Learning of Role-Governed and Thematic Categories

Micah B. Goldwater

University of Sydney

Rebecca Bainbridge Gregory L. Murphy

New York University

Please address correspondence to:

Micah Goldwater
School of Psychology
Brennan MacCallum (A18)
University of Sydney, NSW 2006
Australia

micah.goldwater@sydney.edu.au

Acknowledgements

The research was supported in part by NSF grant BCS 1128769. We are grateful to Arthur

Markman and Mitchell Rabinowitz for helpful comments

Abstract

Natural categories are often based on intrinsic characteristics, such as shared features, but they can also be based on extrinsic relationships to items outside the categories. Examples of relational categories include items that share a *thematic relation* or items that share a common *role*. Five experiments used an artificial category learning paradigm to investigate whether people can learn role-governed and thematic categories without explicit instruction or linguistic support. Participants viewed film clips in which objects engaged in similar actions and then were asked to group together objects that they believed were in the same category. Experiments 1 and 2 demonstrated that while people spontaneously grouped items using both role-governed and thematic relations, when forced to choose between the two, most preferred role-governed categories. In Experiment 3, category labels increased this preference. Experiment 4 found that people failed to group items based on more abstract role relations when the specific relations differed (e.g., objects that prevented different actions). However, Experiment 5 showed that people could identify them with the aid of comparison. We concluded that people can form role-governed categories even with minimal perceptual and linguistic cues.

The psychology of concepts has primarily studied simple object categories, e.g., chairs and dogs, or their even simpler experimental stand-ins: colored shapes, dot patterns, or schematic animals (e.g., Posner & Keele, 1968; Rosch, 1973). Such categories are generally based on descriptive features *intrinsic* to category members, e.g., dogs have fur and four limbs, and that are organized into taxonomic classification hierarchies, e.g., dog, mammal, animal.

Recent research has given increasing attention to relational categories, whose membership is determined by *extrinsic* relationships among items (Gentner, Angorro, & Klibanoff, 2011; Gentner & Kurtz, 2005; Goldwater & Markman, 2011; Goldwater, Markman, & Stilwell, 2011; Kittur, Hummel & Holyoak, 2006; Markman & Stilwell, 2001; Rein, Goldwater, & Markman, 2010; Ross & Murphy, 1999; Tomlinson & Love, 2010). There are many kinds of relational categories, but they are similar in their dependence on context and extrinsic relations. For example, *thematic* categories contain objects that have a specific functional relationship, such as soup and bowl, or honey and bee. Here the items in the same category are not necessarily similar but have a complementary relationship: The soup is eaten out of the bowl; the bee makes the honey (Estes, Golonka, & Jones, 2011; Lewis, Poeppel, & Murphy, 2015; Lin & Murphy, 2001; Markman, 1989). *Script-based* categories group together objects that play the same role in a familiar event structure. For example, desserts are all eaten at the end of the meal, as a treat; greetings are the things we say at the beginning of an interaction (Nelson, 1986; Nguyen & Murphy, 2003; Ross & Murphy, 1999).

Similar to taxonomic categories, some of these relational categories also have overlapping intrinsic features. For example, desserts are usually but not always sweet. However, category membership is more strongly determined by the extrinsic relational properties. A cookie eaten before the main course is not considered dessert, and cheese or an unsweetened biscuit eaten at

the end of the meal can be considered dessert. Someone who only attended to the intrinsic properties of cookies or pancakes would not be able to identify when they occurred as desserts.

The present article focuses on one type of relational category, *role-governed* categories (Goldwater et al., 2011; Markman & Stilwell, 2001). In such categories, membership is determined by the kind of role that the object plays relative to others. For example, a guest is a person who stays in someone else's home; a teacher is someone who instructs someone else; a barrier is something that prevents access to a goal. Although the members of these categories can be similar, the critical determinant of category membership is the exemplar's role in some larger event or structure. A wealthy, domineering person who orders everyone around is not a CEO unless she actually has the proper role relative to others within a corporate structure. Many homes are houses, but a house no one lives in is not a home. Goldwater et al. (2011) empirically demonstrated the link between role-governed categories and relational structures and differentiated them from taxonomic feature-based categories along a number of dimensions (see also Gentner & Kurtz, 2005; Rein et al., 2010). However, the initial formation of role-governed categories has been studied to a much lesser degree (see Gentner et al., 2011; Goldwater & Markman, 2011, and Jones & Love, 2007, for how common roles affect categorization and similarity of familiar items). The initial formation of role-governed categories is the focus of the current work.

It will be important for our purposes to distinguish such role-governed categories from thematic categories. Thematic categories include two or more kinds of objects that are connected by a specific relation, such as leash-and-dog. Note that these two objects do not have a *common* role, but instead play *complementary* roles to one another: The leash holds the dog; the dog is held by the leash. In contrast, objects in role-governed categories have the same roles in

relationship to other objects not in the category. For example, all barriers impede access to a goal; but goals are not in the barrier category.

This example points out that there is often a close relationship between thematic and role-governed categories: One half of a thematic category often forms a role-governed category with similar items. For example, a teacher and student might form a thematic pair. Teachers themselves form a role-governed category, because they all share the same relationship to their students. Students form a role-governed category, because they all share the same relationship to their teachers. Similarly, a barrier and goal are linked thematically, and they are also each members of different role-governed categories. This linkage between the two types of categories does not always hold. Although leashes and dogs might form a thematic category, dogs cannot be said to form a role-governed category, because they are not primarily defined in terms of their relationship to other things.

At this point, some readers may be wondering whether all these groupings really are categories. Not only do they group things that are not at all similar, like leash and dog, they include a diverse set of relationships as the bases for classification—thematic relations, scripts, roles, and similarity. As discussed by Murphy (2010), identification of these kinds of categories is not just a theoretical claim but an empirical observation of the kinds of classifications that people actually make. People group items together in these ways when asked to form categories or to group together the same kinds of things (Estes et al., 2011; Gentner & Brem, 1999; Goldwater & Markman, 2011; Lin & Murphy, 2001; Ross & Murphy, 1999). They can influence memory performance (Rabinowitz & Mandler, 1983). Furthermore, relational categories provide a basis for category-based inferences (Rein et al., 2010; Ross & Murphy, 1999). Thus, such groupings share many of the functions of familiar taxonomic categories.

Another important connection to role-based categories is the creation or comprehension of analogies. Analogical inference is based on shared relational structure (Gentner, 1983; Holyoak & Thagard, 1989). The classic example of comparing the structure of the atom to the solar system works not because the sun and its planets are similar to the parts of the atom, but because the pattern of relationships between the sun and its planets is similar to that of the relations between an atom's nucleus and its electrons. However, analogies are not trivial to notice on one's own, and much research has focused on why analogical reasoning can be so difficult (e.g., Gick & Holyoak, 1980; Keane, 1997; Rottman, Gentner & Goldwater, 2012). This difficulty suggests in turn that role-based categories may be difficult to identify as well. We explore this possibility in the next section.

Learning of Role-Governed Categories

Much early work on role-governed categories and relational categories more generally has focused on distinguishing familiar role-governed categories from familiar taxonomic categories. Exemplars of role categories such as barrier are less similar to each other, e.g., wall and poverty, than are exemplars of taxonomic categories such as fruit (Gentner & Kurtz, 2005). Role-governed categories such as friends elicited extrinsic properties about how category members relate to others, like "they are always there for you," while taxonomic categories such as knife, elicited properties intrinsic to categories members, like "made of metal" (Goldwater et al., 2011). In addition, relational categories' meanings are more contextually mutable (Gentner & Asmuth, 2008).

These properties of role-governed categories seem likely to make them harder to learn than more familiar taxonomic categories. Just as people find it difficult to use the solution from one problem to solve another problem that is relationally identical but superficially different (Gick &

Holyoak, 1980), they may find it difficult to identify a common role like *guest*. If one did not already know this word, would one identify people as being guests across all the different settings in which it is used? Would they identify houseguests, hotel guests (who pay for their services), people invited to a restaurant for dinner (who eat for free), guests at a wedding (who are not served food), or perhaps even a dragonfly visiting a garden (an attested example from Goldwater et al., 2011) all as being the same kind of entity, abstracting across their ages, genders, social classes, and even species? Could they identify the common role across very different events? In contrast to thematic categories, which often involve associated objects, identifying role-governed categories can require considerable abstraction. The fact that we already know the word *guest* makes the similarity of these different entities easy to detect, but we will shortly review literature that suggests that detection of novel role-governed categories is not so easy. That is, perhaps exposure to the word *guest* is important for our learning that category.

