Cognition 118 (2011) 359–376

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The empirical case for role-governed categories

Micah B. Goldwater [a,*], Arthur B. Markman [b], C. Hunt Stilwell [b]

[a] Department of Psychology, Northwestern University, 2029 Sheridan Road – 102 Swift Hall, Evanston, IL 60208-2710, United States
[b] Department of Psychology, University of Texas at Austin, 1 University Station A8000 Austin, Texas 78712-018, United States

ARTICLE INFO

Article history:
Received 31 May 2010
Revised 8 October 2010
Accepted 8 October 2010
Available online 13 December 2010

Keywords:
Categorization
Concepts
Knowledge representation

ABSTRACT

Most theories of categorization posit feature-based representations. Markman and Stilwell (2001) argued that many natural categories name roles in relational systems and therefore they are role-governed categories. There is little extant empirical evidence to support the existence of role-governed categories. Three experiments examine predictions for ways that role-governed categories should differ from feature-based categories. Experiment 1 shows that our knowledge of role-governed categories, in contrast to feature-based categories, is largely about properties extrinsic to category members. Experiment 2 shows that role-governed categories have more prominent ideals than feature-based categories. Experiment 3 demonstrates that novel role-governed categories are licensed by the instantiation of novel relational structures. We then discuss broader implications for the study of categories and concepts.

1. Introduction

Theories of categorization typically assume that categories are represented by some set of features that describe the properties of category members (e.g. Cree, McRae & McNorgan, 1999; Love, Medin, & Gureckis, 2004; Rosch, 1973; Smith, Shoben, & Rips, 1974). Theories differ on whether the set of features is organized around a prototype, exemplars of the category experienced in the past, or subclusters of features that describe category members. New items are classified based on some function of the similarity of the new exemplar to the category representation.

This view of categorization does a nice job of accounting for data from studies of natural categories as well as studies of category learning. However, there is reason to believe that some categories are defined by taking into account the role that an object plays in a situation rather than by just collecting the features that describe category members (Gentner, Anggoro, & Klibanoff, in press; Gentner & Kurtz, 2005; Markman & Stilwell, 2001; McRae, Ferretti, & Amyote, 1997). For example, consider the category *guest*. Guests can certainly be characterized with typical properties (e.g., friendly, polite) but what really defines something as a guest is that it is someone or something that is visiting something else. The concept of *visiting* is a relation, because it relates two objects: one that is doing the visiting (the guest) and one that is being visited (the host). Markman and Stilwell (2001) called categories that name roles in a relational structure (such as guest or host) *role-governed categories*. Despite the interest in role-governed categories, there is little empirical evidence about how role-governed categories might differ from categories represented by feature sets (but see: Barsalou, 1983, 1985; Ferretti, McRae, & Hatherell, 2001; Gentner & Kurtz, 2005; Rehder & Ross, 2001; Rein, Goldwater, & Markman, 2010; Ross & Murphy, 1999).

In this paper, we present three experiments that illuminate key differences between role-governed and feature-based categories. To place these studies in perspective, we first discuss the concept of role-governed categories in more detail. Then, we examine three ways that role-governed categories might differ from feature-based categories and present studies that address them. In addition to the experiments, we also examine the use of role-governed categories on a photo-sharing website.

* Corresponding author.
E-mail address: micahbg@gmail.com (M.B. Goldwater).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.10.009

1.1. Role-governed categories

To place the idea of role-governed categories into perspective, we first consider the simple relation *x visits y* illustrated in Fig. 1. This figure assumes that people represent relationships among elements in their environment, that is, elements are explicitly bound by how they relate. Proposals for knowledge representation use structured relational representations to account for these relations (see Markman (1999), for a review). In the relation x visits y, the relation visits describes a relation between two elements x and y (which serve as the arguments to the relation). The elements x and y in this case are variables, and so they can be bound to a variety of different objects in different settings in order to allow the cognitive system to represent different instances of the event using the same relation.

To illustrate the importance of relational knowledge in thought, consider analogical reasoning processes (Gentner, 1983, 1989; Holyoak & Thagard, 1989). When forming an analogy, people seek similarities in the relations that describe a pair of domains, even if the objects that take part in those relations are not identical. Thus, the atom was seen to be similar to the solar system, because there is one element that revolves around another in each, even though electrons do not look like planets and nuclei do not look like the sun.

Markman and Stilwell (2001) pointed out that the arguments to a particular relation form a class of objects, namely those objects that play that role in some event. If items that play that role were somehow important to an individual then a label could be attached to that role and used as a category. So, a *guest* could be defined as the individual that visits. That is, the category would name the first argument to the relation x visits y in Fig. 1. Similarly, a *barrier* might name a class of items that blocks something, and a *shield* might name a class of items that is added to a system to protect it. On this view, then, whenever someone creates a relational system, they open up the possibility for a category that names the roles within that system. We test this possibility in Experiment 3.

1.2. A framework for distinguishing kinds of categories

Role-governed categories can be distinguished from three other types of categories (see Fig. 1): schema-governed categories, thematic categories and feature-based categories. Schema-governed categories refer to entire relational structures (Markman and Stilwell (2001) and see Gentner and Kurtz (2005) for a very similar discussion). For example, in x *visits* y, x is the role-governed category *guest* and y is the role-governed category *host*, but a *visit* refers to the relational system, or schema, *x visits y*, and thus is a schema-governed category. Another good example is the distinction between the schema-governed category *trip*, and the associated role-governed category *destination*. A trip is the whole relational system of travel, relaxing, the destination, etc. *Destination* picks out just one role within the schema.

Thematic categories (Lin & Murphy, 2001) label a set of items that all take part in a common relational system or event, though they play different roles within that event. So, cereal and milk may be grouped together because they are thematically associated, even though milk probably bears more similarity to other drinks than to cereal. This relationship is different from role-governed categorization because thematic relations group things together that play different roles in the same event, while role-governed categories group things together that play the same role across events. While thematic categories are distinct from role-governed and schema-governed categories, all three are kinds of relational categories because all three are

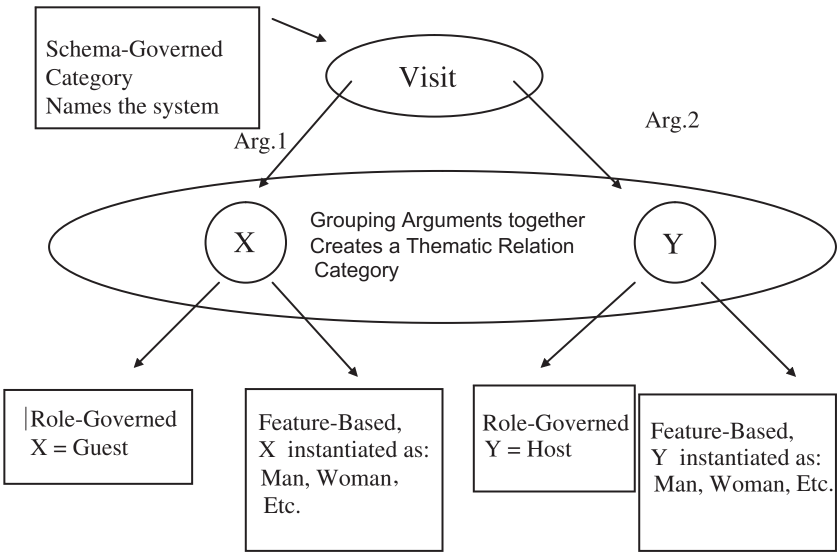

Fig. 1. A relational system and its constituent categories. A schema-governed category, e.g. *visit* names the system. Role-governed categories, e.g. *guest and host*, name individual arguments. Grouping the arguments together, creating a category of entities that includes guests and hosts, define a thematic relation category. In any given relational system, e.g. a visit, the roles are fulfilled with members of feature-based categories, e.g. men and women.

about the relations among entities and not just the entities in their own right.

Distinct from all three kinds of relational categories are feature-based categories. Feature-based categories have been given the most attention in the literature on artificial and natural categories. We will give only a cursory discussion of feature-based categories here to motivate the studies we present (see Murphy (2002), for a detailed review). Feature-based categories are represented as collections of features describing category members, e.g., birds are animals with wings and a beak. These features are primarily about the category members themselves and not about relations among category members and other entities (see Barr & Caplan, 1987, and expanded discussion below). Natural feature-based categories are typically thought to consist of natural kinds and artifacts, e.g. *dog* and *chair*. These categories are the most commonly used stimuli in semantic priming tasks (e.g., Cree, McNorgan, & McRae, 2006). They are used for Rosch's (1973) classic property-listing tasks. In addition, virtually every artificial category learning experiment has subjects learning categories easily represented by features.

While feature-based categories and role-governed categories are distinct, as shown in Fig. 1, their representations can become connected because whenever a role is filled, it is filled with a member of a feature-based category. For example, the roles of *guest* and *host* are most typically played by people. If roles are consistently filled with the same entities, their representations may become associated, as with *dog* and *pet* (see the discussion section).

In addition, many artifacts may seem to be role-governed categories because they are defined by their functional roles. However, in practice, artifact representations may be primarily feature-based because their functional roles are performed by objects that share a consistent set of features. Members of the class of blenders or vacuum cleaners, for example, all tend to look the same. To be a blender, all something has to do is blend stuff, but in reality, blenders all blend stuff in essentially the same way. Even when the function is not performed in the same way, many design features are kept consistent with other products. The Dyson vacuum cleaner uses a very different principle from traditional vacuums, but the general structure of the device is still similar to that of traditional vacuums. This suggests that designers are sensitive to the nature of people's categories. Finally, as we discuss below, there are norms collected by Goldwater, Asmuth, and Gentner (in preparation) that support this operating assumption.

