute when we allow context to vary.

Since relations among features are labeled by their type, it is not possible to employ a simple algorithm that calculates the importance of a feature, since different types of relations are not directly comparable. One could overcome this difficulty by employing a simpler representational scheme that still captured the basic intuitions of the theory-based view.

I propose (Love & Sloman, 1995) that all types of relations can be collapsed to one primitive type, namely the unidirectional relation of *depends*; for the purposes of calculating feature mutabilities. In such a scheme, the

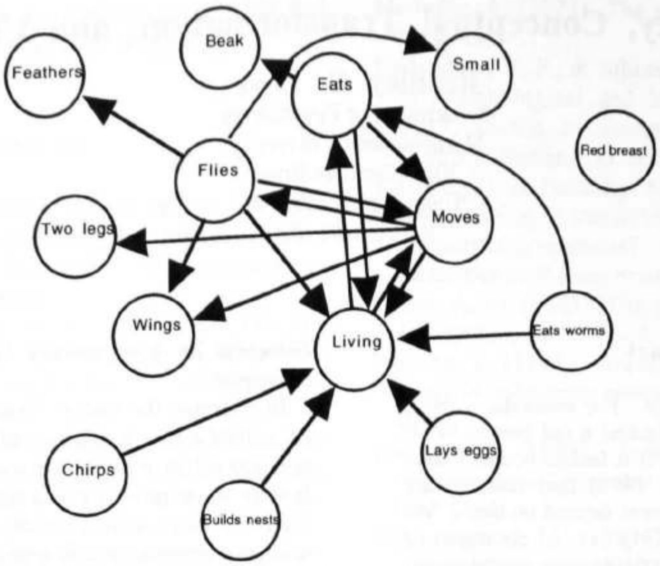


Figure 1: The arrows point from a feature to one that it depends upon, as rated by subjects (Love & Sloman, 1995).

relations *pumps*, *carries*, and *needs* would all be collapsed to the pairwise relation *depends*. Faced with the challenge of equating different types of relations, people may resort to using only dependency information in certain tasks. Figure 1 illustrates a dependency graph, in which relations among features are only represented as dependencies.

Having many features depending upon a given feature will make it more difficult to transform the given feature since the transformation will disrupt the representation of the concept. Other features that depend upon the immutable feature will also change and this can have ramifications for the entire representation. For example, if you were told that a particular robin did not have wings, you would need to update your default assumption that the robin can fly, since "can fly" depends on "has wings". Performing conceptual transformations across mutable features is relatively easy because other features are unaffected by changes in mutable features. We would expect reaction time to be slower for transformations performed across immutable features. In experiment 1, I test this prediction.

Having one type of relation makes it possible to compare all relations on the basis of magnitude. The mutability of a feature can be calculated by summing the number and strength of the other features that depend on it, which is a straightforward computation, yet accounts for subjects' mutability ratings (Love & Sloman, 1995). Obviously, there are tasks that require people to attend to the labels of relations, such as some reasoning tasks, but interestingly, such tasks require considerably more effort and processing time than tasks that do not demand labeled relations (Ratcilff & Mckoon, 1989).

By positing that people employ a dependency-like representation, an explanation of how context affects perceived mutability is suggested. When forming a concept, one draws upon a huge database of knowledge, only using a fraction of it in forming any particular concept (Barsalou, 1993). An individual can conceive of a category in a number of different ways, depending on context and current goals. Studying these effects is critical to our understanding of concept representation as context can dictate which features are included in forming a concept.

Since a feature is immutable to the extent that other features depend upon it, forming a concept from different sets of features (in different contexts) should affect feature mutability in a principled way. More precisely, if a feature is introduced (or highlighted) that depends upon a given feature, the given feature will become more immutable. Concretely, if I speak extensively of the mating practices of robins, and you know that certain aspects of the mating process depend upon the participants' color, then "has a red breast" should be more immutable in this context than in a context centered around flight.