The category of *guest* co-occurs with another role-governed category, *host*, and the two form a thematic category. Much research now shows that many people will classify thematic pairs such as *guests* and *hosts* as being in the same category or as “the same kind of thing” (Estes et al., 2011; Goldwater & Markman, 2011; Lin & Murphy, 2001; Murphy, 2001). In light of that, it is not clear that people will reliably detect role-governed categories themselves in the absence of a linguistic label or an explicit search for commonalities. That is, they might think that a *host* and his *guest* go together but, prior to learning the word *host*, may not perceive that two different *hosts* are the same kind of person. The *host* bears a direct relationship to his or her *guest* but not to a different *host*.

Indeed, relational categories are not always correctly identified as such. For example,

consider Keil and Batterman's finding (see Keil, 1989, ch. 4) that children misidentify the relational term *uncle* to be a feature-based category (a nice man who comes to visit and gives you presents) until they are in 4th grade. Clearly, these common features are easier to acquire than the familial relation that is the true basis of the word. Gentner and Kurtz (2005) point out that relational words are generally learned later, that it takes longer to produce examples of relational categories, and that people who view such examples (barriers, filters, guides, etc.) often have difficulty identifying what they have in common. Such results raise the question of whether people can identify and use relational categories as they do taxonomic categories, without instruction. Indeed, after considering this question, Gentner and Kurtz (p. 165) propose that learning relational categories is at least in part "through cultural guidance: in particular, through linguistic labels." Gentner et al. (2011) showed that 4-5-year-old children identify relational categories based on a single example much better if they are given linguistic support, for example, "this knife is the *dax* for the watermelon." Such sentences not only label the category but use relational language that makes the nature of the category clear.

Another aid to role-governed category learning is comparison of category exemplars. Goldwater and Markman (2011), using linguistic stimuli, showed that without the support of a category label or explicit comparison, people were four times more likely to group thematic pairs, such as bodyguard and celebrity, than they were to group things that played the same role across domains, such as bodyguard and force-field (both are protectors) (see also Gentner et al., 2011). These results are similar to, and perhaps have the same psychological basis, as the finding that analogies are difficult to notice without explicit instruction (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Goldwater & Gentner, 2015).

In sum, it is not clear just how well people can identify and learn these relational categories

under neutral conditions without support from language, a teacher, or a helpful setting. We present a series of studies examining under what conditions people spontaneously form novel role-governed categories from viewing novel events. We contrasted such categories with novel thematic categories, which are more directly associated. To construct novel categories of both types, we created events in which different kinds of objects interacted. Participants viewed films of objects that engaged in similar actions across the films. Some of the objects could be grouped into thematic pairs such as one chasing the other. However, across films, objects could be grouped into role-governed categories such as the chasees. We tested each kind of category independently (following Nguyen & Murphy, 2003) to identify whether people could identify role-governed categories, thematic categories, or both, from the same events. We also included conflict trials, in which one type of category had to be selected over the other, to identify which kind of category is more salient.

As mentioned earlier, role-governed categories often share attributes, which makes sense given that the ability to carry out a role may require certain properties. For example, teachers are usually adults with a college education; barriers are often bulky. Across experiments, we varied the similarity of the members of the role-governed categories. Because the paradigm required people to remember a number of novel objects and what they did in the films, there were always some perceptual cues to allow participants to remember which kinds of objects did what (we discuss the role of perceptual features in relational categories in the discussion). Since thematic categories involve grouping different kinds of objects, these categories did not generally have as much perceptual similarity, just as leash and dog or soup and bowl are not similar. However, these pairs had consistent relationships across the films. For example, a green object might always chase a red object, and participants who attended to the thematic relations could then

group green and red objects into thematic pairs at test.

Experiment 1

In order to discover whether people identify and use novel role-governed categories to classify objects, we constructed film clips in which objects interacted. In each clip, there were two objects that interacted in meaningful ways. In one set of three clips one object crushed the other one; in another set, one object chased another one; in the third, one object lifted another. (These descriptions are shorthands for the specific events that occurred in the clips. The actual actions had distinctive components and motions not included in the neutral meanings of *chase*, *crush*, or *lift*.) The third objects in the clips were *bystanders* that moved around but did not engage in any specific goal-related action. In the classification test, bystanders served as comparison items that could be classified simply because they appeared in the same clip as the target object, and so provided a meaningful response even though they did not represent role or thematic choices. We picked relations that would be readily identifiable from a short film but that did not refer to familiar role-governed categories. That is, no one spontaneously identifies objects as things that are chased or things that crush when they are not participating in such events. People do not classify a cardboard box and a baby as the same thing because both may be lifted. Nor would they out of context categorize a heavy stone and a truck as both things that could crush something else. So, although the events of chasing, crushing, and lifting are themselves familiar enough, they do not form the basis for familiar role-governed categories that are used to classify objects.

The fact that one can add *-er* or *-ee* to the ends of verbs may give the impression that such role-governed categories are trivial to create and identify. After all, if one knows the verb *lift*, one can interpret the word *lifter*. However, this productive use of language is not the same as

actually noticing categories when no such verb is used, as in our studies. Indeed, these suffixes can be added even when one does not know the category. If you hear that Stephanie zonged Jen, you can infer that Stephanie is a zonger and Jen a zongee, even though you have no idea what those categories are. Finally, it should be mentioned that some participant utterly failed to identify role-governed categories in our experiments, so the fact that words like *chaser* exist in English is in fact not sufficient for people to group objects together based on this role.

The two critical objects in each clip represented different roles. As pairs, they represented thematic groupings: crusher-crushee, chaser-chasee, or lifter-liftee. But they could also be grouped by role (across clips): all three crushers or all three chasees. After viewing the clips, we asked people to classify the items using the classic *triad task* used in much prior research on thematic vs. taxonomic categorization in both children and adults (e.g., Gentner & Brem, 1999; Lin & Murphy, 2001; Markman & Hutchinson, 1984; Nguyen & Murphy, 2003; Smiley & Brown, 1979, among many others). In this task, one item was the target, and two other items served as the choices. Participants chose which of the two went together best with the target to make a category.

The experiment asked three related questions about how these items would be classified. First, would people identify two objects that played the same role, for example, two liftees, as being the same kind of thing? Second, would people identify the thematic pairs as being the same kind of thing? These questions were answered by test items that tested for these two relations compared to grouping the target with a bystander. The third question was which of these conceptual groupings is stronger—having the same role or being thematically related? This was tested in a *conflict* trial, in which participants had to choose between two items that had one of these relations (same role or thematic relation) to the target. For example, with a crusher as the

target, the two choices would be another crusher or a crushee. Participants might well identify both of these relations as being important in the trials that measured each one separately, but they could also identify one of them as being much more salient than the other. Together, these tests allowed us to independently test each relation as well as their relative strengths.

Method

Participants. Sixteen New York University students participated in the study for course credit.

Materials. We created nine movies in Adobe Flash to demonstrate the various roles of items in three different types of events. Each event consisted of an object (an agent) acting upon another object (a patient) and a third object moving around on its own (a bystander). The actions in the three events were chasing, crushing, and lifting. Each type of event had three different film clips with different objects performing the same type of action. The objects that performed similar roles looked similar. Figure 1 shows the objects from the chase event. The chasers in the top row of the figure all shared a similar shape and size, but varied in other features. The same is true for the chasees and the bystanders. (We altered the visual similarity in later experiments.) The objects in the same clip were the same color, which was different across the nine clips: red, green, blue, orange, pink, purple, yellow, black, and brown. There were three different roles for each event (counting the bystanders), three objects in each role for each event (across film clips), and three different types of events, making a total of 27 objects overall.

Figure 1. Objects from the chase movies in Experiment 1. The objects on the top row were the chasers, the objects on the second row were the chasees, and the objects on the bottom row were the bystanders. The objects with the same role all shared a similar shape and size, but varied in their features and color. Objects in the same column appeared in the same film clip and were the same color.