There do appear to be many cases in which artifact and role-governed categories have very similar meanings, creating a sort of "minimal pair." One clear example is with *house* and *home*. The extensions of both terms overlap, however the terms suggest different construals of these extensions. *House* highlights a collection of descriptive features of a structure, while *home* highlights the role that an entity may play in a person's life. We propose that this particular kind of difference in construal – featural vs. relational – is rooted in a formal difference in mental representation (Markman, 1999). In addition, the distinction between featural and relational categories is not more than just a difference in generality because there is no strict subset relation. There are houses that are not homes (because nobody lives in them) and homes that are not houses (like apartments). We view it as a strength of our framework that it is possible to isolate the distinction between concepts that seem so similar on the surface.[1]

To summarize, we discussed four kinds of categories: role-governed categories, schema-governed categories, thematic categories, and feature-based categories. The first three are all kinds of relational categories because they are all defined by the relations among entities. The experiments in this paper focus on and are designed to provide empirical evidence for the nature of role-governed categories.

Experiments 1 and 2 distinguish between role-governed categories and feature-based categories. We focus on distinguishing those two, because feature-based categories are the major focus of the literature.[2] Experiment 1 uses a property-listing task to show that when people consider role-governed categories, they focus on ways that their members relate to other entities, whereas when people consider feature-based categories, they focus on just the members themselves and other entities to a much lesser degree. Experiment 2 connects role-governed categories to goal-derived categories (Barsalou, 1983, 1985) by examining whether ideal members are more important for role-governed categories than for feature-based categories (see Rein et al., 2010). Finally, Experiment 3 examines the relationship between role-governed categories and the relations from which they are derived. In this study, we teach people a novel relational category in the form of a new verb. We then ask whether learning this verb supports learning role-governed categories that name the arguments of the verb.

1.3. Motivating Experiments 1 and 2

Because Experiments 1 and 2 are methodologically dependent on each other, we give an overview of both. Experiment 1 examines the kinds of properties listed for both role-governed and feature-based categories. In Experiment 2, we use these property-lists in a forced-choice task to look at the difference in prominence of ideals and prototypes in the representation of role-governed and feature-based categories.

We know a lot about our natural categories, and role-governed categories are no exception. A productive way to investigate our knowledge is through a property-listing task (Barr & Caplan, 1987; McRae et al., 1997; Rosch, 1973). If role-governed categories are represented differently than feature-based categories, then some difference should be measurable in this task. Our proposal is that

[1] A similar example from our items is the role-governed category "drug" and the artifact "beer." While, beers have a number of typical descriptive features, drugs are any chemicals that cause a mental state change. In addition, drugs can be natural kinds or artifacts. A beer is certainly a typical example of a drug, but non-alcoholic beer is not a drug, so again, there is no strict nesting.

[2] For experiments distinguishing role-governed categories from thematic categories, see Goldwater and Markman (in press). Experiments distinguishing role-governed from schema-governed categories are on-going.

role-governed categories are pieces of, and are defined by, larger knowledge structures. Thus, people's knowledge about them should contain a lot of information about the larger knowledge structure of which they are a part and by which they are defined. In contrast, knowledge of feature-based categories should consist primarily of descriptions of category members themselves.

One important thing to note is that when we use the word "property" we mean what is written in a property-listing task. This is distinct from the word "feature," which we use to refer to a particular form of mental representation. The content of both structured relational representations and feature-based representations can be expressed with listed properties.

Barr and Caplan (1987) discuss one useful way to classify the properties listed in property-listing tasks. They distinguish between intrinsic and extrinsic properties. *Intrinsic* properties describe category members (without regard to other entities). For example, having fur is an intrinsic property of dogs. *Extrinsic* properties point outward to other objects. For example, being owned by people is an extrinsic property of dogs.

Barr and Caplan (1987) established a coding criterion for distinguishing extrinsic from intrinsic properties. In Experiment 1, we use their coding criterion, and predict that more extrinsic properties will be listed for role-governed categories than for feature-based categories because role-governed categories are about how their members relate to other entities. The feature-based categories we chose are artifacts. Artifacts exist because of their functions, and functional information is considered extrinsic by Barr and Caplan, so artifacts make a strict control condition, but unlike role-governed categories, which are represented more strictly by relational structures, their exemplars share a considerable number of features. So we predict that artifacts will elicit more intrinsic properties, and fewer extrinsic properties than role-governed categories. See the Methods section below for a more complete description of the materials.

Experiment 1 will extend Barr and Caplan's work. Their research assumed that all categories are represented by features, and so they did not have predictions about which categories were likely to have relatively more extrinsic than intrinsic properties. Gentner and Kurtz (2005) showed that members of feature-based categories share more intrinsic similarity than members of relational categories through the use exemplar generation studies. Experiment 1 extends both lines of research work by generating predictions for – and looking directly at – the types of properties listed for role-governed and feature-based categories.

We claim there is a qualitative difference between role-governed and feature-based categories, but it is important to specify what that means. Any quantitative measure of natural categories will create a continuum upon which all natural categories will vary (the proportion of extrinsic vs. intrinsic properties listed is no exception). This perhaps suggests that no qualitative distinctions can be made. However, to say that role-governed and feature-based categories are representationally distinct is to say that they elicit distinct *distributions* of, for example, extrinsic and intrinsic properties. As an analogy, consider the relationship between gender and height. Men and women are qualitatively different, and one measure of this is their distinct height distributions. While one could create a single height continuum, this would obscure the true pattern. Evaluating whether data sets come from distinct distributions is of course, the function of many statistical tests. Experiment 1 aims to show that role-governed and feature-based are qualitatively different in just this manner.

Experiment 2 is inspired by a common finding in the categorization literature that category members differ in their goodness of membership in a category, e.g. a Toyota Camry is a better example of a car than a Ferrari (even if a Ferrari is a better car). Often, the average member of a category is judged to be the most typical or "best" exemplar of a category. Indeed, prototype and exemplar models first had to explain why central category members are deemed more typical than peripheral category members (e.g., Medin & Schaffer, 1978; Nosofsky, 1986; Posner & Keele, 1968).

Central tendencies are clearly important for determining the typicality of exemplars of feature-based categories. However, research on *ad hoc*, and goal-derived categories (Barsalou, 1983, 1985) and studies of categorization by experts (Lynch, Coley, & Medin, 2000) demonstrate that ideal category members can influence people's beliefs about goodness of membership as well. Ideal members have extreme values on the relevant dimensions. For example, the goal-derived category "diet food" has an ideal value of 0 along the calorie dimension. If one were to average the calories of all the foods classified as diet foods, not only would the average be much larger than 0 but there may not be a single exemplar that has that value, and yet the goodness of membership in the category is based on distance from the ideal, and not the central tendency. Lynch et al. (2000) found that ideal category members also influenced judgments of tree experts who have goals related to their interactions with trees (e.g., arborists).

We suggest that role-governed categories are like goal-derived categories in that they have prominent ideals in their representations (see Rein et al. (2010) for related work). Like all categories, exemplars will vary in their goodness of membership. In role-governed categories, category members will differ in the degree to which they fulfill the role in the relational structure. Because the ability to serve the relational role is crucial, goodness of category membership should be related to this ideal rather than to the average values of the particular items that happen to be part of that category.

In Experiment 2, we assessed the role of typical and ideal category properties as predictors of category goodness. The measure we used for category goodness was the utility of category properties for explaining the category to someone else. Explanatory theories are crucial parts of the information people use to classify items (Ahn, Kim, Lassaline, & Dennis, 2000; Keil, 1989; Murphy & Medin, 1985; Rehder, 2003). If we are correct that ideals are more important for role-governed categories than for feature-based categories, then properties that pick out ideal category values should have a more prominent role for people trying to explain the category to someone else than should properties that pick out typical category values.

To test this hypothesis, some participants in Experiment 1 listed properties of typical category members of feature-based and role-governed categories. Other participants listed the properties of ideal category members. For Experiment 2, we constructed lists of the five most frequently listed ideal properties and the five most frequently listed typical properties. Properties that were listed frequently for both ideal and typical examples were not included. A new group of participants was shown the ideal and typical property lists and were asked which list they would use to explain the category to someone with no knowledge of it. We predict that participants will choose the list of ideal properties more often when explaining role-governed categories than when explaining feature-based categories.

Because we collect lists of ideal and typical properties in Experiment 1, we can examine the role of ideals in Experiment 1 as well. We would expect more extrinsic properties to be listed for ideal exemplars than for typical, because thinking about ideals should focus participants on the functional and extrinsic aspects of these categories. That is, something is ideal because it is ideal for some purpose, and is ideal relative to lesser category members. This is also why artifacts are a strict control for role-governed categories; they have obvious ideals because they are designed with a purpose in mind; a purpose that can be served ideally. In sum, for Experiment 1 we predict two main effects, with ideals and role-governed categories eliciting more extrinsic properties than typical exemplars and feature-based categories. We do not have specific predictions about the potential for interactions. For Experiment 2, we predict that more ideal property lists will be chosen as the best category member for role-governed categories than for feature-based categories.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Fifty-seven university of Texas at Austin students participated in this experiment for course credit.

2.1.2. Materials

Participants were given a packet with an instruction page plus 24 pages, each with a category label at the top and a series of blank lines below. The task was to list characteristics of either typical or ideal instances of the category, manipulated between subjects. There were 12 artifact terms: *television, chair, cell phone, fridge, truck, beer, website, shoes, knife, table, bicycle, microwave*. There were 12 role-governed categories: *guest, job, game, predator, hobby, gift, drug, customer, home, author, friend, pet.* Some of these categories were chosen, because they are clearly linked to relational verbs (see the motivation of Experiment 3 for a long discussion of the link between verbs and role-governed categories). The representations of *guest, author, gift, customer, home*, and *game* are dependent on *visit, write, give, shop, live, and play*. Here is a justification for the rest: *Job* and *hobby* are role-governed because classifying an activity as one or the other depends on the role it plays in your life, e.g. basketball is a job for some, but a hobby for many. A predator is x in x *hunts* y. A pet is an animal a person owns. Clearly, a person is classified as a friend based on the role they play in your social life. Something is a drug based on its ability to change one's state in some way. For example, industrial hemp, because it does not contain the psychoactive chemical THC found in other strains of cannabis, is not a drug. Microwaves and beer can also cause state changes, but similar to the discussion of blenders above, they each have a number of descriptive features that take prominence in their representations. All of the role-governed categories were selected because they bore no apparent derivational morphological relation to a verb, as e.g. *farmer* and *dancer* to avoid that relation as a confound.