## 3 Testing the dependency model

Two studies were conducted to test the following predictions: i. Features rated as mutable should be easier to transform. Subjects should be faster at imagining derivative concepts that vary in a mutable feature than in a immutable feature. ; ii. This transformation is affected by context in a principled way explained by the dependency structure of the representation.

### 3.1 Experiment 1: Mutability as Transformation

If subjects are performing a transformation of the concept robin to a derivative representation of robin when providing ratings for statements like, "How easily can you imagine a robin that does not have wings?", then one would expect that the ratings for such questions would correspond to the actual difficulty of the transformation. Furthermore, the difficult of a transformation should be measurable through reaction time. Transformations of highly immutable features should take longer than transformations over mutable features. In experiment 1, I tested this prediction.

#### Method

*Subjects.* Subjects in the feature mutability rating task were 20 undergraduates from Brown University. They were paid for their participation. Subjects in the reaction time task were 20 undergraduates from Northwestern University. They received course credit in an introductory psychology course for their participation.

*Materials and procedure.* The stimuli consisted of features from 4 categories (pine tree, robin, cucumber, and apple) taken from Dean and Sloman (1995). Mutability ratings were collected by having subjects answer questions like, "How easily can you imagine a robin without wings?" Subjects responded with a number between 0 and 1 that reflected the ease of the transformation. The number of features per category varied from 17 to 25. The 3 most mutable and immutable features from each category were chosen for the reaction time task, for a total of 24 features. Subjects were shown the name of the category and the feature on a Macintosh computer. They pressed the spacebar when they could imagine a member of the category not having the listed feature, but being normal in every other respect.

To ensure that any difference in reaction time between mutable and immutable features could not be attributed to the goodness, accessibility, salience, or reading times of the features; a feature confirmation task was included as a control. The same stimuli were used with the addition of 24 distractors. Subjects were instructed to press "p" if the category had the given feature, and to press "q" if the category did not possess the feature. Since all the features of interest clearly belonged to their category, 32 distractor features that did not belong to the presented category were included to ensure that subjects would not be biased towards an affirmative response.

#### Results

All observations more than 3 standard deviations above the mean were discarded (the cutoffs were 14576 msecs for the imagining task, and 3341 msecs for the feature confirmation task). For analysis, reaction times were separated into two groups: mutable and immutable. Subjects took longer to imagine instances of a category varying in an immutable feature ($t(539)=4.11$, $p<.001$) with a mean of 5153 msecs for the immutable features and a mean of 4355 msecs for the mutable features. The difference was significant.

There was no statistically significant effect in the control task. Feature's were confirmed with a mean response time of 1373 msecs for the immutable features, compared to 1355 msecs for the mutable features ($t(459)=.39$, $p>.70$).

#### Discussion

As predicted, the time to perform a conceptual transformation varied with the immutability of the feature transformed. The effect cannot be accounted for by any combination of frequency, accessibility, salience, reading time, etc., since any such effect would be manifest in the feature confirmation control task.

The results suggest that mutability ratings indicate how easy it is to perform a conceptual transformation across a feature's dimension. Such a position is consistent with the finding that immutable features tend to have other features depending on them (Love & Sloman, 1995). Performing a transformation across an immutable feature is more difficult because the other features depending on the immutable feature will be affected by the transformation.

### 3.2 Experiment 2: Transformation in Context

If the immutability of a feature is determined by the other features that depend upon it, then introducing (or highlighting) features that depend upon a feature should increase the immutability of the feature being depended on. For example, a context of "can fly" should make "has wings" more immutable, but since *depends* is a directional relation, "has wings" should have little effect on the immutability of "can fly". Experiment 2 tested this prediction.

#### Method

*Subjects.* Subjects were 20 undergraduates from Northwestern University who received course credit in an introductory psychology course for their participation.