Event norming. We conducted a pilot study to evaluate the event animations for this and the rest of the experiments (five in total) to discover whether watching a single example of each event category was sufficient for subjects to identify the relational commonality that defined

each category. Ten participants were asked questions about nine animations each. They were asked, “Have you seen this kind of event before?” and how they would label an individual participant in each event (it was always one of the actors in the event, not the bystander). Out of ten pilot participants, only two focused on the events’ relational structure at all. The primary descriptors for the event participants were names focusing on their appearance, e.g., “the alien vacuum cleaner” or “the blue haired woman.” The two who did focus on the relations among the actors in the events did not consistently pick out the category defining relation. Instead, they offered descriptions such as “those two are friends, but they are excluding the other one” (“the other one” being the bystander). By making statements such as “those two are friends” they were focusing on a mutual symmetric relationship that would not distinguish roles from each other. In addition, only a single participant stated that any of the events were familiar. She said a couple of the events (from Experiment 4 and 5) “were like Pong” (the classic video game). Given this pilot study, we are confident that the pattern of data across these experiments reflects learning. That is, the categorization judgments made throughout the experiments rely on finding generalizations across multiple exemplars and not simply identifying familiar kinds of events.

Procedure. Participants were told that they would be watching several movie clips in which various objects would be performing several actions. They were encouraged to pay close attention to the objects and their actions, as they would be asked about them later. They watched the movies in two identical blocks of all nine clips, with a break in between. Half of the participants saw the movies in the order chase, crush, lift, chase, crush, lift, etc., and the other half saw them in the reverse order.

After participants watched the movies, they performed a classification test. In this test, three of the objects from the movies were arranged in a triangle on the computer screen. The object on

the top was the target, and participants were asked to choose which of the two other objects went best with that object to form a category. The instructions defined a category as “a set of things or people that share some commonalities—be it genetic make up, functions, purposes, physical and perceptual characteristics, or behavioral predispositions,” from Lin and Murphy’s (2001) study. We have no reason to think these instructions would bias participants towards one kind of category or the other because Lin and Murphy (2001) found that participants were quite capable of choosing both taxonomic and thematic choices (and their preferences changed with other task factors). That is, these instructions did not appear to prevent any kind of categorization of interest. Further, our participants were told that there were no right or wrong answers and to choose whichever choice made sense to them. They responded by pressing either the Q or P on the keyboard for the left or right choice.

There were four types of test trials. The first two types tested whether participants learned role and thematic relations. In *role triads*, the target object and its match performed the same role in the movies (e.g., both chasers) and the third item was a bystander from the same type of movie (but not the same movie as the target). In *thematic triads*, the target object was either an agent or patient, its match was its corresponding patient or agent (e.g., a chasee and chaser), and the third item was again a bystander from the same type of movie. Figure 2 shows the third type of triad, *conflict triads*, which pitted these two types of categories against each other to see which relationship was stronger. The target was either an agent or patient from one of the movies and the two choices were an object that played the same role from a different movie (role match) and the type of object that it interacted with (thematic match), also from a different movie. The fourth type of test trial was *fillers*. The target was a bystander and the two choices were an object from the same type of movie and an object from a different type of movie. These trials were included

so that bystanders would sometimes be targets; they did not enter into the theoretical predictions.

Figure 2. A conflict triad. The target was the chasee at the top, its role match was the chasee on the right, and its thematic match was the chaser on the left.

In all four triad types, the items were from different clips, so they could only be classified by kind—not by direct association of appearing together. There were 21 test trials, two of each triad type per event type, except for the fillers, which only had one per event type. The triads were presented in a random order, and the order of the choices was determined randomly by the

computer.

Results and Discussion

Overall, participants answered role and thematic trials by choosing the correct match 75% and 79% of the time, respectively, suggesting that they spontaneously extracted both relationships, about equally often. Participants on average chose the role match in conflict trials 62% of the time, which was reliably greater than chance, $t(15) = 5.98$, $p < .001$. (In filler trials, participants chose the item from the same event 83% of the time.) Closer examination revealed that participants either chose primarily role matches or primarily thematic matches in conflict trials. Ten participants chose role matches two-thirds or more of the time, five participants chose primarily thematic matches, and one participant was evenly split.

This distinction between thematic and role-responders does not reflect random responding. In this and all following experiments, participants had six conflict trials and so could either be evenly split between the two (thematic and role choices) or have at least four out of six favoring one or the other. If participants were responding randomly, then being evenly split would be the most likely outcome (31% of responses according to the binomial theorem). However, across all experiments just 6 out of 126 (under 5%) were evenly split, suggesting that people are much more likely to show a preference either for thematic categories or role categories than to have no preference.

As seen in Table 1, role responders chose role matches on role trials 95% of the time, and were quite accurate in thematic trials as well, choosing the match 75% of the time. Thematic responders chose thematic matches on thematic trials 83% of the time, but were strikingly inaccurate on the role trials, choosing the role match only 30% of the time (and a bystander the majority of the time). That is, role responders detected thematic relations but thematic responders

did not detect (or use) role commonality. We withhold discussion on this point until it is replicated, but this reinforces the point that the preferences are not simply random variation.

Table 1

Experiment 1. Mean Percentages of Correct Response by Participant Group and Triad Type

Triad Type	<i>Role Responders (n = 10)</i>	<i>Thematic Responders (n = 5)</i>	<i>No preference (n = 1)</i>	<i>Overall M</i>
Conflict (role responses)	91	6	50	62
Role	95	30	100	75
Thematic	75	83	100	79
Filler	83	80	100	83

Note. Each entry represents the percentage of choices of the stated type of that triad, e.g., of thematic choices in the thematic triads.

Overall, the results show that people are sensitive to both kinds of relations. This gives the first experimental evidence that people identify and use novel role-governed categories without instruction or linguistic support. However, the test for identifying these relations was somewhat weak in Experiment 1. For example, in the role trials, the two options were between something that played the same role and a bystander that had no strong relationship to the target item. Because the same-role item had to come from a different clip, we used a bystander that also came from a different clip as the other choice. The same was true of the thematic trials, in which both items came from a different clip from that of the target. As a result, the bystander had no specific relationship with the target. In the next experiment, we used a bystander that came from the same clip as that of the target, thereby providing a stronger test of how important the two relations are perceived to be.

Overall, there seemed to be a preference for role relations over thematic relations. People

were more likely to choose the role item over the thematic item in the conflict tests, but this proportion is misleading, because it averages two groups of participants who consistently chose one or the other. Nonetheless, there were twice as many people who predominately selected roles over thematic relations. Experiment 2 investigates a possible basis of this difference.

Experiment 2

Participants clearly identified the roles and thematic relations in Experiment 1 and used them to classify the objects. There was a general preference for role responses, but it is not clear whether this is due to a pure preference for roles over thematic relations, as the role-governed categories shared common shapes as well as common roles. This is probably naturalistic, as things that do the same actions and engage in the same relations must have common parts and other properties in order to accomplish those shared actions (Tversky & Hemenway, 1984). For example, a Jersey barrier and a fallen tree are both barriers, and both are long, some feet high, and very heavy. However, such stimulus structure also raises the question of whether the results are primarily due to the role itself or to the shared shapes. It has long been known that shared shapes are particularly important in determining category membership (e.g., Markman's, 1989, shape bias; Landau, Smith, & Jones, 1988; Tversky & Hemenway, 1984). It may well be that both factors—common role and common shape—influenced the groupings in Experiment 1. Even if the two often co-occur in natural categories, it is important to know to what degree the roles themselves provide a basis for classification.

In Experiment 2, we constructed new stimuli to take part in the same events as in Experiment 1. Here, all the objects had different shapes. Objects that carried out the same role had the same color, and objects from the same film clip had the same texture. This allowed participants to identify an object's role by virtue of its perceptual attributes, but it removed any effect of same-

shape bias. (Surface properties like color and texture are generally dispreferred bases for classification; Landau et al., 1988.) Furthermore, on the role test trials, the role match shared color and the bystander item shared texture with the test item. Thus, mere feature matching would not lead to above-chance same-role selection. Will people still identify the objects' roles and use them to make role-governed and thematic categories in such circumstances? The thematic classifications of Experiment 1 already grouped together items that did not share shape, so the new stimuli would certainly not hurt thematic grouping and might increase it.

Method

Participants. Twenty-eight New York University students participated in this study for course credit.

Materials. We created new items, which consisted of 27 different shapes with varying colors and textures (see Figure 3). Nine new movies were created using these shapes and the events from the previous experiment. Objects with the same role in these clips had the same color, and objects from the same movie had the same texture. However, objects playing the same role no longer had the same shape. Table 2 (top panel) describes this structure for the chasing event. Note that each object had a unique shape (S), but texture (T) and color (C) were shared.

