

Categorizing entities by common role

Micah B. Goldwater · Arthur B. Markman

Published online: 8 February 2011
© Psychonomic Society, Inc. 2011

Abstract Many categories group together entities that play a common role across situations. For example, *guest* and *host* refer to complementary roles in *visiting* situations and, thus, are *role-governed categories* (A. B. Markman & Stilwell, *Journal of Experiment & Theoretical Artificial Intelligence*, 13, 329–358, 2001). However, categorizing an entity by role is one of many possible classification strategies. This article examines factors that promote role-governed categorization over thematic-relation-based categorization (Lin & Murphy, *Journal of Experimental Psychology: General*, 130, 3–28, 2001). In [Experiments 1a and 1b](#), we demonstrate that the use of novel category labels facilitates role-governed categorization. In [Experiments 2a and 2b](#), we demonstrate that analogical comparison facilitates role-governed categorization. In [Experiments 1b and 2b](#), we show that these facilitatory factors induce a general sensitivity to role information, as opposed to only promoting role-governed categorization on an item-by-item basis.

Keywords Categorization · Concepts · Analogy · Relational categories · Role-governed categories

Categorizing entities by common role

How do people judge two entities to be members of the same category? Most accounts have suggested that people categorize entities together when they share common

descriptive features—for example, having four legs and fur (e.g., Rosch, 1973). More recently, Lin and Murphy (2001) showed that people often prefer to categorize together entities that are thematically related—that is, when they co-occur or interact within situations (e.g., *dog* and *bone*). However, A. B. Markman and Stilwell (2001) discussed the fact that many categories group together entities that play a common role across situations. For example, *guest* and *host* name different roles of *visiting* situations; thus, they are *role-governed categories*. That is, the label *guest* classifies together people who play a common role across different examples of visiting situations and distinguishes these people from the members of the complimentary role-governed category *host*. In contrast, thematic relation categories group together all entities that co-occur in situations, ignoring differences in individual roles.

Some research has contrasted the representations of role-governed categories with those of other kinds of categories (Asmuth & Gentner, 2005; Gentner & Kurtz, 2005; Goldwater, Markman, & Stilwell, *in press*; Rein, Goldwater, & Markman, 2010). In the present experiments, we were interested in factors that promote role-governed categorization. We were particularly interested in the roles of labels and structural comparisons on the likelihood of classifying items together on the basis of a common role.

Categorizing two entities by common role requires (1) representing each entity by its role within its domain and (2) representing the relational commonality across domains. People have a tendency to focus on the features of objects, rather than on their relational information, all else being equal (A. B. Markman & Gentner, 1993), and so role-governed classification requires that people's attention be attracted to relational information.

To explore factors that promote attention to relational structure in classification, we adopted a straightforward two-alternative forced choice (2AFC) procedure. In our

M. B. Goldwater (✉)
Northwestern University,
Evanston, IL, USA
e-mail: micahbg@gmail.com

A. B. Markman
University of Texas at Austin,
Austin, TX, USA

experiments, participants were shown a target category (e.g., *bodyguard*). Then they were given one alternative that was thematically related (e.g., *celebrity*)—that is, another object that took part in the same contexts. The second alternative was an entity that played the same role in a different context (e.g., *force field*). In this example, both bodyguards and force fields are protectors (bodyguards protect celebrities; force fields protect spaceships), although they share minimal intrinsic features and occur in different contexts. Previous studies have suggested that children and adults often form thematic groupings in tasks like this (Lin & Murphy, 2001; E. M. Markman, 1989).

Factors that promote role-governed classification

Labels are crucial to categories. They signal category essences (Gelman & Heyman, 1999; Yamauchi & Markman, 2000), and they promote category coherence (Yamauchi, 2005). Category labels evoke a predisposition that category members will have qualities in common and that these qualities are interrelated and causally connected (Yamauchi, 2009). Labeling an entity's category increases the likelihood of attributing properties to that entity and increases the likelihood of projecting properties to and promoting abstraction across other entities given the same label (Yamauchi, 2005, 2009; Yamauchi & Yu, 2008).