In addition, Goldwater, Asmuth, et al. (in preparation) developed a large-scale norming of feature-based and relational (both role-governed and schema-governed) terms used in the current experiment, from Gentner and Kurtz (2005) and Gentner and Asmuth (2008). Naive raters judged the words on their degree of "role-ness" and "entity-ness." The ratings showed an interaction because the feature-based set from this experiment had reliably higher entity ratings (standardized scores from the entire set of 190 words, FB $M = 1.35$; RG $M = 0.23$) and the role-governed set had reliably higher role ratings (FB $M = -0.83$; RG $M = 1.17$), $p < .01$.[3]

For this experiment, we adapted the instructions used by Rosch (1973) for the Typical condition.

We are interested in how you think about everyday things in the world. Particularly we are interested in what you think are characteristics of **typical** examples of these things in the world. For example, a typical car seats 4–5 people comfortably, costs around $20,000 new, has about 175 horsepower, and gets around 20 miles per gallon of gasoline.

It is important that you do not just free associate. For example, if you had your first date in a car, do not list "first date."

There are 24 everyday things in the world we want you to write about in this packet. There is plenty space on each sheet for you to list characteristics.

The critical part of the instructions for the ideal condition read:

We are interested in how you think about everyday things in the world. Particularly we are interested in what you think are characteristics of what **ideal** versions of these things in the world would be. For example, an ideal car would seat 6 people comfortably, costs around $10,000 new, has about 300 horsepower, yet still gets 40 miles per gallon of gasoline.

It is important that you do not just free associate. For example, if you had your first date in a car, do not list "first date."

[3] This pattern held even for terms that appear very similar across sets, e.g., *drug* and *beer*.

There are 24 everyday things in the world we want you to write about in this packet. There is plenty space on each sheet for you to list characteristics.

2.1.3. Procedure

Participants sat at desks individually and were given the packet to fill out with a pen. They read the instructions and wrote down characteristics for all 24 categories. It took approximately 30 min.

2.2. Results

2.2.1. Coding procedure

All the listed characteristics were read carefully and paraphrases were re-worded to match each other to be able to properly calculate the most frequently listed properties. For example, *fuel inefficient* and *gas-guzzling* for *truck*, were considered paraphrases as they refer to the same property, just with a different register, while near synonyms such as *nice* and *kind* for *friend* were not considered paraphrases as their meanings have subtle differences.

We then determined which properties were listed by at least 10% of subjects for each category, typical and ideal.[4] Then, a different experimenter coded them as "intrinsic" or "extrinsic" blind to which category was the source of the property. A second coder also rated the properties. The two coders agreed with 89% of the judgments and all differences were easily resolved after brief discussion. We used the criterion established by Barr and Caplan (1987). To be intrinsic, the property had to be one that does not refer to any entities outside of the category member. For example, *big backyard* was an intrinsic property of *home*, because the backyard is part of the home; it's contained within the home's boundaries and no other entity needs to be considered. However, a property such as *comfortable* for home would be an extrinsic property because for something to be comfortable, it has to be comfortable for someone. Absolute physical dimensions were coded as intrinsic, e.g. *small*, but relative physical dimensions were coded as extrinsic, e.g. *right size*. Characteristics of people that were something just about that person, such as *cute* and *intelligent* were coded as intrinsic, but characteristics about how they behave towards others, e.g. *polite* and *boring*, were coded as extrinsic. Characteristics of artifacts that dealt with their internal parts, e.g. *screen* and *antenna*, were coded as intrinsic, but characteristics related to functions, e.g. *text messaging capabilities* and *clear reception* were coded as extrinsic.

2.2.2. Data

The two kinds of categories did not elicit different total numbers of properties across conditions, (Role-governed categories $M = 11.9$, and feature-based categories $M = 11.8$) so any differences in the numbers of intrinsic or extrinsic properties listed cannot be due simply to differing ease of listing properties of any type for either kind of category. In addition (see below), more extrinsic properties are listed overall across conditions (perhaps because they come from an unbounded set), but what matters for our framework is the relative number of the different kinds of properties elicited by the different kinds of categories. Because the total number of properties across kinds is nearly identical, any differences between category and instruction types cannot be due to the ease of listing different kinds of properties.

We present items analyses, because the properties to be analyzed were determined by averaging across participants. First we analyze the elicitation of intrinsic properties, see Fig. 2a. A 2 Category Type (Role-governed vs. Feature-based) × 2 Instructions (Typical vs. Ideal) Mixed ANOVA revealed two main effects and an interaction. Feature-based categories ($M = 6.5$), elicited more intrinsic properties than did role-governed categories ($M = 3.0$) $F(1, 22) = 6.56$, $p < .05$. The Ideal condition elicited fewer intrinsic properties ($M = 3.7$) than did the Typical condition ($M = 5.9$), $F(1, 22) = 22.68$, $p < .01$. In addition, Category Type and Instructions interacted because the decrease of intrinsic properties across instruction type was due to Feature-Based Categories (see Fig. 2a), $F(1, 22) = 21.00$, $p < .01$.

Next, we analyze the elicitation of extrinsic properties (see Fig. 2b). A 2 Category Type (Role-governed vs. Feature-based) × 2 Instructions (Typical vs. Ideal) Mixed

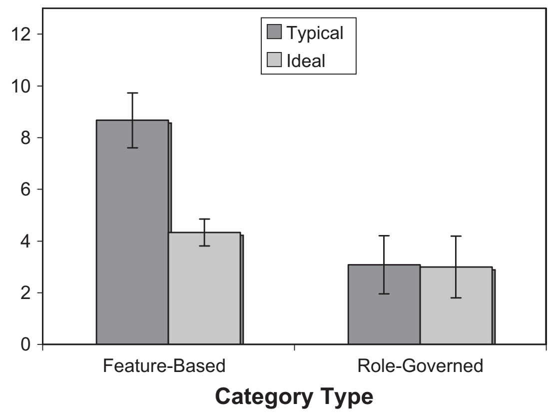

Fig. 2a. Items' means and standard errors of intrinsic properties listed in Experiment 1.

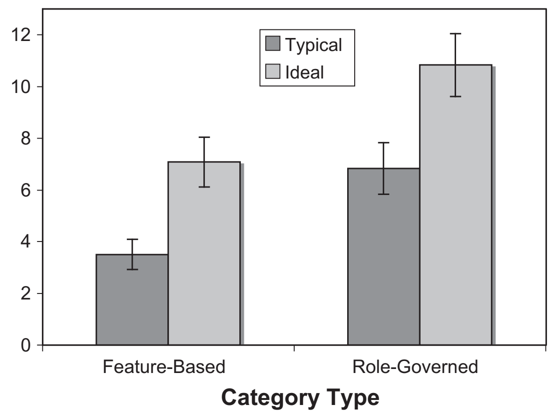

Fig. 2b. Items' means and standard errors of extrinsic properties listed in Experiment 1.

[4] It is not uncommon in the field to use a 20% cut off. Using that cutoff here does not change the reported pattern of results.

ANOVA revealed two main effects, but no interaction. Role-governed categories ($M = 8.8$), elicited more extrinsic properties than did feature-based categories ($M = 5.3$) $F(1, 22) = 9.31$, $p < .01$. The Ideal condition elicited more extrinsic properties ($M = 9.0$) than did the Typical condition ($M = 5.2$), $F(1, 22) = 27.60$, $p < .01$. Category Type and Instructions did not interact, because the increase in the number of extrinsic properties listed across instruction type was similar for both kinds of categories (see Fig. 2b), $F(1, 22) = 0.08$, $p > .75$.

Overall, the predicted pattern is borne out. Role-governed categories elicit more extrinsic properties and fewer intrinsic properties than feature-based categories. This held even though the feature-based categories were artifacts with widely known functions that elicited more extrinsic properties than the roles did intrinsic. In addition, the Ideal instructions elicited more extrinsic and fewer intrinsic properties than the Typical instructions did.

While the pattern of results is clear and the effects are large, there are additional differences between the two sets of items that could be accounting for these effects. First, the role-governed set has a mix of items that have typical animate and inanimate exemplars; for example, guests tend to be animate, and drugs are inanimate. In contrast, all the feature-based categories are inanimate. However, this difference does not account for the advantage of elicited extrinsic properties, or disadvantage of elicited intrinsic properties for role-governed categories. As can be seen in Table 1, the differences between the animate and inanimate role-governed categories is negligible, confirmed by ANOVA's, main effect F's < 0.1, p's > .8.

Second, and more theoretically relevant, role-governed categories tend to be more abstract than feature-based categories. There are many forms of abstractness, but one possibility of concern is imageability. If people list extrinsic properties for role-governed categories just because they cannot think of any intrinsic properties, then that would not be strong support for our hypothesized differences between feature-based and role-governed categories.

To explore this possibility, we obtained imageability ratings from a separate set of participants for our set of items. On a 7 point Likert scale, the feature-based categories were rated as more imageable ($M = 6.6$) than the role-governed categories ($M = 5.2$), $t(22) = 6.02$, $p < .01$. Nonetheless, tests of mediation suggest that this difference does not account for the difference in the number of extrinsic properties listed for feature-based and role-governed categories. "A variable is considered a mediator to the extent that it carries the influence of a given independent variable to a given dependent variable. Mediation can be said to occur when (1) the IV significantly affects the mediator, (2) the IV significantly affects the DV in the absence of the mediator, (3) the mediator has a significant unique effect on the DV, and (4) the effect of the IV on the DV shrinks upon the addition of the mediator to the model" (http://people.ku.edu/~preacher/sobel/sobel.htm).

Table 1
Means and standard deviations of numbers of properties listed for animate and inanimate role-governed categories.

	Intrinsic		Extrinsic	
	Typical	Ideal	Typical	Ideal
Inanimate	3.7 (5.3)	2.5 (5.2)	6.2 (2.5)	11.0 (3.0)
Animate	2.5 (2.1)	3.5 (3.2)	7.5 (4.4)	10.7 (5.5)

We conducted Sobel tests of mediation to examine whether imageability accounted for the effect of category type (Feature-Based vs. Role-Governed) on the kinds of properties listed. The tests showed that imageability mediated the effect of category type on intrinsic properties in the typical condition, $z = 2.19$, $p < .05$, and in the ideal condition $z = 1.87$, $p = .06$. However, imageability did not mediate the effect of category type on extrinsic properties in the typical condition, $z = 0.54$, $p = .59$ nor in the ideal condition $z = 1.19$, $p = .23$. Thus, while the imageability of our items is strongly connected to the number of intrinsic properties they elicit, this same connection does not hold for the number of extrinsic properties, i.e., the properties referring to how the category members relate to other entities.