Figure 3. Objects from the chase movies in Experiment 2. The objects on the top row were the chasers, the objects on the second row were the chasees, and the objects on the bottom row were the bystanders. The objects with the same role were all the same color, but varied in their shape and texture. Objects in the same column appeared in the same film clip and had the same texture.

Table 2

Example of Objects' Features from the Chase Event Type in Experiment 2 and 3

Film Clip	Type of Object		
Experiments 2 and 3	<i>Chaser</i>	<i>Chasee</i>	<i>Bystander</i>
1	S1 T1 C1	S2 T2 C1	S3 T3 C1
2	S4 T1 C2	S5 T2 C2	S6 T3 C2
3	S7 T1 C3	S8 T2 C3	S9 T3 C3
Experiments 4 and 5	Preventer	Preventee	Bystander
1	F1 H1 C1	F2 H2 C1	F3 H3 C1
2	F4 H1 C2	F5 H2 C2	F6 H3 C2
3	F7 H1 C3	F8 H2 C3	F9 H3 C3

Note. Each section presents the feature structure of the objects from one type of film clip, the chasing event (Experiments 2 and 3) with objects or the preventing event (Experiments 4 and 5) with avatars. For chasing, S = shape, T = texture, C = color. For preventing, F = facial features, H = hair color, C = clothing color. Some features were common to a given role and others to the clip.

Procedure. The procedure was largely the same as Experiment 1. To aid learning, the participants were told that the shapes of the objects were random, but the color and textures were relevant to understanding what each object is and what it does. They then watched the movies three times. In role and thematic trials in the classification test, the bystander choice in the triad was now from same movie as the target, but the matches came from different clips, as before.

Results and Discussion

Participants chose role matches slightly more often than thematic matches on the conflict trials (58% overall). Seventeen participants were primarily role responders and ten were

primarily thematic responders. (A remaining participant did not seem to learn the categories.) As seen in Table 3, role responders chose the role match 74% of the time on role trials, and the thematic match 53% of the time on thematic trials. Thematic responders chose thematic matches 70% of the time on thematic trials, and chose the role match only 28% of the time on role trials. These results replicate the finding of Experiment 1 that role responders had some knowledge of the thematic relations, but thematic responders had little recognition of the role relations.

Table 3

Experiment 2. Mean Percentages of Correct Responses by Participant Group and Triad Type

Triad Type	<i>Role Responders (n = 17)</i>	<i>Thematic Responders (n = 10)</i>	<i>Overall M (n = 28)</i>
Conflict*	86	10	57
Role	73	28	56
Thematic	52	70	58
Filler	58	56	57

* Percentage role responses is shown.

Note. The overall means include an additional participant who showed no preference.

The results of Experiment 2 are generally quite similar to those of Experiment 1, in spite of changing the stimuli so that the same-role items did not have common shapes. The overall proportion of role responses on the conflict trials was only 5% lower in Experiment 2, which was not significant, $t(42) = .37$. Thus, it appears that common shapes are not necessary for people to identify role-governed categories in such stimuli.

Furthermore, we changed the foils in the role and thematic triads so that the bystander was from the same film clip as the target was. Therefore, the choice was a more meaningful and difficult one between an object that had the specified relation but was from a different clip or an

object with no specific relation but from the same clip. This change may have reduced the selection of thematic and role items in the nonconflict triads (compare the overall rates in Tables 1 and 3). People who consistently selected thematic or role items in the conflict triads also showed clear use of those relations on their individual test trials even in competition with a bystander from the same film. The role responders chose the role item 73% of time on those trials, and the thematic responders chose the thematic item 70% on its trials in spite of a foil that shared a feature with the target item.

What is more surprising is the performance on the nonpreferred tests—especially the thematic responders on the role trials. Their responding was well below chance, 28%, $t(9) = 2.43$, $p < .05$ (and 30% in Experiment 1, though with only five participants in that group), suggesting that these participants were not simply oblivious to the shared roles but that they were actively rejecting them as the basis for categorization. It is important to realize that thematic responding is only possible in our task if participants have learned the roles, at least to some degree. That is, they must identify the target as a chaser, say, and one of the choices as a chasee in order to consistently make thematic responses. Since these items did not share colors or textures, thematic responders must have used the complementary roles to make their choices. Therefore, when they did not choose the two chasers as being in the same category, it cannot be due to an inability to recognize that they were each chasers. This leaves three options. Perhaps participants simply did not view sharing a role as being the basis for classification. Alternatively, their encoding of the role-relations were context-specific, and so they did not represent the two specific chasers as sharing a more general chaser role, and thus, this more general role was not available as a basis for categorization (Gentner, 2010). Third, it is possible that these responders saw the bystander as engaging in another thematic relation: scene co-occurrence. Given the

thematic preference in conflict and thematic trials, clearly they find causal-based thematic relations more compelling, but we cannot rule out that co-occurrence in the same scene has some value to these responders. We will address this surprising behavior in the General Discussion.

Experiment 3

Ellen Markman (1989) has discussed the paradoxical finding that children often make thematic groupings but that their category names are generally not thematically based. She suggested that children have a *taxonomic assumption* to interpret nouns as being feature-based. This makes sense, as objects that have complementary relations generally have different properties, reflecting their different roles. Soups and bowls differ greatly in their origins, functions, parts, and constituent ingredients. Children must learn that soup should go into a bowl, but calling the two by the same name would surely confuse them. As a result, when generalizing noun labels, even young children choose items that have similar shapes and properties over items that have thematic relations with previously labeled objects (Markman & Hutchinson, 1984). Adults also show evidence of this taxonomic bias, greatly reducing their thematic groupings when choosing items with the same name (Gentner & Brem, 1999; Lin & Murphy, 2001).

Another reason the use of category labels may increase role-governed categorization is that labels can induce *structural alignment*, a process that aligns relational representations and puts items into correspondence by role (Gentner & Markman, 1997; Namy & Gentner, 2002; Yamauchi, 2009). By aligning the events' structures, role-based commonalities should become more salient (Goldwater and Markman, 2011).

Experiment 3 used the same stimuli and procedure as Experiment 2, but the test questions differed. Now people were told that the target item had a novel name and were asked to choose an item with the same name. Our prediction was that this would increase the proportion of role

responses, as natural language has a number of nouns that denote role-governed categories (Goldwater et al., 2011), such as *victim*, *barrier*, *guest*, and *teacher*. In these examples, the categories may also share common properties (e.g., teachers often are adult humans with certain educational degrees, levels of knowledge, working in schools, and so on). In our experiment, the same-role items had minimal features in common—they could be identified primarily because of the kinds of actions they engaged in and how those actions related to another object. Therefore, our test examined whether people have the assumption that nouns would be more likely to label role-governed than thematic categories, above and beyond any effects of similarity.

In our earlier experiments, most participants showed some recognition of thematic relations in the thematic trials. In this experiment, however, labeling a chaser, say, as a *dwak* would not be expected to solicit the chatee as also being named a *dwak*. That leaves the bystander from the same clip as the target as the only other option in the test. Although calling the chaser and bystander by the same name would also not be optimal, our prediction was that people would prefer that over the thematic relation, where the objects are complementary. However, to take into account that in a forced-choice test neither option may be very good, we also obtained ratings of answer acceptability. Thus, even if people consistently chose one option, we can discover whether they believed that this is actually a good answer.

Method

Participants. Thirty New York University students participated in this study for course credit.

Materials. Items and movies were the same as in Experiment 2.

Procedure. The procedure was largely the same as Experiment 2. At test, instead of being asked to choose which object went best with the target to form a category, participants were told

a nonsense name of the target item (e.g., *dwak*) and asked which of the objects was most likely called that name. After choosing, they rated how likely they thought that their choice was also called that name on a scale from 1 to 7, where 1 was very unlikely and 7 was very likely.

Participants were told that even though they chose one object as most likely, they might not think this was very likely or they might think that it was very likely, and that they should give a rating that indicated their opinion. Each test trial used a different nonsense name.