However, all kinds of labels do not have such inferential potency. Labeling the properties of entities has a smaller influence on category coherence, property projection, and forming abstractions than does labeling the entity (Gelman & Heyman, 1999; Yamauchi, 2005; Yamauchi & Yu, 2008).

While category labels are clearly important, this importance alone does not suggest why they would preferentially select for role-governed categories over thematic categories. If labels promote abstraction, they might focus people on the thematic relatedness of the items. However, there are a few reasons to believe that category labels elicit a preference for role-governed categories.

Developmental research suggests that novel labels allow children to go beyond tempting thematic alternatives to classify on the basis of properties that ultimately allow children to create taxonomic categories (Brown, 1957; Imai, Gentner, & Uchida, 1994; E. M. Markman, 1989; E. M. Markman & Hutchinson, 1984). Taxonomic categories have category labels, while thematic categories tend not to. Like taxonomic categories, role-governed categories are often given labels. Therefore, knowledge of word use should promote a focus on common role information and, thus, should allow role-governed categories to overcome tempting thematic alternatives.

In addition, category labels often promote analogical comparison (Christie, 2010; Gentner, 2003; Loewenstein &

Gentner, 2005; Namy & Gentner, 2002; Rattermann & Gentner, 1998; Son, Doumas, & Goldstone, 2010; Yamauchi, 2009). When forming an analogy, people seek similarities in the relations that describe pairs of domains, even if the objects that take part in those relations are not identical (Gentner & Markman, 1997; Hummel & Holyoak, 1997). For example, atoms can be seen as similar to the solar system because both involve one object that revolves around a second, even though electrons do not look like planets and nuclei are not hot and gaseous like the sun. Analogical comparison is a process that aligns relational representations by putting entities into correspondence by role, ignoring featural dissimilarities, resulting in this perceived similarity.

Evidence that labels induce structural comparison has come from a variety of sources. For example, Rattermann and Gentner (1998), Loewenstein and Gentner (2005), and Christie (2010) showed how labels facilitate relational mappings in favor of mappings based on superficial similarities in young children. In adults, Son et al. (2010) demonstrated that relational labels promote analogical transfer of signal detection principles to novel problems.

Experiments 1a and 1b explored the influence of labels on the formation of role-governed categories. On each trial, people heard a novel label applied to a target object. Then they performed a 2AFC between a role-governed match and a thematic match. We were interested in whether a novel category label would increase sensitivity to the role-governed match over a control condition in which we presented the novel label as an adjective, rather than as a noun. Because adjectives describe or label the properties of objects, they should not increase people's likelihood of making the role-governed match to the same degree (Gelman & Heyman, 1999; Yamauchi, 2005; Yamauchi & Yu, 2008).

While we assumed that label extension tasks implicitly induced comparison of the items, **Experiments 2a and 2b** used similarity judgments to test directly whether analogical comparison was a mechanism underlying role-governed categorization. A. B. Markman and Gentner (1993) showed that rating the similarity of a pair induced analogical comparison. We expected that if people judged the similarity of a pair of examples of a role-governed category (e.g., *bodyguard* and *force field*), they would recognize that both played the common role of protection within their respective scenarios. Thus, similarity comparisons should have increased sensitivity to role-governed categorization.

In **Experiments 2a and 2b**, prior to making a 2AFC, participants judged the similarity of the target category to each of the two alternatives. In a control condition, participants rated the relative imageability of the items. Imageability acted as a control because these ratings should not have induced structural comparison. If similarity

comparisons increased sensitivity to the basis of role-governed categories, participants should have selected the role-governed alternate following similarity ratings more often than they did following imageability ratings.

Finally, there was one additional question of interest in these experiments. One possibility was that the presence of novel labels and similarity comparisons influenced judgments for specific items. A second possibility, though, was that we created a generalized sensitivity to role-governed categorization. To test this prospect, Experiments 1b and 2b repeated the label and similarity manipulations (respectively) in a transfer block with a new 2AFC for which no label or judgment was given. If people who were given labels or similarity comparisons continued to make role-governed category choices in the transfer block, this would suggest that the task induced a general sensitivity to role-governed categories.