2.3. Discussion

Experiment 1 tested two hypotheses. First, we examined the prediction that extrinsic properties are elicited more often by role-governed categories than by feature-based categories because role-governed categories are about how their members relate to other entities. This prediction follows straightforwardly from the framework for relational categories, though it is not predicted by other standard theories of categorization. Second, we explored the prediction that extrinsic properties would also be listed more often by people thinking about ideal category members than by people thinking about typical category members because ideals are ideal for some purpose beyond the category member itself, and are ideal relative to other category members. Both hypotheses were supported.

We were able to rule out the possibility that these findings were due primarily to differences in imageability of the items. Our framework suggests that a crucial difference between role-governed categories and feature-based categories is the degree to which their representations reflect the relationship between category members and other entities. However, in addition to being relational, role-governed categories tend to lack descriptive features. A lack of descriptive features is one definition of abstractness, and as the mediation tests show, there is a strong connection between imageability and elicited *intrinsic* properties. However, while relationality is associated with a lack of concrete features, relations are not just about the features they lack. Crucially, the effect of category type on elicited extrinsic properties was not mediated by imageability.[5] This analysis showed that relationality is not identical to abstractness; that is, relationality influences category representation beyond abstractness.

[5] This pattern is quite sensible, as things denoted by intrinsic properties, e.g., visual features, are quite imageable, while things denoted by extrinsic properties, e.g., social relations, are not. But again, things such as social relations are about more than just what they don't contain.

In addition, this data set makes predictions about ''categorization in the wild,'' (Glushko, Maglio, Matlock, & Barsalou, 2008) which we turn to next. To further distinguish role-governed from feature-based categories on the basis of the prominence of extrinsic properties, we examined a naturalistic data set obtained by analyzing the photo-sharing website flickr.com. While laboratory experiments will always be the primary way to investigate category representation, the world-wide-web provides behavioral scientists with extremely rich data sets of how category labels are used (specifically, how visual images are labeled) in the real world. Clearly, predicting real world use of categories should be a goal of all theories of categorization.

On Flickr.com, people upload their pictures, give them titles and describe them. Any user then can ''tag'' any of the uploaded photos with a label. For example, there is a photo of a dragonfly on a flower titled ''My regular guest'' (see Fig. 3) that has been tagged with *dragonfly*, *nature*, and *flower* among other things. The image also has a description associated with it, describing that these dragonflies live in the photographer's garden, etc. Essentially, flickr.com can be seen as a corpus of natural categorization free of artificial laboratory conditions. People tag photos because they feel like it, not because they need to fulfill a course requirement. Tagging is relevant to us, because we think tagging will be based more on descriptive features than titling or describing a photo will be. The person who titles and describes the photo took the picture, experienced the moment it captured, and has a representation of the situation that the picture depicts. However, the taggers just come along see a picture and name it, without all the relational situation knowledge. They probably just name what they see, i.e. the descriptive features of the objects pictured. ''My regular guest'' exemplifies this proposal. The photographer experienced this dragonfly as her guest, i.e. the dragonfly was seen as visiting her. Looking at the picture, not being there to be visited, the tagger just sees a dragonfly and a flower, two feature-based categories.

Flickr.com provides the user with a very easy way to verify our proposal. There are two ways to search for photos on flickr.com. One type of search uses all the text associated with the photo, which includes the titles, descriptions, and the tags, and the other just uses the tags (see Fig. 3). We searched for all our categories using both search types. Definitionally, the first type of search will get more hits because it is a superset of the second. We predicted that the proportion of the superset included in the tag-only subset (the ''flickr tagged proportion'') should be lower for role-governed category labels on the photos than it is for the feature-based category labels on the photos. As predicted, a lower proportion of the role-governed category labels were included in the tag-only subset ($M = .21$) than the feature-based category labels ($M = .34$),

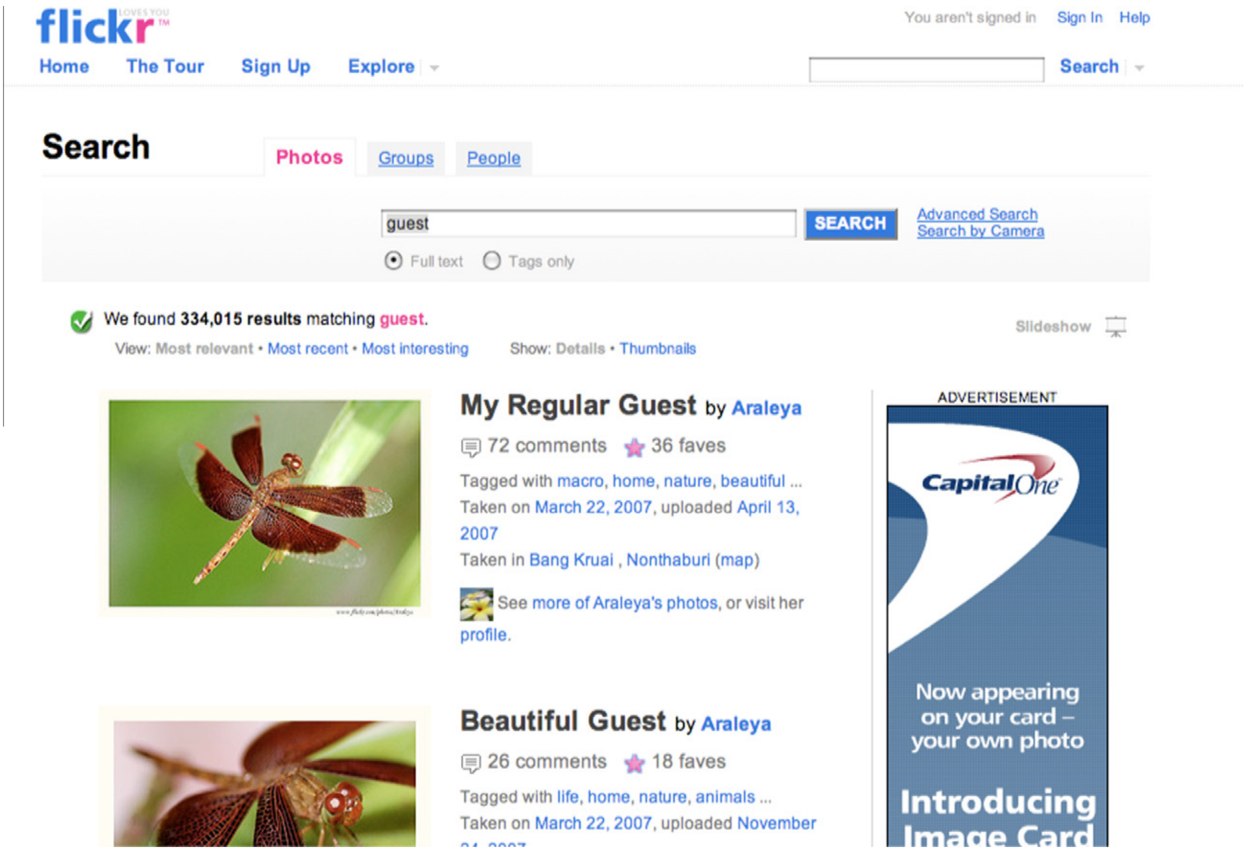

Fig. 3. Screen shot of full text search for guest.

Table 2
Flickr search hits on 4/30/2008 and overall proportion extrinsic properties.

Category	All text hits	Tagged only	Tagged only (%)	Extrinsic (%)
Beer	10,81,889	564,961	0.52	0.56
Bicycle	582,075	388,066	0.67	0.41
Cell phone	174,399	57,224	0.33	0.61
Chair	553,629	136,730	0.25	0.34
Fridge	79,476	18,648	0.23	0.39
Knife	102,813	31,522	0.31	0.37
Microwave	24,798	5725	0.23	0.33
Shoes	682,369	234,912	0.34	0.39
Table	826,758	152,115	0.18	0.24
Television	122,220	65,555	0.54	0.42
Truck	639,376	225,427	0.35	0.54
Website	256,133	43,423	0.17	0.93
Author	97,941	14,283	0.15	0.65
Customer	69,373	7143	0.10	0.88
Drug	97,090	6427	0.07	0.81
Friend	59,91,136	269,683	0.05	0.94
Game	19,46,118	394,008	0.20	0.94
Gift	592,433	92,053	0.16	0.79
Guest	287,132	16,089	0.06	0.96
Hobby	92,150	47,793	0.52	0.95
Home	33,69,970	10,99,971	0.33	0.36
Job	331,765	32,968	0.10	1.00
Pet	10,19,514	449,931	0.44	0.58
Predator	41,819	13,813	0.33	0.35
Feature-based means (SD)			0.34 (.16)	0.46 (.18)
Role-governed means (SD)			0.21 (.16)	0.77 (.23)

$t(22) = 2.11$, $p < .05$. Looking at Table 2, which lists the number of hits from each search for each category, it is heartening that this effect reflects a large number of actual items that occurred in a natural setting.

To further relate the findings of Experiment 1 to the flickr.com data, we regressed the proportion of extrinsic properties collapsed across the Typical and Ideal instructions against the flickr tagged proportion. We used both Typical and Ideal conditions because there is no good apriori reason to think one would be a better lens into how people categorize "in the wild" (and we used a proportion of extrinsic properties, instead of intrinsic and extrinsic separately, for simplicity). However, the pattern of results is unaffected if you just use one instruction type or the other. Looking at the scatterplot (Fig. 4), two things are apparent. One is that, as predicted, an increased proportion of extrinsic features predicts a decrease in the flickr tagged proportion, as shown by the linear regression, $y = -0.30x + 0.46$ $R^2 = 0.21$. The second is that the relationship is not linear, but better fit by an exponential function. While an increase in proportion of extrinsic properties predicts a decrease in flickr tagged%, the flickr tagged% decreases at a lesser rate when the proportion of extrinsic properties approaches ceiling. This leads to a substantial increase in the amount of variance accounted for when using an exponential regression, $y = 0.62e^{-1.67x}$ $R^2 = 0.34$. The flickr tagged proportion is based on millions of naturally occurring behaviors. Over a third of the variance is accounted for by the constrained, artificial, and superficially unrelated experimental task. We believe this result reflects that both phenomena arise from the use of role-governed categories.