Results

Overall, participants chose the role match in conflict trials 74% of the time. This was marginally higher than in Experiment 2 ($M = 57\%$), which used the identical stimuli, $t(56) = 1.67, p = .10$. (Because of the consistent strategies people used, in which many means are near 100% or 0%, the SDs are very high.) Twenty-two participants were primarily role responders, six were primarily thematic responders, and two were evenly split. Role responders chose the role match in conflict trials 92% of the time, while thematic responders chose it only 14% of the time. Role responders were fairly accurate in role trials, choosing the role match 69% of the time but did not classify the thematic relation in thematic trials, choosing the thematic match only 36% of the time, which was below chance, $t(21) = 4.96, p < .001$ (see Table 4). Thematic responders chose the thematic match in the thematic trials 78% of the time, and made the role match in role trials 56% of the time. These results differ from the findings of Experiment 1 and 2 where role responders recognized both role and thematic relations, but thematic responders only recognized thematic relations. However, since the task here was to identify items with the same name, the rejection of thematic responses was expected, as discussed above. Table 5 shows the mean likelihood ratings for thematic and role responders by choice in each trial type.

Table 4

Experiment 3. Mean Percentages of Correct Responses by Participant Group and Triad Type

Triad Type	Role Responders (n = 22)	Thematic Responders (n = 6)	Overall M (n = 30)
Conflict*	92	14	74
Role	69	55	64
Thematic	36	68	47
Filler	67	68	67

* Percentage role responses is shown.

Note. The overall means include two additional participants who showed no preference.

Table 5

Experiment 3. Mean Likelihood Ratings of Participant Groups by Match Made in Each Triad Type

	Triad Type							
	Conflict		Role		Thematic		Filler	
			Bystan		Bystan		Same	
	Role Match	Thematic Match	Role Match	der Match	Thematic Match	der Match	Movie Match	Different Movie Match
Role Responders (n = 22)	5.27	4.14*	5.34	4.64	3.30	4.08	3.04	2.90
Thematic Responders (n = 6)	5.25*	4.78	5.54	5.17	4.75	4.44*	3.97	4.33*
No Preference (n = 2)	4.17	5.50	2.50	4.08	4.03	4.83	5.50	4.75
Overall (n = 30)	5.19	4.58	5.25	4.70	3.70	4.18	3.38	3.27

*The means in these cells are based on very few data and therefore may not be stable.

We collected the likelihood ratings because we imagined that some participants would not want to choose either option in the thematic triads. The thematic choice did not overlap with the target in its features or role; the bystander only appeared in the same film clip. These are not

strong bases for sharing a name. The forced-choice test requires one of these options to be selected, so the ratings could reveal any reluctance with the selected option. Indeed, the role responders gave much lower ratings to their thematic choices, 3.3 on average, compared to 5.3 for the role choice in the role trials, $t(18) = 4.89, p < .001$. Clearly, they did not view the thematic choice as likely to share a name with the target even when they were forced to choose it. The thematic responders showed a much smaller difference, with ratings of 4.8 and 5.5 to the thematic and role responses, $t(5) = 1.11, p = .16$, indicating that they believed thematic relations a reasonable basis for naming.

Discussion

The change in the task from choosing an item that goes with the target to make a category to choosing an item that has the same name should bias participants towards taxonomic and away from thematic responding (Markman, 1989). Assuming that role-governed categories are more like taxonomic categories in their sharing of relevant features, this should have resulted in a switch towards role responding. That is in fact what happened. The proportion of role responders increased slightly in this experiment (73% compared to 62% and 61% in Experiments 1 and 2). More importantly, their classifications in the triad task changed.

This was most evident in the thematic trials. In earlier experiments, the role responders preferred role matches, but in the thematic trials where there was no such match, they were able to respond based on the thematic relation (e.g., 75% and 52% in Experiments 1 and 2). However, in Experiment 3, this proportion was driven well below chance, to 36%, indicating that these responders thought that a bystander with no fixed relation to the target was more likely to have the same name than the thematically related object. This intuition seems to be correct, as in fact, we do not use the same name for hammer and nail, bowl and soup, or shoe and foot. When

participants did choose the thematic item, their ratings indicated that they did not think that it actually was likely to have the same name.

Thematic responders did not show the same sensitivity to the name. They chose the thematic response in the thematic trials 78% of the time, suggesting that they saw no problems in giving things like hammers and nails the same name. (And of course they chose the thematically related item the majority of the time on the conflict trials.) This suggests that these participants are less sensitive to the instructions and task demands. When two items are thematically related, these participants want to group them together, and the instructions about what kind of grouping should be made do not seem to influence them very much. Another possibility is that using a different nonce name on every trial might have weakened the constraint posed by naming, since an item named on one trial might receive a different name on another trial (as pointed out by a reviewer). We made this choice so as to not bias participants to commit to one way of categorizing after an initial trial.

In sum, changing the task to generalization of a label did have an effect, but it was predominantly on the participants who were already making role-governed categories. There was still a minority of participants who grouped items thematically in spite of the change in task. The number of thematic choices did decrease overall, but the effect was not dramatic. Lin and Murphy (2001, Experiment 6) and Goldwater and Markman (2011) found a stronger decrease in thematic choices of real-world categories (e.g., bee-honey) when given the task of generalizing a name. However, the initial level of thematic grouping was much higher in their studies with familiar categories (50-80%), giving more room for a decline. Lin and Murphy also found a minority of participants (2 of 18) who still strongly preferred thematic choices when identifying items with the same name. Overall, the impression is that the use of names with familiar stimuli

in past studies had a stronger effect than the same manipulation with novel stimuli in this one. Although cross-experimental comparisons are obviously tenuous, we speculate that the novelty of the present stimuli does not provide as strong a lexical constraint. All our items were unfamiliar, so naming some of them together thematically doesn't seem as bad as, say, calling honey and bee by the same name.

Experiment 4

The same-role and thematic pairs tested in the previous experiments were all based on very similar actions within a set of clips. The chasing or lifting or crushing were all quite similar, even though they differed in their location on the screen, the shape and specific motions of objects carrying them out, and so on. Each event category can be represented by first-order relations that directly relate objects, e.g., one object closely follows another. Some roles we identify in everyday life are abstract, however, such as the notion of a barrier. This form of role-governed category, unlike chasers, is derived from higher-order relations, that is, relations that take other relations as arguments. To define a barrier, there must first be an agent who wants to achieve a goal; that is the first relation. The barrier is an entity that prevents that relation from taking place; this defines a second relation. Therefore, the overall structure is: Prevents (Barrier, (Achieve (agent, goal))). (See Wolff, 2007, for a formal representation scheme for prevention and other kinds of causal relations.) This same higher-order relation can incorporate an unbounded set of first-order relations (agents and their goals) allowing for vast differences between examples of barriers.

Identifying such roles is similar to identifying analogies: Superficial properties must be ignored and underlying relations must be discovered, with the best analogies sharing higher-order relational structure (Gentner, 1983). A boulder and a promise to one's parents might both

represent barriers, but recognizing this requires ignoring the hundreds of differences between these two things and focusing on their similar roles of preventing the accomplishment of some desired goal. All theories of analogy focus on matching relations across the two domains (e.g., Falkenhainer, Forbus, & Gentner, 1989; Hummel & Holyoak, 1997).

Our experiments are not testing the discovery of analogy, but rather the correspondence of elements in relational structures. In Experiments 1-3, the relational structures were very similar. In Experiments 4 and 5, the relational structures differed in concrete detail, requiring detection of more abstract, higher-order relations. For each role, we constructed three films that involved different actions. Each clip had objects whose abstract role could also be found in another clip: initiating the actions, preventing a desired action, or helping a desired action. As in the previous experiments, one type of object was the actor and another type was the recipient (e.g., preventer-preventee). However, unlike previous experiments, the three clips had very different actions. That is, although three clips showed a desired action being prevented, the action and its means of prevention differed across the clips.

Pilot testing suggested that people had difficulty identifying the three types of films and their roles. Because the films had very different actions, this created a memory load in order to remember, compare, and identify the similarities across the different clips. Therefore, in Experiment 4, we presented each set of three related clips together (all the prevention, all the initiation, and all the helping clips) and gave the classification tests after each set. Given the abstract relations used, we also felt that perception of the goals and their relations might be improved if the objects were clearly animate. Therefore, we changed the objects to avatars who might be perceived as being helpful, mean, or agentive more than abstract shapes would be. When different shapes all crushed different objects, one did not need to perceive psychological

properties in them to identify the roles. But for more abstract roles like prevention, a psychological goal might be necessary—or useful—to induce the underlying role. One advantage of using the avatars is that they were all clearly humanoid, with very similar shapes, and so there was no issue of shape differences in this experiment.

Method

Participants. Twenty-seven New York University students participated in this study for course credit.