Stimulus norming and pilot studies

We created a set of 20 triads consisting of a target category, an item that was thematically related to the target, and a second item that was a member of the same role-governed category as the target (see Table 1 for the complete set). These items were constructed so that any taxonomic categories that the role match shared with the target were

also shared by the thematic match. For example, if the target and role match were both humans, the thematic match was also a human. In a few items, the targets and thematic alternates were both humans, while the role-match was not. For these items, the similarities would bias against our predicted pattern of results.

As a further check on the stimulus set, we asked a group of 10 participants to list commonalities between the target and each of the other items. The purpose was to ensure that what was common between the target and role matches was role based, and not that they had more intrinsic features than the thematic match (see Barr & Caplan, 1987, for a discussion of intrinsic and extrinsic features). This check validated our intuitions. Role matches elicited an average of 0.70 intrinsic commonalities, while the thematic matches elicited 0.76 intrinsic commonalities. Role matches elicited slightly more extrinsic commonalities to the target ($M = 2.43$) than did the thematic match ($M = 2.18$). Neither of these differences was statistically reliable.

To establish that the thematic match was a tempting alternative in the 2AFC, we had a group of 30 participants select the alternative that best made a category with the target for each of the 20 pairs. Participants selected the thematic match on 80% of the trials, suggesting that it does not just make a good alternative, but is strongly preferred (at baseline).

Experiments 1a and 1b

As was discussed above, on each trial of this study, participants performed a series of 2AFC trials on which there was a role-governed alternate and a thematic alternate. In these experiments, the 2AFC was presented as a novel word extension. Participants were given either a label or a description framing for the trial. On label trials, participants were told the following: “The < target > is a *goppin*. Which of these other two is better called *goppin*?” On description trials, participants were told the following: “It’s a *goppin* < target >. Which of these other two is better called *goppin*?” If novel labels promoted attention to common roles, more role matches would be selected on label trials than on description trials.

In Experiment 1b, unlike in Experiment 1a, the label and description conditions were run between subjects, such that a single mode of response could potentially be induced and transferred. After 10 trials had presented a label or description query, there was a transfer block of 10 2AFC trials where the query was worded as it was in the pilot study that established the baseline preference for the thematic match: “What goes best with < target > to make a category?” This version of the experiment enabled us to assess whether the tendency to make role-governed

Table 1 Materials for pilot studies and Experiments 1–4

Target	Role Match	Thematic Match
Appetizer	Opening Act	Entrée
Bird’s Nest	House	Tree
Bodyguard	Force-field	Celebrity
Bouncer	Tollbooth	People in line outside a club
Coach	Manager	Athlete
Coffee	Red Bull	Cream
Customer	Patient	Salesman
Director	Author	Actor
Drawer	Wallet	Socks
General	CEO	Private
Hard Drive	Filing Cabinet	Electronic Files
Jar	Closet	Pickles
Lock	Security Guard	Key
Paralegal	Nurse	Lawyer
Police Chief	Admiral	Beat Cop
Quilt	Campfire	Bed
Real Estate Agent	Drug Dealer	House-Buyer
Teacher	Camp Counselor	Student
Tripod	Cast and Sling	Camera
Vacuum	Soap	Carpet

judgments was maintained even after the label manipulation was no longer present.

Method

Participants There were 27 participants in [Experiment 1a](#) and 96 in [Experiment 1b](#). All received course credit.

Materials The materials are presented in [Table 1](#), using the role and thematic alternates. In [Experiment 1](#), there were two novel words to be extended, *goppin* and *blicket*. These nonce words were randomly assigned to the label and description conditions for each participant. In [Experiment 1](#), all 10 items in the label condition were given one label, and all 10 items in the description condition were given the other. In the label condition, the trial query read

A < target > is a *goppin*. Which of these other two is better called *goppin*.

In the description condition, the trial query read

It's a *blicket* < target >. Which of these other two is better called *blicket*.