3. Experiment 2

The goal of Experiment 2 is to build on the findings of Experiment 1 that ideals are a more prominent part of the representation of role-governed categories than of the representation of feature-based categories. Experiment 1 hints at this in two ways. First, the interaction between Category Type and Instruction Type occurs because feature-based categories typically elicit a low proportion of extrinsic properties, though the proportion rises substantially when people are forced to think about ideal members of feature-based categories. This finding suggests that people know how to express ideal properties of feature-based categories, but do not think of them as typical. In contrast, the typical properties elicited by role-governed categories are extrinsic as are the ideal properties. This commonality may be a cause or an effect of ideals being important for role-governed categories.

Because Experiment 1 was not designed to investigate the prominence of ideals, the results of Experiment 1 are not sufficient to draw conclusions about the importance of ideals for role-governed categories. Experiment 2 investigates this question directly by using the methodology

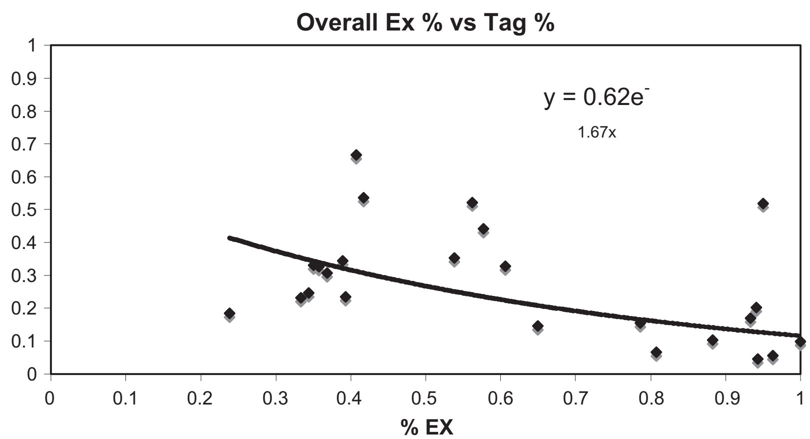

Fig. 4. Scatterplot of the overall proportion of extrinsic properties from experiment 2 and the flickr tagged proportion.

described above. Participants choose between a list of ideal and typical features as the best description of a category. We expect that people will select exemplars with ideal features as the best description of role-governed categories, and exemplars with typical features as the best description of feature-based categories.

3.1. Methods

3.1.1. Participants

Thirty-one university of Texas at Austin students participated for course credit.

3.1.2. Materials

Each of the 24 category labels from Experiment 2 was presented on a screen, one at a time. Below the label were two lists of characteristics. These lists were labeled A and B. The two lists of characteristics were created by using the properties listed in Experiment 1. One list consisted of the five most frequent properties listed in the ideal condition of Experiment 1 (that were not among the five most frequent items listed in the typical condition). The other list consisted of the five most frequent properties listed in the typical condition (excluding those items that were among the five most typical listed in the ideal condition). Thus, each list had a set of five unique properties. The labels A and B were randomly assigned to a list for each subject. See Appendix A for a complete listing of properties.

The instructions read:

> We are interested in how you think about everyday things in the world, e.g., a car, so that you could EXPLAIN these things to a person with no knowledge of them, as if this person has spent their whole life living in the woods, or perhaps someone who has amnesia. At the top of each page there is a word for an everyday thing in the world. Underneath each word there are two separate lists of characteristics (one on the left and one on the right) that could describe these everyday things. It is your job to decide which of these two lists of characteristics you would use to describe an ILLUSTRATIVE EXAMPLE of these everyday things as a way to EXPLAIN them to the person with no knowledge.

3.1.3. Procedure

Participants were tested individually, seated in front of a computer. Instructions and stimuli were presented using *E-Prime*. There were 24 trials. Participants either pressed A or Z to indicate which characteristic set they thought more explanatory for each trial. The task took about 10 min.

3.2. Results and Discussion

As predicted a higher proportion of ideal characteristic sets were chosen for the role-governed categories ($M = .45$) than for the feature-based categories ($M = .30$), $t(30) = 6.18$, $p < .001$ by participants, $t(22) = 2.41$, $p < .05$ by items.

The goal of this experiment was to show that ideals are more prominent in the representation of role-governed categories than they are for feature based-categories as defined by their ability to illustrate and explain the category to a novice. A strength of this experiment is that the characteristic sets were generated by the participants in Experiment 1 and not from the intuition of the experimenters. It is worth noting however, that even for the role-governed categories the typical lists were still chosen more frequently than the ideals. We believe that this result makes sense, because ideals do not necessarily exist in the real world.

We showed an ideal advantage for role-governed categories in a 2AFC task. Rein et al. (2010) replicated this finding with the same materials using continuous representativeness ratings. Instead of construing the property list as a potential explanation for a category, Rein et al. told participants they were the characteristics of a particular member of the category. Participants rated how representative of the category that member was. Feature-based exemplars made from the typical list had higher representative ratings than the typical role-governed exemplars, while role-governed exemplars made from the ideal lists were rated as more representative than the ideal feature-based exemplars.

The psychological prominence of particular features is clearly important for feature-based categories (tautologically so), as it plays a clear role in determining the goodness of category membership. However, the prominence of particular properties should be less important for role-governed categories, because they are defined by whether an entity fills a relational role rather than the feature they contain. Thus, the effectiveness of the object at filling the role is the key determiner of goodness of category membership for role-governed categories.

To test this proposal, we examined the production frequencies of the properties from Experiment 1 as a proxy for their psychological prominence. The average production frequencies of the items for all categories were calculated for both the typical and ideal condition. Then each item was given a typical-ideal difference score, which was simply the difference between the average production frequency of the five properties of the typical condition and the five properties of the ideal condition. This difference was then correlated with each of the proportion of times that the list of characteristics for the typical condition was chosen in Experiment 2. There was a significant correlation for the feature-based categories, $r(11) = .41$, $p < .05$, but no such correlation for the role-governed categories, $r(11) = .16$, $p > .2$. That is, property prominence predicted the likelihood that an item would be considered a good exemplar of a feature-based category, but not for a role-governed category.

4. Experiment 3

Experiments 1 and 2 demonstrated two basic properties of role-governed categories: that they point outward to other objects and that they are more driven by ideal characteristics than feature-based categories. If role-governed categories are as ubiquitous as we claim, however, then it should also be straightforward to acquire them. In Experiment 3, we examine how role-governed categories might be learned in the context of learning new verbs.

Our participants instantiated relational structures by interpreting novel denominal verbs (i.e. novel verbs derived from nouns). Many familiar verbs are denominal, e.g. *dust*, *shelve*, *saddle*, *google*, etc. (see Clark and Clark (1979) for a comprehensive taxonomy). In addition, *novel* denominal verbs are readily understood in their first encounter in on-line sentence comprehension (Goldwater & Markman, 2009, see Kaschak & Glenberg, 2000 for an off-line comprehension task) suggesting they are a reasonable object of study for the instantiation of novel semantic representations.[6]

The use of novel denominal verbs also allows the instantiation of relational representations that are rooted in pre-existing knowledge, allowing concept learning to be rapid and to be embedded in the simple reading of short passages. For example, consider the sentence: "At Mardi Gras, Paul *whiskied* himself stupid." We can test whether this instantiation licenses a novel role-governed category by using the –er morpheme in English to create a novel agent term. So, later in a passage containing the sentence about Paul, we could refer to him as *the whiskier*. If people are able to form novel role-governed concepts when they instantiate a relational structure, then being exposed to a novel denominal verb should immediately allow them to understand the role-governed category.

Before giving more detailed description of the materials and procedure, we describe the process of interpreting novel denominal verbs and agents and explain why we think this is an appropriate way to examine the link between relational structures and role-governed categories.

A novel denominal verb changes a noun (usually one that refers to an object or substance) into a verb. Nouns and verbs have quite different representations. As discussed above, object nouns are usually thought to be represented by collections of features. In contrast, verbs provide relational structures that connect the elements of sentences (Gentner, 1982). Verb representations are not decomposed into features, but rather into more primitive relations (Gentner, 1975; Jackendoff, 1990, 2002; Levin & Rappaport-Hovav, 2006; McKoon & MacFarland, 2002; Wechsler, 1995). We give examples of verb relational structures for *give* and *exchange* to illustrate this point (for a more detailed analysis of the semantics of verbs of possession see Gentner, 1975). Give involves a transfer of possession of some object Z from person X to person Y.

Give: X CAUSE [Y POSSESS Z]

Exchange is a reciprocal giving event, so when one person (X) gives some object (Z) to a second person (Y), Y in turn gives a second object (W) to X. That leads to the rather complex relational structure

Table 3
Verbs and corresponding role-governed categories.

Verbs	Role-governed categories
x steals y	x = thief
x visits y	x = guest, y = host
x trains/adivses y	x = mentor, y = protégé
x gives birth to y	x = mother, y = son or daughter
x defeats y	x = victor/winner, y = loser
x eats y	y = food
x plays y	y = game
x writes y	x = author
x gives y to z	y = gift
x shops in y	x = customer
x lives in y	y = home

Exchange: [X CAUSE [Y POSSESS Z]] CAUSE [Y CAUSE [X POSSESS W]]

The relations underlying verb meanings are predicates that bind arguments. This means that these representations include more than just the aspects of the events; they include placeholders for potential arguments, known as thematic roles. These thematic roles are frequently represented by variables as in the examples just presented, but we believe that these variables are not just empty placeholders. Instead, they include rich conceptual information, i.e. the stuff of role-governed categories.