Materials. We created 27 new items using an online avatar generator (see Figure 4). Each avatar had unique facial features, expression, hairstyle, and clothing. Avatars in the same movie had the same color clothing, and avatars with the same role had the same color hair, as shown in the bottom panel of Table 2, similar to the stimulus structure of Experiments 2 and 3. There were three new types of events for these movies: helping, preventing, and initiating. Each of the event types had three movie clips with distinct events. For example, the three helping movies had one movie where an avatar helped another up a hill, one where an avatar helped another get a balloon, and another where an avatar helped another through an opening. That is, the events all differed in their first-order relations and objects, but shared the helping higher-order relational structure.

Figure 4. Avatars from the helping movies in Experiment 4. The avatars on the top row were the helpers, the avatars on the second row were the helpes, and the objects on the bottom row were the bystanders. The avatars with the same role all shared the same color hair. Avatars in the same column appeared in the same film clip and had the same color clothing. These were created using Portrait Avatar Maker (http://avatarmaker.abi-station.com/index_en.shtml)

Procedure. The procedure was the same as in Experiment 2 except that the task was divided into three blocks, one for each event type. In each block, participants watched the three movie clips from one event type three times then performed the classification test for the avatars in those clips. The order of the blocks and the order of the movies within each block were counterbalanced.

Test trials were generally similar to those of the previous experiments, except that the foils had to be avatars from the current set of films, since each set was presented separately. (Clearly, no one would choose an item that had not yet been seen.) For filler trials, the target was a bystander and the two choices were another bystander and either an agent or patient, both from different movies of the same event type. These trials do not test any specific hypothesis but served to use bystanders as test items some of the time.

Results and Discussion

In contrast to the previous experiments, participants chose the role match far more than the thematic match in conflict trials (82% of the time). The majority of participants were primarily role responders with 25 participants choosing mainly the role match on conflict trials, only 2 choosing primarily the thematic match, and 1 being evenly split. Role and thematic responders chose the role match in conflict trials 89% and 16% of the time, respectively. On role trials, role responders chose the role match 52% of the time (i.e., chance) and the thematic match on thematic trials 23% of the time. (With only two participants, we do not address the thematic responders' patterns.)

Both of these results are surprising. Given that so many participants chose the role matches in the conflict trials, why was responding at chance in the role trials, which should be even easier? The answer seems to be that people did not really identify a strong common role within the films.

For example, they did not identify one character as the helper and another as the helpee in the three helping films. The strong role response in the conflict trials arose from the fact that the items with the same role has the same color hair, whereas the thematic items had no feature in common with the test item (see Table 2). Choice declined to chance in the role trials because there, one item (the role selections) had the same color hair and the other (bystander) had the same color clothing and appeared in the same clip. Role responders chose the two options about equally often, suggesting that their choices were based on shared surface properties. This result was never found in previous experiments.

The thematic trials show that most people did not identify the target as having a particular relationship with its abstract thematic pair. Instead, they chose the bystander the large majority of the time. Like the results for role matching, this may reflect the fact that the bystander (from the same clip) shared clothing color with the target, but the thematic item did not share any feature (Table 2). In the previous experiments, participants consistently paired together items with the same role or thematic pairs, showing that they were sensitive to their conceptual relations. However, the present pattern, in which even “role responders” in conflict trials did not pair items with the same roles together, does not give evidence of use of higher-order relations.

Experiment 5

People’s difficulty in identifying common abstract roles in Experiment 4 seems related to the general finding that people find analogy difficult when superficial features and first-order relations differ across items (e.g., Gick & Holyoak, 1980). In one preventing clip, a character stops another character from reaching a balloon; in another, the preventer stops another character from exiting through a gap in the wall. Although both are abstractly good examples of preventing, they involve different actions by the preventer who is attempting to stop different

goals of the preventee. Our results suggest that in such a condition, people do not identify the abstract roles that we used to construct the stimuli and fall back on more superficial commonalities such as having the same color hair or clothes to group the items.

Of course, null results have a number of possible explanations. Possibly our clips with the same roles were simply not perceived as having similar roles. Perhaps the avatar who prevented another from getting a balloon was interpreted as angry or clumsy, whereas the one who prevented someone from exiting was interpreted as protecting the preventee, or as simply being in the wrong place, and so on. For the results of Experiment 4 to be informative, we need to establish that people can identify common roles across the clips. One possible way to do that would be to simply ask people about the actions of the avatars and see whether similar relations appear in the supposedly related clips. However, people's descriptions might not do justice to their perceptions, and this measure is very far removed from the classification measure of Experiment 4. Therefore, we repeated Experiment 4 under conditions that should encourage use of role-governed categories if the intended relations could be perceived in the stimuli.

In Experiment 5, participants first viewed a set of three related clips. Then they engaged in a comparison task, in which they viewed two of those clips and wrote down their commonalities; they repeated this with another pair of clips. They then performed the triad task. Much research has shown that identifying commonalities induces structural alignment, which can shift people away from superficial dissimilarities and highlight deeper relational structure (e.g., Markman & Gentner, 1993). Therefore, if our clips are perceived as having similar roles within each group, participants should be able to identify them during the comparison phase and then use those roles in the subsequent classification task. Our dependent measure was not what people wrote in their comparisons but rather the same measure as in Experiment 4: the classification of items based on

common roles or thematic relations.

If comparison fosters role-governed categorization, it suggests that structural alignment is a crucial process in learning such categories. Such results would also corroborate Goldwater and Markman's (2011) findings on the importance of structural alignment for role-based classification with familiar verbal stimuli.

Method

Twenty-four New York University students participated in this study for course credit. Items and movies were identical to those in Experiment 4. The procedure was similar to that of Experiment 4 except that participants performed a comparison task before the classification test. After participants watched the three clips from an event type twice, they compared two pairs of clips. (As a result, one clip was used in both comparisons, and the others appeared in one comparison.) For each pair, participants watched both clips one after the other with a 1s gap in between and then explained how the clips were similar to each other by typing into a text box. The classification task was the same as in Experiment 4; in particular, participants viewed and compared the clips from one event type and were immediately tested on that type before going on to the next set of clips.

Results and Discussion

Recall that in Experiment 4, people chose the role matches in the conflict trials but then showed no sensitivity to the same or complementary roles in other trials. In contrast, participants in Experiment 5 again chose the role match in conflict trials but also showed strong role matching in the other trials.

Participants chose the role option in the conflict trials 87% of the time, and 21 out of 24 participants were predominantly role responders. (Given the small number of others, we do not

present their data separately.) In contrast to Experiment 4, people performed well on the role trials, with 74% correct (79% for the role responders). They also performed extremely well on the filler trials, choosing to match the bystanders together 90% of the time. However, all participants avoided choosing the thematic item on the thematic trials, with lower than chance levels of selection, $M = 35\%$, $t(23) = -2.42$, $p < .05$. The role responders (based on the conflict trials) similarly did not choose the thematic item, $M = 36\%$.

In Experiment 4, we argued that people did not extract the role structure of the movie clips, because their results could be explained completely by color-matching of the items. Even though people chose same-role items in the conflict trials, they did not do so in the role trials when there were two different color matches available (hair and clothing). In contrast, after comparing movies in Experiment 5, people consistently identified items as being in the same role and selected them as the same kind. Informal examination of the comparison data (the similarities that participants typed in) shows that a number of them did identify the same roles across movie clips, though not always stated in our terms. For example, people described the initiators as initiating but also as helping; the helpers were sometimes perceived as accompanying or following the helpees. In some cases, participants began by noting only superficial similarities in the first comparison (e.g., “the red avatar jumped over the yellow one twice...”) and then showed evidence of identifying the abstract commonalities in the second comparison (“the blue avatar always disrupts the red one from doing something”). However, even in cases where people seemed to focus on superficial similarities in their verbal responses, they may have noted the similar roles the items played and then used them in classification.