Ten trials contained descriptor extensions, and 10 contained label extensions. Items were randomly assigned to conditions between subjects. Description and label conditions were run within subjects. Trial order was randomly determined for each participant.

[Experiment 1b](#) was identical to [Experiment 1a](#), except that there were 10 nonce labels that were randomly assigned to 10 of the triads.

Procedure The participants were tested individually at computers using the experimental software E-Prime.

Participants were told:

In this experiment we are interested in how you think about everyday things in the world. Specifically, we are interested in what things should be described or named with the same words. In this experiment, you will read statements about everyday things that are described or named with a new word. Your job is to indicate what else should be named or described with this same new word.

On each trial, a query was presented with two alternatives marked with the letters A and B. The assignment of the role match and thematic match alternatives was assigned pseudorandomly, so that the role match was alternative A on half the trials and alternative B on the other half. Participants responded by pressing A or B on the keyboard. The entire experiment took about 5 min.

Results and discussion

As was predicted, in [Experiment 1a](#), the label condition produced a higher rate of selections of the role match ($M = .66$) than did the description condition ($M = .38$). This difference was significant both by participants, $t_1(26) = 4.46$, $p < .001$, $d = 0.86$, and by items, $t_2(19) = 9.53$, $p < .00001$, $d = 2.13$ (see [Table 2](#)). As a reminder, the pilot study showed that without novel labels, people selected the role match on .20 of the trials.

The first 10 trials of [Experiment 1b](#) yielded a similar pattern of data. Role matches were selected on a higher proportion of trials in the label condition ($M = .59$) than in the description condition ($M = .33$), $t_1(95) = 4.48$, $p < .0001$, $d = 0.45$, $t_2(19) = 6.81$, $p < .0001$, $d = 1.52$.

Of interest, this difference between conditions was also observed in the transfer block, although the overall rate of selection of the role match was lower. More role matches were selected in the transfer block in the label condition ($M = .35$) than in the description condition ($M = .20$), $t_1(95) = 2.73$, $p < .05$, $d = 0.49$, and $t_2(19) = 8.45$, $p < .0001$, $d = 1.89$.

These results support the proposal that category labels, more so than descriptions or property labels, promote the construction of categories (Gelman & Heyman, 1999; Yamauchi, 2005) and that category labels select for role-governed category formation over thematic-based categorization strategies. In addition, the data from [Experiment 1b](#) suggest that the label manipulation also made people generally more sensitive to role information, since the difference in the rate of selection of the role match persisted into the transfer block.

We suggest that labels promote role-governed categorization because of knowledge of language usage patterns and because labels induce analogical comparison (Gentner, 2003; Namy & Gentner, 2002; Yamauchi, 2009), which aligns elements on the basis of a common relational role. The role of comparisons was examined explicitly in the next experiments.

Experiments 2a and 2b

[Experiments 2a and 2b](#) examined whether analogical comparison is an underlying mechanism of role-governed

Table 2 Summary of results for Experiments 1a and 1b. Means and standard deviations of the proportion of role matches selected by items shown

	Experiment 1a	Experiment 1b Block 1	Experiment 1b Transfer
Label	.66 (.22)	.59 (.24)	.35 (.12)
Description	.38 (.18)	.33 (.23)	.20 (.11)

categorization. A. B. Markman and Gentner (1993) showed that rating pairs of items for their similarity induces a structural comparison of the items and increases sensitivity to relational commonalities of the pair. Building from this finding, we had half the participants in [Experiment 2a](#) rate the similarity of the target words to both the role match and thematic match before judging which of the two better formed a category with the target. To control for the possibility that making ratings influences category judgments, the other half of the participants rated the relative imageability of the target with each of the alternates before making the categorization judgment.

[Experiment 2b](#) was identical to [Experiment 2a](#), with the addition of a transfer block to test the persistence of the influence of similarity comparisons on sensitivity to role-governed categorization.

Method

Participants There were 30 participants in [Experiment 2a](#) and 40 in [Experiment 2b](#). All were University of Texas undergraduates who received course credit.