*Materials and procedure.* The stimuli consisted of a subset of features from 4 categories (apple, chair, guitar, and robin) taken from Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976). The subset was chosen to allow A-B-C feature triads to be formed from the same category such that feature A depends upon feature B, but feature B does not depend upon feature A, and feature C shares no dependency relations with either features A or B. The dependency ratings were collected from subjects by Love and Sloman (1995). An example of such an A-B-C triad for the category robin is "can fly", "has wings", and "has a red breast".

Subjects rated the mutability of the features in the A and B sets by answering questions on a Macintosh computer such as, "How easily can you imagine the robin without wings?" Responses were made by pressing a number labeled key, with 1 indicating that the modified token was very easy to imagine and 9 indicating that the token was very difficult to imagine. The context was varied by having a feature in the triad precede the rated feature for 3 seconds in a statement like, "The robin can fly." Each subject saw only half of the stimuli to avoid having any subject rate a feature twice.

#### Results

The results are listed in table 1. When a feature was preceded by another feature that depends on it, the feature was judged more immutable than when it was preceded by a feature it did not share a dependency relation, with a mean of 7.37 compared to 5.52 ($t(197)=4.85$, $p<.001$).

This result did not hold for the reversal. When a feature was preceded by another feature that the rated feature depends on, the feature was not judged more immutable than when it was preceded by a feature that it shared no dependency relation: mean of 5.16 compared to 5.06 ($t(198)=.25$, $p>.80$).

| *preceding context* | *feature rated* | *mean of rated features* |
|---|---|---|
| *A* ( -------> ) | *B* | *7.37* |
| *C* (no relation) | *B* | *5.52* |
| *B* ( <------- ) | *A* | *5.16* |
| *C* (no relation) | *A* | *5.06* |

Table 1: Feature A depends upon feature B, but feature B does not depend upon feature A. Features A and B share no dependency relations with feature C. Higher ratings indicate that the feature is more immutable.

### Discussion

Experiment 2 supports the prediction that mutability context effects are mediated through dependency relations. Furthermore, it dispels the notion that dependencies are simple associations and that mutability is nothing more than the connectivity of relations, since dependency priming was shown to be asymmetric (e.g. the context of "has wings" did not make "can fly" more immutable, but the context "can fly" did make "has wings" more immutable.).

It may seem strange to some readers to call a feature presented on a computer screen a context. The rationale for this decision is that a richer context would activate the preceding feature. It seems reasonable to assume that a rich context about reproduction and birds would activate the feature "can lay eggs". By presenting the phrase "The robin can lay eggs.", I circumvent the need for a rich context.

## 4. The Role of Mutability and Dependency in Cognitive Processes

### 4.1 Categorization and Similarity

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.

Love and Sloman 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 was highly correlated with mutability judgments in all four categories. I predict that context affects these judgments in the same way that it affects mutability ratings.

Results in a similarity rating task mimic the categorization results. Subjects rate a token lacking a mutable feature as more similar to an ideal category member than a token lacking an immutable feature. This result holds over the entire continuum of feature mutabilities (as in the categorization result). This result suggests that models of similarity should not limit themselves to considering feature matches and mismatches (Tversky, 1977), but should also take into account the mutability of features when calculating similarity. Some features seem to count more than others in determining similarity. Commonalties and differences should be weighted by their immutabilities.

In contrast, Gentner, Markman, and Medin (submitted) propose that similarity comparisons involve an analogy-like process in which representations are aligned based on labeled relations. Distant analogies do require representations that have labeled relations. Calculating the similarity between an atom and a solar system may be an entirely different type of similarity process than calculating the similarity between a baseball and a tennis ball. A baseball and a tennis ball are easily comparable on perceptual properties and do not demand a comparison process that utilizes labeled relations. I would expect that the first similarity task would take longer because it requires the use of labeled relations for comparison. The distinction between using labeled relations and collapsing across relations in favor of dependencies is deeply related to dual processing theories (Sloman, 1996) which pit slow symbolic processes against fast associative processes. Perhaps, instead of proposing two distinct reasoning systems, a theorist only needs to outline the conditions under which relation labels are used instead of being discarded in favor of dependency information.