The results show that the clips within each event type did have similar roles as intended and that people could group items because they were both victims or initiators, and so on. Therefore,

the failure of participants to do so in Experiment 4 cannot be attributed to poor stimulus construction. This raises the interesting question of why participants failed to group items thematically in the present experiment. If they could identify two avatars as helpers, why were they loath to group together a helper and helpee, when these were the only closely related items in a triad? Instead, the large majority of times they chose the other item, which was always a bystander from the same clip as the target. An intriguing possibility is that thematic relations must be specific, based on first-order relations, in order to serve as the basis for classification. A character who helped someone in doing X is not perceived as related to a character who has been helped in doing Y. In the earlier experiments, the actions were similar across related film clips, so a chaser and a chatee were judged as being related (by many participants) even though they did not have a personal relationship. In a real-life analogy, a doctor and a patient might be grouped together even though that doctor did not treat that particular patient. However, a doctor and student would probably not be grouped together because one provides help (to someone else) and the other receives help (from someone else). In Experiments 1 and 2, participants often thematically grouped a character who chased or lifted another with a different character who was chased or lifted, respectively. However, in Experiment 5, a character who helped someone catch a balloon was *not* classified with a character who had been helped to get up a hill, even though the two helpers *were* classified together.

Common roles were seen as a good basis for classification when induced from first-order and higher-order relations, while thematic associates were only seen as a good basis for classification when connected through first-order relations. Thus, identifying thematic categories may be more grounded in specific attributes and relations than identifying role categories. This suggests that role relationships are more robust and are perceived as being stronger bases for classification

than are thematic relations, and it also suggests that different processes are at work when categories are based on roles vs. thematic associates, as we discuss below.

General Discussion

There is growing evidence for the prominence of role-governed categories in our conceptual system. This article is first to show that people can abstract novel role-governed categories from viewing novel events, without tuition or guidance from linguistic labels. Experiments 1 and 2 found that people can readily detect role-governed categories as well as thematic categories based on first-order relational commonalities. Experiment 3 showed that category labels shifted focus from thematic to role-governed categories. Experiment 4 showed that unlike with roles based on first-order relations, more abstract roles based on higher-order relations were not readily detected from the mere viewing of events. However, when participants were asked to compare events, inducing an alignment of the events' relational structures (Experiment 5), these more abstract role-governed categories were also formed.

It is important to remember that participants formed these categories with little instruction or prior knowledge. They were told to watch the films so that they could answer questions about the objects later. They were not told to learn categories or given any hints about how the objects were related. Therefore, people's later classification performance indicates that they had noticed the shared roles and encoded them during learning simply as a matter of attending to the events in the films. This is presumably similar to real-world learning without instruction, in which a learner must notice spontaneously that some novel trees seen are of the same type, different from other known trees. Such unsupervised learning of taxonomic feature-based categories is difficult unless the new categories are very distinct or are related to prior knowledge (Billman & Knutson, 1996; Spalding & Murphy, 1996). The fact that role-governed categories were formed after

viewing nine films of three different types of actions suggests that the roles involved in each action were perceived as distinct and salient.

Throughout the five experiments, roles were correlated with one or more perceptual features to serve as memory aids during test and because in the real world roles and features are often correlated. One may question whether these perceptual features were the basis for the classification judgments and not the roles. We believe this to be very unlikely. First, color alone tends not be a compelling basis for categorization. Second, after Experiment 1, bystanders overlapped with the target as much as the role choice did in the text, and yet most people strongly preferred the same-role item (though granted the bystander was probably the least salient object on the screen). Third, there was surprisingly little effect of shape, in that role responding in Experiment 2 was not significantly different from the level in Experiment 1, where the same-role items shared shape. It is possible that the event structure of the films drew attention more to the objects' actions than to their inherent properties. We did not design an experiment to specifically test the influence of shape, so we do not draw strong conclusions about this apparent lack of influence. However, the data do suggest that perceptual overlap is not driving role responding. The exact role of perceptual features is not yet clear. We have focused on the fact that people must have some way of identifying which objects served which roles, and if there were no common perceptual features, that would have been extremely difficult. But would people identify two objects as both lifters, even though they had no features in common? Or if people formed relational categories based on items that shared some features, could they apply those categories to new items that did not possess those features? These are an important questions for future research.

In natural categories, we would expect role-governed categories to share perceptible features,

albeit in a family-resemblance structure. The ability to play a given role depends on an item's shape, material, size, and other important properties. A hair cannot serve as a barrier because its size, weight, and thickness do not prevent most agents from achieving some goal. Although it is possible experimentally to separate intrinsic properties from extrinsic properties in experiments such as ours, in real life, there is a close connection between the two, just as there is between function and structure (Kemler Nelson, 1995). Our experiments permit us to conclude that people can identify common roles even with minimal perceptual support. Real-world role-governed categories that also incorporate substantive features would be identified and used even more strongly. Indeed, people intentionally add perceptual features to relational categories for ease of identification, for example, having different uniforms for different positions within an organization.

In both our experiments and in the real-world, the perceptual features are not the “definitional” part of the category representation or the central basis for category judgments, but we think that typical perceptual features are very much part of role-governed category representational content. Computational models that abstract relations from exemplars maintain perceptual features in the relational representation to the degree to which they consistently co-occur with those relations (Doumas, Hummel, & Sandhofer, 2008; Kuehne, Forbus, Gentner, & Quinn, 2000).

With the concrete categories in Experiments 1-3, nontrivial numbers of participants preferred thematic associates as a basis for classification—about a third in Experiments 1 and 2 (fewer in Experiment 3 when the task was to select objects called by the same name). They chose them not only in the trials specifically testing thematic relations but also when they had to choose between thematic and role matches. Thematic associates play complementary roles. One might ask

whether forming categories from complementary roles stems from the same processes as forming categories based on a single role. The current results suggest that they may not. Consistently across experiments, people who preferred thematically based categories *could not identify role-governed categories*: They consistently selected bystanders instead of the same-role item in the triad task. This is surprising, because logically, the detection of thematic associations and role-governed relations are opposite sides of the same coin. Because we asked people to match items that never appeared in the same clip, even the thematic matches were abstractly related—one chaser would be paired with a chasee from a different clip. Thematic responders grouped those items together but not two chasers or two chasees, also from different clips. If they could identify that chasers in general went with chasees in general, why could they not group together chasers in general?

We propose that role-governed categories were formed via the alignment of relational structures whereas thematic categories were formed by much simpler associative processes. In the experiments, each role was correlated with a perceptual feature. However, the importance of this shared feature could only be identified by comparing clips and explicitly identifying the role that the items had in common. In contrast, forming a thematic category could simply rely on associating simultaneously present agent and patient features. This association could be learned incidentally: After seeing nine clips in which a red thing lifted a blue thing, participants could easily learn that red goes with blue without necessarily remembering or clearly representing exactly what their relationship was or that red figures all played a similar role in their event. Structural alignment is a much more computationally expensive process than associative learning or simply matching surface features (Forbus, Gentner, & Law, 1995; Goldstone & Medin, 1994; Waltz, Lau, Grewal, & Holyoak, 2000). Although thematic relations involve more than

associations (see Estes et al., 2011, for discussion), an association can initiate their recognition. It is not clear whether our thematic responders did use the specific thematic relationship as the basis for grouping objects, or if they only used a simpler red-blue association.

Gentner and Brem (1999) found that thematic grouping was more likely to occur under time pressure, suggesting that it is a simpler judgment. Our task was not speeded, and RTs were extremely variable. However, thematic responders were generally faster than role responders, and in Experiment 1, thematic responders were significantly faster, $t(13) = 2.16, p < .05$. Combined with the finding that thematic responders did not form role-governed categories, this suggests that thematic categorization is an easier, more basic process that comes relatively “for free,” while role-governed categorization via structural alignment is a more effortful process that some participants may avoid engaging in. In addition, when structural alignment was experimentally induced via category labels in Experiment 3 and comparison in Experiment 5, role responders no longer viewed thematic associates as a good basis for categorization. Recent findings have shown that structural alignment processes, given the proper perceptual grounding can be performed implicitly (Day & Gentner, 2007; Day & Goldstone, 2011). However, whether role-governed categories can also be formed without explicit comparison and inference in an open question.

Lastly, we note that identifying different categorization strategies across individuals in these experiments joins a growing trend in the categorization literature. For example, there appear to be stable differences across learning tasks as to whether any individual’s goal is to find abstract rules underlying all the exemplars or instead to memorize exemplars and rely on more superficial forms of similarity (Little & McDaniel, 2014; McDaniel, Cahill, Robbins, & Wiener 2014). While Little and McDaniel (2014) found that rule-seekers are not limited to relational rules, it

seems quite possible that rule-seekers are more likely to engage in structural alignment unprompted and be role responders in the current tasks, while exemplar learners may simply rely on associative learning mechanisms and perhaps just learn the thematic-relations in the current tasks. Likewise, Rottman et al. (2012) and Goldwater and Gentner (2015) identified individual differences in whether individuals categorize natural phenomena by their abstract causal structure or whether they simply use more superficial domain-specific information (such as two phenomena being biological). Tests of thematic vs. taxonomic categorization find strong individual differences in which is preferred, which Estes et al. (2011) attribute to differing cognitive styles. How connected these individual strategies are across all these disparate tasks is an interesting question for future research.