Materials The items used were the targets, role matches, and thematic alternates from [Table 1](#). These materials were used for three judgments: categorization, similarity, and imageability. The categorization judgment was queried identically to the pilot studies: “Which of these two better go with [target] to make a category?” Then, below, were two alternates, marked by A and B. On half of the trials, the role match was choice A, and on half, choice B.

For similarity ratings, the target and one of the alternates was presented. The query read, “How similar are < target > and < alternate > ?” The target and alternate were presented in large font on their own lines. There was a 7-point Likert scale, that ranged from *not similar* to *very similar*. For imageability ratings, two words were presented, the target and one alternate. The query read, “Which is easier to picture in your head: A < target > or B < alternate > ?” There was a 7 point Likert scale, that ranged from *A is easier* at 1 to *both are equal* at 4 and *B is easier* at 7. Half of the participants made similarity ratings, and half made imageability ratings.

In [Experiment 2b](#), participants rated either similarity or imageability for half of the items. Then they performed a transfer block without ratings for the other half of the items. We counterbalanced which items appeared in the rating and transfer blocks between subjects.

Procedure Participants were tested individually at computers using the experimental software E-Prime. Their instructions differed depending on whether they were in

the similarity or the imageability condition. Participants were told that we were interested in how they think about everyday things in the world—specifically, that we were interested in what things they found similar, what things were easy to picture in their heads, and, in addition, what things went together to make categories. For both conditions, participants were instructed on the use of the Likert scale. For the category judgments, we defined a category as “a set of things or people that share some commonalities—be it genetic makeup, functions, physical or perceptual characteristics, purposes, or behavioral predispositions” (the definition used by Lin & Murphy, 2001).

Each judgment trial consisted of three parts. First, participants made two judgments that were either similarity ratings, comparing the target with each of the two alternatives, or relative imageability ratings, also comparing the target with each of the two alternatives. Half of the trials presented the judgment comparing the target to the role match first, and half the thematic match first. The third judgment was the triad categorization judgment for all participants.

In [Experiment 2a](#), participants performed 20 judgment trials. In [Experiment 2b](#), participants performed 10 judgment trials and then 10 transfer trials with no judgments preceding them. The study took about 15 min. to complete.

Results and discussion

The similarity ratings to the thematic and role-governed alternatives were about the same in each condition in both [Experiment 2a](#) (thematic $M = 4.32$; role $M = 4.18$) and [2b](#) (thematic $M = 4.77$; role $M = 4.52$), all $|ts| < 1$, $ps > .3$.

If analogical comparison is a mechanism of role-governed categorization, participants should have selected the role match more often after making similarity ratings than after making imageability ratings. In [Experiment 2a](#), the role match was selected on a significantly higher proportion of the trials following a similarity judgment ($M = .41$) than following an imageability judgment ($M = .21$), $t_1(28) = 2.24$, $p < .05$, $d = 0.84$, and $t_2(19) = 4.18$, $p < .001$, $d = 0.93$ (see [Table 3](#)). The proportion of role match selections following image-

Table 3 Summary of results for [Experiments 2a and 2b](#). Means and standard deviations of the proportion of role matches selected by items shown

	Experiment 2a	Experiment 2b Block 1	Experiment 2b Transfer
Similarity	.41 (.19)	.43 (.21)	.39 (.23)
Imageability	.21 (.16)	.13 (.13)	.22 (.13)

ability ratings was about the same as that in the baseline from the pilot study.

Experiment 2b replicated the basic finding of Experiment 2a. We observed a higher proportion of role match selections following similarity judgments ($M = .43$) than following imageability judgments ($M = .13$), $t_1(38) = 3.85$, $p < .001$, $d = 1.25$, and $t_2(19) = 6.31$, $p < .0001$, $d = 1.41$.

The transfer block provided evidence that this difference persisted. The similarity comparison condition elicited role match selections on a higher proportion of trials ($M = .39$) than did the imageability rating condition ($M = .22$), $t_1(38) = 1.78$, $p = .08$, $d = 0.58$, and $t_2(19) = 4.28$, $p < .001$, $d = 0.96$ (see Fig. 4). It should not be surprising that the items analysis has a larger effect size than did the participants analysis. The items were matched across conditions, while the participants were not, reducing the power of the participant analysis. However, given the very large effect size of the items analysis, there is no reason to doubt the reliability of the similarity condition's advantage.