There seems to be a special relationship between verbs. It seems that every role-governed category noun has a corresponding verb or verb phrase. As an additional example to *guest* above, the concept of a "thief" relies crucially on verbs like "steal," because the defining characteristic of a thief is that this individual is the first argument to the relation x *steals* y. Table 3 lists more corresponding verbs and role-governed categories. In addition to these example sets that do not bear a morphological relation, the "–er" morpheme in English allows us to freely derive terms for typical agents, a type of role-governed category, from verbs, e.g. *dance* and *dancer*.

In addition to the linguistic evidence, there is experimental evidence supporting the idea that verb representations include the conceptual information about their typical arguments. Ferretti et al. (2001) used a lexical decision task to show that verbs prime their typical agents, e.g. *arrest* primes *cops,* their typical patients, e.g. *arrest* primes *criminal,* typical instruments, e.g. *stir* primes *spoon*, and features of patients, *manipulate* primes *naive*. Interestingly, computational models of semantic memory currently cannot account for these findings because they rely on featural overlap between concepts (Cree et al., 1999). Thus, they cannot represent relational structures.[7]

In Experiment 3 we test the link between relational and role-governed categories by having participants read passages with novel denominal verbs and novel agent terms

[6] Goldwater and Markman showed that novel denominal verbs are comprehended well enough the first time they are encountered to enable readers to detect subtle distinctions in their sentences' event semantics, an aspect of semantics thought to be rooted in verb representation (Jackendoff, 1990; Mauner & Koenig, 2000 among others, and see the Discussion section below for an explanation of this process).

[7] It is possible to model these priming results with semantic networks, but the "agent" and "patient" links between nodes are not built in (Collins & Loftus, 1975), and it would be post hoc to add them (Ferretti et al., 2001). The problem with semantic networks is that anything could be built into them post hoc, and so if this practice is used, then they become non-falsifiable.

derived from those novel verbs. The passages contained five sentences, (see Table 4 for a complete example, and Appendix B for more sample passages). The second sentence introduced the novel denominal verb, e.g. "On the first night, Paul *whiskied* himself stupid." The fifth sentence presented the novel agent term, e.g. "The next day, the *whiskier* rested to get ready for the third night." The comprehension of the agent term should be facilitated when compared to reading it after one of two control conditions.

The control condition passages were identical except for the second sentence. The first control uses a paraphrase of the sentence with the novel denominal verb, but without the lexical innovation, e.g. "On the first night, Paul used *whiskey* to drink himself stupid." The second control condition uses a novel adjective derived from the same root noun, e.g. "On the first night, Paul had a *whiskeyful* time." This second control condition was included to rule out a "general novelty effect." One could argue that the novel agent term is read faster after the novel verb just because once one has read a novel word derived from a root word, it is then easier to understand any other novel word derived from that same root word. If the novel denominal verb condition has an advantage over the novel adjective condition, then this general novelty effect cannot explain our results in entirety, and instead we will have found support for the special link between verbs and agents.

In all conditions, the event consisting of Paul drinking a lot of whiskey is understood, and so the term *whiskier* is understandable. Because of the special link between verbs and agents (and thus between relational and role-governed categories), understanding the same novel agent term after the novel denominal verb should be easiest. We measure ease of understanding by using the standard self-paced reading paradigm (Just, Carpenter, & Woolley, 1982) where reading time is a marker of ease. Detailed methods explained below.

4.1. Methods

4.1.1. Participants

Sixty university of Texas at Austin undergraduates participated for course credit.

4.1.2. Materials

Sample critical passages can be seen in Appendix B. There were three passage types for each of the 27 novel agent terms. Nine contained novel denominal verbs, nine contained a paraphrase without a lexical innovation, and nine referred to the same events but with novel denominal adjectives (see Table 4 and Appendix B). Many of the novel denominal verbs are based on examples from Kaschak and Glenberg (2000). The paraphrases all included the root noun of the novel verbs in approximately the same position in the sentence. Three lists were created such that each participant only saw one passage for each agent term and so the passages were counterbalanced across lists. The final sentence of each critical passage shared the same form, see Table 5. There were 43 filler passages with no lexical innovations, making 70 in total. There were also 70 comprehension questions about the passages. Appendix C shows typical questions. Sixteen comprehension questions asked about the meaning of the novel agent term, sampled from all conditions.

4.1.3. Procedure

Participants were tested individually, seated in front of a computer. Stimuli were presented using *E-Prime*. Participants were told that we were interested in what was remembered from text after reading at a natural pace; they were instructed not to study the text. They were informed that they would be reading passages and then they would be asked questions about what they remembered and understood about the passages. It was explained that the first four sentences of each passage would be presented as whole but then the final sentence would be presented one or a few words at time. They were instructed to press the spacebar as soon as they comprehended what was on the screen. During the passage final sentences, "#"s masked all the preceding and subsequent regions, modeled on the self-paced reading time procedure from Just et al. (1982). The regions are as defined in Table 5. They had 15 s to answer each question. There were 14 blocks of five sentences followed by five comprehension questions each. Block order and order of passages and questions within blocks were randomized. In between blocks they could

Table 4
Critical passage types for Experiment 3.

Novel verb
Paul took a week off from school to go to New Orleans for Mardi Gras. On the first night, he whiskied himself stupid. He got very sick and slept in the next day. Paul's friends gave him a lot of grief. On the second night, the whiskier took it easy to get ready for the next night
Paraphrase
Paul took a week off from school to go to New Orleans for Mardi Gras. On the first night, he used whiskey to drink himself stupid. He got very sick and slept in the next day. Paul's friends gave him a lot of grief. On the second night, the whiskier took it easy to get ready for the next night
Novel adjective
Paul took a week off from school to go to New Orleans for Mardi Gras. On the first night, he had a whiskeyful time. He got very sick and slept in the next day. Paul's friends gave him a lot of grief. On the second night, the whiskier took it easy to get ready for the next night

Table 5
Structure of novel agent term sentences for Experiment 3.

[On the second night,]$_1$ [the whiskier]$_2$ [took]$_3$ [it]$_4$ [easy]$_5$ [to]$_6$ [get ready]$_7$ [for the next night.]$_8$
[Temporal phrase,]$_1$ [the agent term]$_2$ [predicate]$_3$ [argument]$_4$ [argument]$_5$ [to]$_6$ [predicate]$_7$ [argument]$_8$

take a break for as long as they wanted. The session lasted around 45 min.

4.2. Results

4.2.1. Comprehension questions

Questions about the meaning of the novel agents were included to make sure their meaning was comprehensible. The accuracy for these questions was high at 88% and there was no difference across lexical innovation conditions. The overall accuracy on comprehension questions was 83%.

4.2.2. Reading time

In reading time studies that examine complex semantic processes, effects may not emerge until the region following the critical word (e.g. Traxler, Pickering, & McElree, 2002). So, here we analyze the novel agent term and the following predicate. As can be seen in Fig. 5, the predicted differences emerge at the predicate following the agent term. We conducted a 3 (Lexical Innovation: Novel Verb, Paraphrase, Novel Adjective) × 2 (Sentence Region: Agent Term, Predicate) repeated measures ANOVA that revealed a main effect of Sentence Region because the novel agent term was read slower than the following predicate $F(1, 59) = 20.34$, $p < .01$ by participants, and $F(1, 26) = 41.4$, $p < .01$ by items. In addition there was a main effect of Lexical Innovation, $F(2, 118) = 3.19$, $p < .05$ by participants, but this did not reach significance by items $F(2, 52) = 2.54$, $p < .1$ by items.

There was an interaction between Lexical Innovation and Sentence Region, $F(2, 118) = 3.08$, $p < .05$ by participants, and $F(2, 52) = 3.20$, $p < .05$ by items, because at the agent term, (NV M = 672 ms, P M = 674 ms, ADJ = 658 ms), no effects approached significance by participants or items, all t's < 1., all p's > .4, while at the predicate term, the Novel Verb condition (M = 513 ms) was significantly faster than the both the Adjective (M = 544 ms) and the Paraphrase conditions (M = 574 ms), $t(59) = 2.58$, $p < .05$, $D = .33$; $t(59) = 3.88$, $p < .01$, $D = .5$ by participants, and $t(26) = 2.45$, $p < .05$, $D = .47$; $t(26) = 3.82$, $p < .01$, $D = .73$ by items. The Paraphrase and Novel Adjective conditions were not reliably different from each other by participants, $t(59) = 1.66$, $p > .1$, $D = .21$ or items, $t(26) = 1.52$, $p > .1$, $D = .29$.

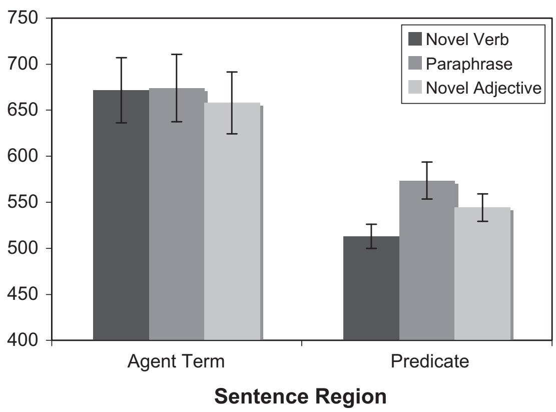

Fig. 5. Participants' means and standard errors of reading times for Experiment 3.

4.3. Discussion

The goal of Experiment 3 was to show that interpreting a novel denominal verb licensed the creation of a novel agent concept, and that this was specific to the relationship between verbs and agents and not just between any two novel words derived from the same root. Experiment 3 achieved those goals by showing faster reading times for the predicate following the novel agent term in the novel verb condition than the novel adjective and paraphrase conditions. This provides crucial evidence for our proposal about the representation of role-governed categories. Role-governed categories are pieces of relational structures, so part of instantiating a relational structure is creating placeholders for role-governed categories. This result also speaks to the ubiquity of role-governed categories. They are licensed by every verb we have learned, and we know thousands of verbs.