The Nature of Relational Categories

As we noted in the Introduction, one might question whether relational categories are indeed categories at all. Without a doubt, role-governed categories are relatively common in natural language (Gentner & Kurtz, 2005). To some degree they overlap with taxonomic categories, which can have both relational components and shared attributes (*teacher, host, carnivore, pet*). Thus, role-governed categories are clearly “real,” by any reasonable standard. This is less clear for thematic categories however. One view might be to take the task at face value: Anything people will say is a category is a category. That seems a bit extreme, however, as the decision about what counts as a category is a theoretical one based on evidence concerning what kinds of mental representations are used to perform conceptual tasks (see Murphy, 2010, for discussion).

In addition to classification behavior, most theorists would argue that categories must provide inferential information. When you classify an object as being in some category, you must thereby learn something about it that you didn’t know just based on its appearance—e.g.,

its origins, composition, function, internal parts, behavior. By this standard, thematic categories do have some inductive strength (e.g., Lin & Murphy, 2001), but it is often limited to situations involving the objects' co-occurrence. For example, leash-and-dog don't have common features involving their origins, parts, or composition, but they are likely to be found in similar settings. If the leash got rained on, it is likely that the dog also got rained on. Soup-and-bowl are not made out of the same stuff, but if soup was served, it is likely that bowls were involved too.

Inferences based on co-occurrence prompts the further question: Is a functional relationship even necessary for thematic categories, or is mere co-occurrence sufficient? This form of induction certainly seems much weaker than that found in taxonomic and indeed role-governed categories, which is perhaps why thematic pairs typically do not have linguistic labels. Nonetheless, people do rely on thematic relations to group and sometimes reason about objects, and some people see such relations as being particularly salient (our thematic responders; elderly people in Smiley & Brown, 1979). Further, inferences with the categories do appear to go beyond co-occurrence to focusing on particular causal relations among thematically associated entities. Medin, Coley, Storms, and Hayes (2003) found that people would induce items from a causally related source to a target (e.g., from banana to monkey) as much as they would between two similar items (from mouse to monkey). It is not always clear in other tests of induction between thematic items (e.g., Lin & Murphy, 2001) whether a specifically causal relation or mere co-occurrence was the basis. However, one can argue that even co-occurrence implies a weak causal relationship in which a property found on one object can literally rub off onto another (e.g., a virus). We suspect that without such a potential causal link, mere co-occurrence would have very weak inductive strength.

Even with the potential causal reasoning underlying thematic categories, one could argue, as

Gentner and Brem (1999) do, that choosing thematic categories in classification tasks is basically an error, in which the association between the items overrides the task instructions to form a category (cf. our own instructions, which emphasized shared properties). However, as Markman (1989) and Estes et al. (2011) have argued, understanding thematic relations is very important for understanding events (and we certainly represent event categories). Furthermore, outside Western culture, thematic relations may be even more salient and important. Eastern cultures may focus more on thematic relations in category judgments than do Westerners, (Ji, Zhang, & Nisbett, 2004), and Native Americans are more likely to group natural categories relationally (specifically, ecologically) than are European-Americans (Medin et al., 2006).

Perhaps our thematic responders were event-focused, similar to the members of these other cultures. That is, if the event as whole constitutes the important unit to these individuals, then thematic responding could be seen as a way of completing a category exemplar. In this case, the processes behind thematic responding bear some similarity to inferring the missing features from exemplars of feature-based categories (as suggested by a reviewer; see Markman & Ross, 2003).

Despite these differences across individuals and cultures, it seems likely that all adult humans can learn about and think with both kinds of relational information in forming categories and reasoning analogically. Indeed, some have argued that relational thought is at the center of human-unique cognition and all of higher-level cognition (Gentner, 2003; Halford, Wilson, & Phillips, 2010; Penn, Holyoak & Povinelli, 2008). The current set of studies demonstrate further that relational categories can be spontaneously noticed and used to classify objects, so long as the relations are sufficiently specific.

References

- Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 458-475.
- Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological review*, 115, 1-43.
- Day, S., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition*, 35, 39-49.
- Day, S., & Goldstone, R. L. (2011). Analogical transfer from a simulated physical system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 551-567.
- Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: The apprehension and consequences of thematic relations. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 249-294). San Diego: Academic Press.
- Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19, 141-205.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195-235). Cambridge, MA: MIT Press.
- Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, 34(5), 752-775.
- Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational

language support children's learning of relational categories. *Child Development*, 82, 1173-1188.

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., & Brem, S. K. (1999). Is snow really like shovel? Distinguishing similarity from thematic relatedness. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual meeting of the Cognitive Science Society* (pp. 179-184). Mahwah, NJ: Erlbaum.

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W.-k. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 151-175). Washington, DC: APA.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45-56.

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

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 3-28.

Goldstone, R. L., Day, S., & Son, J. Y. (2010). Comparison. In B. Glatzeder, V. Goel, & A. von Müller (Eds.), *On thinking: Volume II, towards a theory of thinking* (pp. 103-122). Heidelberg, Germany: Springer.

- Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, 137, 137-153.
- Goldwater, M. B., & Markman, A. B. (2011). Categorizing entities by common role. *Psychonomic Bulletin & Review*, 18, 406-413.
- Goldwater, M. B., Markman, A. B., Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, 118, 359-376.
- Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in cognitive sciences*, 14(11), 497-505.
- Holyoak, K. J., & Thagard, P. (1989). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 242-266). New York: Cambridge University Press.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427-466.
- Ji, L.-J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, 87, 57-65.
- 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. (2011). Progressive alignment facilitates learning of deterministic but not probabilistic relational categories. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*.
- Keane, M. T. (1997). What makes an analogy difficult? The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

23(4), 946.

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

Kemler Nelson, D. G. (1995). Principle-based inferences in young children's categorization:

Revisiting the impact of function on the naming of artifacts. *Cognitive Development*, 10, 347-380.

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2006). Ideals aren't always typical: Dissociating

goodness-of-exemplar from typicality judgments. In R. Sun & N. Miyake (Eds.),

Proceedings of the 28th Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

Kuehne, S., Forbus, K., Gentner, D., & Quinn, B. (2000, August). SEQL: Category learning as

progressive abstraction using structure mapping. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 770-775).

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical

learning. *Cognitive Development*, 3, 299-321.

Lewis, G. A., Poeppel, D., & Murphy, G. L. (2015). The neural Bases of taxonomic and thematic

conceptual Relations: An MEG study. *Neuropsychologia*.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of*

Experimental Psychology: General, 130, 3-28.

Little, J. L., & McDaniel, M. A. (2014). Individual differences in category learning:

Memorization versus rule abstraction. *Memory & cognition*, 1-15.

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*

and Theoretical Artificial Intelligence, 13, 329-358.

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

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs thematic relations. *Cognitive Psychology, 16, 1-27.*

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143, 668-693.*

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. L. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10, 517-532.*

Medin, D. L., Ross, N. O., Atran, S., Cox, D., Coley, J., Proffitt, J. B., et al. (2006). Folkbiology of freshwater fish. *Cognition, 99, 237-273.*

Murphy, G. L. (2001). Fast-mapping children vs. slow-mapping adults: Assumptions about words and concepts in two literatures [commentary]. *The Behavioral and Brain Sciences, 24, 1112-1113.*

Murphy, G. L. (2010). What are categories and concepts? In D. Mareschal, P. C. Quinn & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 11-28). Oxford: Oxford University Press.

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.*

Nelson, K. (1986). *Event knowledge: Structure and function in development.* Hillsdale, NJ:

Erlbaum.

Nguyen, S. P., & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development*, 74, 1783-1806.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31(02), 109-130.

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

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

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive science*, 36(5), 919-932.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

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.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preferences for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, 28, 249-257.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22,

525-538.

Tomlinson, M. T., & Love, B. C. (2010). When learning to classify by relations is easier than by features. *Thinking & Reasoning*, 16, 372-401.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, 113, 169-193.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, 136, 82-111.

Yamauchi, T. (2009). Finding abstract commonalities of category members. *Journal of Experimental and Theoretical Artificial Intelligence*, 21, 155-180.