### 4.2 Mapping Processes in Metaphor and Conceptual Combination

Mutability may play a key role in the mapping process involved in interpreting nominative metaphors and noun-noun combinations. One can view the interpretation of a nominative metaphor as involving processes that transfer a property from the base noun to the target noun. Mutability theory constrains this mapping process. Immutable features of the target noun will resist conceptual change, while mutable features are more likely to accede to change.

For instance, the metaphor "This desk is a junkyard" can be interpreted as meaning the desk is messy since "is orderly" is a mutable feature of desk. Notice the mapping does not drastically transform or discard an immutable feature of desk, like "has a flat surface". Potential mappings that destroy immutable features of the target are rejected. This application of mutability theory does not suggest mappings, but constrains which mappings can be actualized.

An additional constraint on mapping processes is that the dependency structure of the target must be able to support the feature mapped over from the base. It is much more plausible to interpret the noun-noun combination *frog car*, as a green car, than as a car that hops, because the feature "can hop" has dependencies that are not satisfied in the target. Most cars can't hop. The dependency structure of car cannot support the addition of the feature "can hop". This constraint may also prove useful in searching for possible mappings.

## 5. Conclusion

I have given an account of how different types of relations can be collapsed to a dependency relation for the purposes of calculating feature mutability. I have also given an account of how context affects the transformation of features.

A question remains: If what is immutable varies with context, then what is the core of a category, if anything? Everything is not as slippery as it seems. Certain features will be present across contexts. Working cars always have engines, people always have brains (even if it doesn't always appear that way), etc. Some features are immutable across contexts. There are also strong constraints on what can be seen as immutable, such as a temporal order constraint (Byrne et. al., 1995; Kahneman & Miller, 1986) which can be viewed as another type of dependency (current events depend on past events). Ahn & Lassaline (1995) have shown that effects are more mutable than causes.

Also, not everything at the core of a category is related to immutability. Some categories have defining properties. For instance, a *red truck* has to be red, but no features of *red truck* depend on the feature "is red". Still, internal structure and dependencies are what scaffolds are understanding of categories.

## Acknowledgments

I thank Steven Sloman, Edward Wisniewski, Dedre Gentner, Lance Rips, and the anonymous reviewers for their comments.

## References

Ahn, W., & Lassaline, M.E. (1995). Causal Structure in Categorization, *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, 521-526. Pittsburgh, PA.

Barsalou, L.W. (1993). Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a System of Perceptual Symbols. In A.F. Collins, S.E. Gathercole, M.A. Conway, & P.E. Morris (Eds.), *Theories of memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Byrne, R.M., Culhane, R., & Tass, A. (1995). The temporality Effect in Thinking about What Might Have Been, *Proceedings of the Seventeenth Annual Conference of the CognitiveScience Society*, 385-390. Pittsburgh, PA.

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.

Dean, W. & Sloman, S. A. work in progress.

Gentner, G., Markman, A. B., & Medin, D. L. (submitted). The Structure of Similarity.

Gerrig, R.J., & Murphy, G.L. (1992). Contextual influences on the comprehension of complex concepts. *Language & Cognitive Processes.*, 7, 205-230.

Gick, M. L., Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.

Gildea, P., Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning & Verbal Behavior*, 22, 577-590.

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.

Love, B.C., & Sloman, S.A. (1995). Mutability and the Determinants of Conceptual Transformability. *Proceedings of the Seventeenth Annual Conference of the CognitiveScience Society*, 654-659. Pittsburgh, PA.

Medin, D. L., Goldstone, R. L., Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.

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.

Ratcliff, R. & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval.*Cognitive Psychology*, 2, 139-155.

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. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.

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

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