Across the four experiments, the effect sizes were larger by item analysis than by participants. This was due simply to the greater variability across participants than across items, not because the participants were distributed bimodally; that is, there were no distinct distributions of *role responders* and *thematic responders*.

The results of **Experiments 2a and 2b** supported the claim that one mechanism that promotes role-governed categorization is structural comparison. When participants rated the similarity of the items prior to making classification decisions, they were significantly more likely to make role-governed classifications than when they made an imageability rating, which does not require a comparison of the elements of the mental representation of the concepts. Finally, the results of the transfer block in **Experiment 2b** suggest that similarity comparisons, like labels, created a general sensitivity to role-governed categories that persisted beyond the specific items on which the judgments were made.

General discussion

These experiments explored factors that increase people's sensitivity to role-governed categories. In a simple 2AFC task that pit role-governed classification against thematic classification, people were strongly biased toward thematic classification. Thematic classification can rely on simple associative mechanisms, because both category members co-occur, whereas role-governed categorization requires noticing relational commonalities across contexts. These experiments demonstrate that both novel labels and similarity comparisons increase people's propensity to make role-governed classifications.

Experiments 1a and 1b extended previous work on learning novel labels. Research with children has suggested that labels promote attention to factors that are associated with taxonomic categorization, such as shape (Imai et al., 1994; E. M. Markman, 1989;). These experiments suggest that people believe that novel nouns are more likely to refer to a role-governed category than to a thematic category.

While the category label condition promoted role-governed categorization to a greater degree than did the description condition, the description condition also showed a greater degree of role-governed categorization (38% in **Experiment 1a**) than did the baseline preference established in the pilot (20%). In the studies examining the role of category labels in categorization (e.g., Gelman & Heyman, 1999; Yamauchi & Yu, 2008), the control to a category label has often been a property label, not the absence of a label. While property labels and descriptions of entities clearly lack the inferential potency of category labels, these results suggest that they still have some ability to promote role-governed categories (consistent with Yamauchi & Yu, 2008). Common labels invite comparison of the entities being labeled, promoting abstraction and generalization (Gentner, 2003). There is no reason to assume that property labels do not induce these processes to any degree.

While labels have an important role in categorization, this alone does not answer why they preferentially select for role-governed categories over thematic relation categories. Labels have also been shown to induce analogical comparison (e.g., Christie, 2010; Gentner & Rattermann, 1991; Loewenstein & Gentner, 2005; Namy & Gentner, 2002; Yamauchi, 2009). Analogical comparison aligns elements on the basis of a common role across domains. This facilitates role choices and also reduces the likelihood of within-domain thematic choices. This can explain the large boost in role selections for the category label condition over the baseline and can explain why there was even a sizable boost for the description condition.

Label extension tasks implicitly induce comparison, but **Experiments 2a and 2b** extended these initial studies by directly examining whether comparison supports role-governed categorization. They demonstrated that people can be encouraged to process information about the roles that objects play within situations by having them compare the concepts with the aim of rating their similarity. Previous work showed that similarity comparisons promote attention to information that relates objects to each other (A. B. Markman & Gentner, 1993). Attending to these relations also seemed to make people more sensitive to the relational roles played by objects.

These data suggest that the effect of labels on categorization is not just an effect of inducing comparison. Participants were much more likely to select the role option given a label (66% in **Experiment 1a**) than when making a

similarity comparison (41% in [Experiment 2a](#)). It could be that the label is a more reliable or more powerful inducer of structural alignment than is a similarity comparison. Alternatively, labels may induce other processes that also support role-governed categorization. While more research needs to be done comparing the two methods, recent work by [Christie \(2010\)](#) has shown that labels are also a more powerful inducer of structural alignment in 2- and 3-year-old children than are other forms of comparison.