One issue that was not addressed when motivating this study was the process of making an object category a relational category. That is, where do the relations come from? Kaschak and Glenberg (2000) and Goldwater and Markman (2009) examine the way novel denominal verbs acquire relational structures. Novel denominal verbs assume the relational structure of their sentences' event semantics. The event structure of a sentence is the "who did what to whom" aspect of a sentence's meaning, e.g. an agent acting on a patient causing it to change its state. Construction Grammar (e.g. Goldberg, 1995) posits that syntactic forms are united with event structure, forming "Grammatical Constructions."[8] When one interprets a novel denominal verb, the verb assumes the event structure of the sentence's grammatical construction. An example from Kaschak and Glenberg, and co-opted for the present study is the verb *crutch* in, "John *crutched* Bill the apple." Here, *crutch* has an element of possession transfer that is not in the sentence "John *crutched* Bill" or " John *crutched* the apple." This is because the sentence form "[Noun Phrase]$_1$ [Verb Phrase [Noun Phrase]$_2$ [Noun Phrase]$_3$]", the ditransitive construction, is united with the relational event structure of "X_1 CAUSE [Y_2 HAVE Z_3]." The sentence form, "[Noun Phrase] [Verb Phrase [Noun Phrase]]" is not.

Our interpretation of the current results is that the novel verb interacts with the syntactic form of the sentence to inherit the sentence's event structure. In the sentences we have used, this event structure contains a number of thematic roles, including a role for an agent. Then, we use the familiar "-er" morpheme that signals that the term refers to an agent. People then map this agent term onto the event structure created by the initial presentation of the novel verb.

This is clearly an explanation of relational structure instantiation that relies on linguistic representations, as it

[8] Other conceptions of Construction Grammar, e.g. Kay and Fillmore (1999) and Kay (2005), focus more on other semantic domains than event structure, and have a slightly different view of how verbs and constructions unify.

depends on grammatical semantics. However, we do believe that this is still informative about concepts generally, and not just a reflection of linguistic and grammatical processing. In an event-related potential replication of Experiment 3 (Goldwater, Markman, Trujilli, & Schnyer, in preparation), the reading time advantage for the novel agent following the verb was reflected in ERP's consistent with an N400 effect, a marker of general semantic processing (e.g., Kutas & Hilyard, 1980). If the advantage for the novel verb condition was, for example, simply rooted in morpho-syntactic derivation, as the "-er" morpheme is most often used to derive agent terms from verbs, then the agent terms in the paraphrase condition would have elicited an increased P600, a marker of detecting a morpho-syntactic anomaly (see Kim and Osterhout (2005) for a review).

Experiment 3 shows that a relational category can be drawn from a feature-based category by applying a known schema, often associated with that feature-based category. This schema instantiates a novel relational structure. Then roles within that relational structure are licensed, and can be reified with a label. These processes are likely to be related to those that produce novel metaphoric extensions of familiar words (Bowdle & Gentner, 2005). We believe this could be a domain-general process.

5. General discussion

The purpose of this paper was to provide evidence for the idea of role-governed concepts (Markman & Stilwell, 2001). These studies tested and supported four predictions from this account:

1. Role-governed categories are defined by their place in a larger knowledge structure in which they take part. As a result, most of what people list about role-governed categories reflects relationships between that object and other elements of the relational system. Thus, people are more likely to list extrinsic properties when considering members of role-governed categories than when considering members of feature-based categories.
2. The knowledge structures in which role-governed categories are embedded are properties of a situation not properties of an object. Thus, one can only view an object as a member of a role-governed category by knowing about or assuming that situation.
   We tested this prediction using the website flickr.com. People who posted a picture, and thus knew the situation in which the picture was taken were far more likely to label the picture using a role-governed category than were those who simply saw the picture.
3. Role-governed categories are defined by their place in a relational system. As a result, properties that make an object ideal for fulfilling that role are considered to be illustrative of role-governed categories. In contrast, categories that are typical for members of feature-based categories are judged to be illustrative of feature-based categories.
4. Role-governed categories label particular roles in a relational structure. Thus, creating a new relational structure opens up the possibility for new role-governed categories. We found that when people acquire a new verb (which labels a relational structure), they are immediately able to acquire labels for the roles within that relational structure. Learning a novel adjective (which labels a property rather than a relational structure) does not ease the acquisition of a novel role-governed category relating to it.

In the rest of this paper, we examine the ubiquity of relational knowledge, and how the present work strives towards unifying the study of language and categories.

5.1. The ubiquity of relational knowledge

The dominant view of categories has been a feature-based view. From this, one might conclude that categories that name relations and relational roles form a relatively small subset of people's natural categories, however this is far from being accurate. In fact, informal ratings of the 100 most common nouns by Gentner and Kurtz (2005) suggest that half are feature-based categories and half are relational. Furthermore, verbs and prepositions are (nearly) all relational. Thus, there are a significant number of relational and role-governed concepts among our natural categories.

While many of our concepts are relational, the ubiquity of relational knowledge even spreads to feature-based categories. We are able to use feature-based categories in analogies, and as arguments for verbs in sentences, showing they are frequently parts of relational representations. In addition, when different feature-based categories share the same role across situations, their perceived similarity increases (Jones & Love, 2007). As discussed above, we understand that members of a feature-based category, like *dog*, can play a role for us, like *pet*. It seems reasonable to assume, that when a feature-based category consistently plays the same role like *dog* and *pet*, relational knowledge will become a part of its representation and affect how the category behaves. On the flip side, role-governed categories may start to behave as feature-based if the role is fulfilled consistently with the same features, as appears to be the case with many common artifacts invented for their extrinsic properties, i.e. their functions.[9]

Relational knowledge is also crucial for acquiring expertise in most domains. For example, expert physicists differ from novices in their ability to classify problems on the basis of relational similarities across items rather than based on featural (Chi, Feltovich, & Glaser, 1981). If our framework is correct, then part of gaining such expertise will be forming novel role-governed categories. Future work must investigate these processes.

We believe that the study of relational knowledge and how it interacts with featural knowledge should no longer

[9] Another way features become associated with role-governed categories are through the type-restrictions of their "root" relational structures. For example, a beverage has to be a liquid, because it is the object of drinking, which places restrictions on its role fillers.

take the back seat in the study categorization. The field is moving in this direction, investigating causal relations among features of category members (e.g. Sloman, Love, & Ahn, 1998), how relational information can cohere categories whose exemplars share no consistent sets of features (Rehder & Ross, 2001), showing that categories rooted in event script representations differ than feature-based categories (Ross & Murphy, 1999), and showing that creating relational structures on the fly allow for *ad hoc* categories (Barsalou, 1983, 1985). We suggest giving further prominence to relational information by investigating how categories can be defined by systems of extrinsic relations between entities. We think that a focus on relations is crucial for the progress of the study of categorization.

A broader point is that theories of the representation of categories and concepts will be intimately tied to the tasks with which we choose to study them (Markman & Ross, 2003). Most studies of category acquisition involve some form of classification task. When we use inductive classification to study categories built from unstructured and uncorrelated features, our models of categories represent them with unstructured and uncorrelated features (e.g. Nosofsky, 1986). When we use language comprehension including on-line sentence processing and analogy/metaphor understanding to study categories and concepts, our models will need to represent relational structures. When using inductive classification, concept learning often appears to take hundreds of solitary reinforced learning trials. When using language and communication, concept learning is shown as an extremely rapid and collaborative process that needs as little as a single exposure to the concept (e.g. current Experiment 3, Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Markman & Makin, 1998). Thus, laboratory studies must look at category use broadly in order to better understand the range of category types present in natural categories.

5.2. Open questions

This paper presents preliminary empirical studies of the differences between role-governed and feature-based categories. There are many aspects of role-governed categorization still to be explored. For example, we do not know what mechanisms may license novel role-governed categories in nonlinguistic contexts. More work must be done to understand the accumulation of relational and featural information associated with a category over time and how that knowledge accretion affects category representation and processing.

In addition, more work must be done to distinguish role-governed categories from other category types such as schema-governed categories. This work must explore what makes schema-governed and role-governed categories cohere, and what sort of structure persists in long-term memory. This work will help us to address the question of what allows two things to be considered members of the same role-governed category. For example, a tripod and an AA sponsor, are both types of supports. What enables this common category to be recognized? Perhaps analogical reasoning processes are required to make this categorization (Gentner, 1983; Goldwater & Markman, in press).

In many ways, then, providing evidence for the distinction between role-governed categories and feature-based categories raises more questions than it answers. Because these category types do appear to be distinct, a substantial amount of additional research must be done to better understand how these types of categories are formed and used.

Appendix A. Ideal and typical characteristic sets used for Experiment 2

A.1. Role-governed categories

Author
Ideal: Good writer, Funny, Books aren't too long, Clear, Relatable
Typical: Middle-aged, Imaginative, Intellectual, Reclusive, Large Vocabulary

Customer
Ideal: Has a lot of money, Polite, Friendly, Nice, Knows what they want
Typical: Female, Impatient, Has money, Demanding, Buying something

Drug
Ideal: No side effects, Inexpensive, Not addicting, Legal, Not Impairing
Typical: Cures, Alters mind, Pain killer, Addicting, Helpful

Friend
Ideal: Understanding, Intelligent, Sympathetic, Able to keep secrets, Compatible
Typical: Nice, Helpful, Kind, Always there for you, Fun

Game
Ideal: Interesting, Challenging, Active, Changes frequently, Easy to learn
Typical: Competitive, Rules, Athletic, Winners, Losers

Gift
Ideal: Practical, Something recipient wants, Liked by recipient, Hand-made, Reusable
Typical: Wrapped, Card, Small, On special occasions, On birthdays

Guest:
Ideal: Clean, Courteous, Has manners, Fun, Unobtrusive
Typical: Family, Friend, Kind, Dressed up, Invited

Home
Ideal: Looks good, Good location, Big, Comfortable, Pool

(*continued on next page*)

Typical: Has a kitchen, Where family is, Roof, Windows, Made of brick

Hobby
Ideal: Inexpensive, Social, Intellectually stimulating, Exciting, Healthy
Typical: Time-consuming, Relaxing, Done Alone, Requires a skill, Done regularly

Job
Ideal: High paying, Flexible hours, Good boss, Good co-workers, Fun
Typical: Boring, Long hours, Low wages, Make money, Time consuming

Pet
Ideal: Playful, Easy to care for, Loyal, Obedient, Friendly
Typical: Furry, Loveable, Loving, Soft, Fun

Predator
Ideal: Smart, Strong, Agile, Cunning, Camouflage
Typical: Mean, Sharp teeth, Big, Has Claws, Hungry

A.2. Feature-based categories

Beer
Ideal: Tastes good, Inexpensive, Non-fattening, Healthy, Flavorful
Typical: Bottled, or Canned, Carbonated, Yellow, Alcoholic

Bicycle
Ideal: Comfortable, Fast, Looks Good, Cheap, Durable
Typical: Handlebars, Seat, Made of Metal, Pedals, Reflectors

Cell Phone
Ideal: Camera, Durable, Easy to use, Inexpensive, Has lots of memory
Typical: Screen, Key pad, Ringtones, Buttons, Flip phone

Chair
Ideal: Adjustable, Reclining, Wheels, Leather, Rolls around/has wheels
Typical: Wooden, Four legs, Backrest, Arms, Seat

Fridge
Ideal: Spacious, Energy efficient, Goes well with room, Looks good, Has a water dispenser
Typical: Has a freezer, Stores food, White, Has two doors, Has an automatic light inside

Knife
Ideal: Easy to hold, Comfortable, Cuts well, Durable, Light weight
Typical: Dangerous, Metal blade, Silver blade, Shiny, Black handle

Microwave
Ideal: Quiet, Fast, Energy efficient, Inexpensive, Easy to use
Typical: Black, Heats food, Buttons, Turning plate, Light inside

Shoes
Ideal: Look good, Inexpensive, Match everything, Cute, Last a long time
Typical: Laces, Rubber, White, Soles, Leather

Table
Ideal: Sturdy, Looks good, Strong, Adjustable, Big
Typical: Four legs, Flat top, Round, or Rectangular, Glass top

Television
Ideal: Clear reception, High Definition, Big, Light weight, Inexpensive
Typical: Black, Color picture, Square, Remote control, Screen

Truck
Ideal: Fuel efficient, Powerful, Carries a lot, Roomy, Strong
Typical: Four wheels, Bed, Gas-guzzling, Two-door, Loud

Website
Ideal: Easy to use, Looks good, Useful, Entertaining, Fast-loading
Typical: Links, Colorful, Pictures, Ads, Search

Appendix B. Sample critical passages for Experiment 3

B.1. Cake

Novel Verb: Beth works at a bakery in the town where she grew up. Every morning, she comes in at 5:00 and cakes the flour, sugar and butter. The citizens of the town appreciate her skill. Beth has even won an award for what she does. Her boss noticed that this caker worked extra hard to get a Christmas bonus.

Paraphrase: Beth works at a bakery in the town where she grew up. Every morning, she comes in at 5:00 and makes cakes out of flour, sugar and butter. The citizens of the town appreciate her skill. Beth has even won an award for what she does. Her boss noticed that this caker worked extra hard to get a Christmas bonus.

Adjective: Beth works at a bakery in the town where she grew up. Every morning, she comes in at 5:00 and begins her cakeful morning. The citizens of the town appreciate her skill. Beth has even won an award for what she does. Her boss noticed that this caker worked extra hard to get a Christmas bonus.

B.2. Tomato-sauce

Novel Verb: Ilene was walking home when she was taken by surprise and sprayed by a skunk. She took some strange advice and tomato-sauced herself clean in her bathtub. She at first tried shampoo and soap, but she still stunk. Ilene was desperate. The next day, the tomato-saucer called the plumber to unclog the drain.

Paraphrase: Ilene was walking home when she was taken by surprise and sprayed by a skunk. She took some strange advice and washed herself with tomato-sauce until she was clean in her bathtub. She at first tried shampoo and soap, but she still stunk. Ilene was desperate. The next day, the tomato-saucer called the plumber to unclog the drain.

Adjective: Ilene was walking home when she was taken by surprise and sprayed by a skunk. She took some strange advice and used a tomato-sauceful shower to wash herself clean. She at first tried shampoo and soap, but she still stunk. Ilene was desperate. The next day, the tomato-saucer called the plumber to unclog the drain.

B.3. Plastic-sheet

Novel Verb: Sheila wants to paint her apartment and is concerned about messing up her furniture. She decided to plastic-sheet the furniture. She remembered how her grandmother kept her furniture. Sheila had to go back to Home Depot. The next day, the plastic-sheeter called her grandma to say hello.

Paraphrase: Sheila wants to paint her apartment and is concerned about messing up her furniture. She decided to use plastic-sheet to cover the furniture. She remembered how her grandmother kept her furniture. Sheila had to go back to Home Depot. The next day, the plastic-sheeter called her grandma to say hello.

Adjective: Sheila wants to paint her apartment and is concerned about messing up her furniture. She decided to make her furniture plastic-sheetful. She remembered how her grandmother kept her furniture. Sheila had to go back to Home Depot. The next day, the plastic-sheeter called her grandma to say hello.

B.4. Glass

Novel Verb: Fred was sitting at home when an intruder broke the window. He picked up a shard and glassed the intruder in the throat. He wasn't sure what to do. Fred had to act quickly. After calming down, the glasser called the police to get some help.

Paraphrase: Fred was sitting at home when an intruder broke the window. He picked up a shard and used the glass to cut the intruder's throat. He wasn't sure what to do. Fred had to act quickly. After calming down, the glasser called the police to get some help.

Adjective: Fred was sitting at home when an intruder broke the window. He picked up a shard and made a glassful slice in the intruder's throat. He wasn't sure what to do. Fred had to act quickly. After calming down, the glasser called the police to get some help.

B.5. Crutch

Novel Verb: Lyn works in a cider mill and recently broke her leg. She puts apples right on the floor and crutches them over to her coworkers. She's been in a cast for two weeks. She had to figure out ways to do her job. Yesterday, the crutcher developed a technique to increase her accuracy.

Paraphrase: Lyn works in a cider mill and recently broke her leg. She puts apples right on the floor and uses the crutch to send them to her coworkers. She's been in a cast for two weeks. She had to figure out ways to do her job. Yesterday, the crutcher developed a technique to increase her accuracy.

Adjective: Lyn works in a cider mill and recently broke her leg. She puts apples right on the floor and makes crutchful passes to send them over to her coworkers. She's been in a cast for two weeks. She had to figure out ways to do her job. Yesterday, the crutcher developed a technique to increase her accuracy.

Appendix C. A sample of the questions from Experiment 3

1. What does the plastic-sheeter do?
A. Wraps people in plastic sheets
B. Covers the furniture with plastic sheets
C. Sleeps in a bed with plastic sheets

2. Who got sick at work?
A. Jack
B. Sheila
C. Kelly

3. What did the prisoner use to build a tunnel?
A. A shovel
B. A fork
C. A spoon

4. What does the crutcher do?
A. transfer things to others with a crutch
B. hits people with a crutch
C. gets things down from high places with a crutch

5. What is the boxer training for?
A. A title defense
B. A shot at the title
C. His first big fight

References

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361–416.
Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory and Cognition, 15*, 397–418.
Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*, 211–227.
Barsalou, L. W. (1985). Ideas, central tendency, and frequency of instantiation as determinants of graded structures in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629–654.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*, 193–216.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1482–1493.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*(4), 767–811.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic memory. *Psychological Review, 82*, 407–428.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32*, 643–658.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science, 23*, 371–414.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*, 516–547.

Gentner, D. (1975). Evidence for the psychological reality of semantic components: The verbs of possession. In D. A. Norman, D. E. Rumelhart, & the LNR Research Group (Eds.), *Explorations in cognition* (pp. 211–246). San Francisco: Freeman.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.). *Language development: Language, thought and culture* (Vol. 2, pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press.

Gentner, D., & Asmuth, J. (2008). Can relationality be distinguished from abstractness in noun mutability? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 863–868). Austin, TX: Cognitive Science Society.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (in press). Structure-mapping and relational language support children's learning of relational categories. *Child Development*.

Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the lab* (pp. 151–175). Washington, DC: APA.

Glushko, R. J., Maglio, P. P., Matlock, T., & Barsalou, L. W. (2008). Categorization in the wild. *Trends in Cognitive Sciences, 12*, 129–135.

Goldwater, M. B., & Markman, A. B. (in press). Categorizing entities by common role. *Psychonomic Bulletin & Review*.

Goldwater, M. B., Asmuth, J., & Gentner, D. (in preparation). Norms distinguishing schema-governed categories, role-governed categories, and feature-based categories.

Goldwater, M. B., Markman, A. B., Trujilli, L., & Schnyer, D. (in preparation). An ERP analysis of licensing novel role-governed categories.

Goldwater, M. B., & Markman, A. B. (2009). Constructional sources of implicit agents in sentence comprehension. *Cognitive Linguistics, 20*, 675–702.

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

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology, 55*, 196–231.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology General, 11*, 228–238.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*, 508–529.

Kay, P. (2005). Argument structure constructions and the argument–adjunct distinction. In M. Fried & H. C. Boas (Eds.), *Grammatical constructions – Back to the roots*. Amsterdam/Philadelphia: Benjamins.

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The *what's X doing Y?* construction. *Language, 75*, 1–33.

Keil, F. C. (1989). Spiders in the web of belief: The tangled relations between concepts and theories. *Mind and Language, 4*, 43–45.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205–225.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Levin, B., & Rappaport-Hovav, M. (2005). *Argument realization*. Cambridge, UK: Cambridge University Press.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3–28.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions and graded category structure among tree experts and novices. *Memory and Cognition, 28*, 41–50.

Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127*, 331–354.

Markman, A. B., & Ross, B. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592–613.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence, 13*, 329–358.

Mauner, G., & Koenig, J. P. (2000). Linguistic vs. conceptual sources of implicit agents in sentence comprehension. *Journal of Memory & Language, 43*, 110–234.

McKoon, G., & MacFarland, T. (2002). Event templates in the lexical representations of verbs. *Cognitive Psychology, 45*, 1–44.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12*, 137–176.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review, 85*, 207–238.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

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

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Rehder, B., & Ross, B. H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1261–1275.

Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What is typical about the typicality effect in category-based induction? *Memory and Cognition, 38*, 377–388.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*, 495–553.

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189–228.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*, 214–241.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language, 47*, 530–547.

Wechsler, S. (1995). *The semantic basis of argument structure: A study of the relationship between word meaning and syntax*. Stanford University: CSLI Publications.