[Experiments 1b](#) and [2b](#) also suggest that people can be induced into a mind-set in which they attend to role information. These experiments included blocks of transfer trials in which there were no labels or similarity comparisons. Those people who began to classify by roles in the first blocks of these experiments were also more likely to classify by role in the transfer blocks than were those who were not induced to classify by role. These results provide strong support for the framework laid out by [A. B. Markman and Stilwell \(2001\)](#). Inducing a general role-governed mode of responding implies that role-governed categorization employs different information or different processes than does thematic relation categorization.

It is important to acknowledge that the manipulations in these experiments influenced people's tendency to make role-governed classifications, but they did not drive everyone to make these classifications all the time. Thematic groupings are often important ways to match items. Thus, we were interested in factors that increase the rate of role-governed classification, relative to thematic grouping.

These experiments were the first to examine whether and how people judge two entities that play a common role across situations to be members of the same category (for similar work with children, see [Gentner, Anggoro, & Klibanoff, in press](#)). This work joins a growing body of evidence demonstrating the importance of role-governed categories in our conceptual system ([Gentner & Asmuth, 2008](#); [Gentner & Kurtz, 2005](#); [Goldwater et al., in press](#); [Jones & Love, 2007](#); [Jung & Hummel, 2009](#); [McRae, Ferretti, & Amyote, 1997](#); [Rein et al., 2010](#); [Ross & Murphy, 1999](#)). We should note that these experiments used verbal stimuli, and so the entities to be categorized were categories themselves. In fact, categories such as *bodyguard* and *force field* are also role governed. One could think of the categories formed in these experiments (e.g., *protector*) as more general role-governed categories. There is no reason to assume that the patterns identified in this article will not generalize to the learning of more specific role-governed categories, such as *bodyguard*, although this is an open empirical question. Important next steps are to examine how people categorize specific entities together by their common role across examples of novel event categories. We predict that category labels and analogical comparison will play a critical role in these settings as well.

Acknowledgments We thank Cathy Echols, Jeff Loewenstein, David Schyner, Zeni Griffin, and three anonymous reviewers for valuable discussion of this work. In addition, we acknowledge a fellowship from the IC2 Institute given to A.B.M.

References

- Asmuth, J. A., & Gentner, D. (2005). Context sensitivity of relational nouns. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual meeting of the cognitive science society* (pp. 163–168). Mahwah: Erlbaum.
- Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition, 15*, 397–418.
- Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55*, 1–5.
- Christie, S. (2010). *How simple is same: Perception of relational similarity in young children* (Doctoral dissertation). Available from Dissertations and Theses database. (UMI No. 3402146)
- Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*, 489–493.
- Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge: MIT Press.
- 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., & 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: Cognitive Science Society.
- 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: American Psychological Association.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *The American Psychologist, 52*, 45–56.
- Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London: Cambridge University Press.
- Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (in press). The empirical case for role-governed categories. *Cognition*.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427–466.
- Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45–75.
- Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology, 55*, 196–231.
- Jung, W., & Hummel, J. E. (2009). Learning probabilistic relational categories. In B. Kokinov, K. Holyoak, & D. Gentner (Eds.), *New frontiers in analogy research: Proceedings of the second international conference on analogy*. Bulgaria: Sofia.
- Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3–28.
- Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315–353.

- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge: MIT Press.
- Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*, 1–27.
- Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.
- Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*, 329–358.
- McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176.
- Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*, 5–15.
- Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 274–282). Sophia: New Bulgarian University.
- Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What's typical about the typicality effect in category-based induction? *Memory & 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.
- Son, J. Y., Doumas, L. A., & Goldstone, R. L. (2010). When do words promote analogical transfer? *Journal of Problem Solving*, *3*, 52–92.
- Yamauchi, T. (2005). Labeling bias and categorical induction: Generative aspects of category information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 538–553.
- Yamauchi, T. (2009). Finding abstract commonalities of category members. *Journal of Experimental & Theoretical Artificial Intelligence*, *21*, 155–180.
- Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 776–795.
- Yamauchi, T., & Yu, N. (2008). Category labels versus feature labels: Category labels polarize inferential predictions. *Memory & Cognition*, *36*, 544–553.