Structural Thinking about Social Categories: Evidence from Formal Explanations, Generics, and Generalization

Nadya Vasilyeva (nadezdav@princeton.edu)

Tania Lombrozo (lombrozo@princeton.edu)

Department of Psychology, Princeton University, Princeton, NJ, 08540, USA

Abstract

Most theories of kind representation suggest that people posit internal, essence-like factors believed to underlie kind membership and the observable properties of members. Across two studies (N = 234), we show that adults can construe properties of social kinds as products of both internal and *structural* (stable external) factors. Internalist and structural construals are similar in that both support formal explanations (i.e., "category member has property P due to category membership C"), generic claims ("Cs have P"), and a particular pattern of generalization to individuals when the individuals' category membership and structural position are preserved. Our findings thus challenge these phenomena as signatures of essentialist thinking. However, once category membership and structural position are unconfounded, different patterns of generalization emerge across internalist and structural construals, as do different judgments concerning category definitions and property mutability. These findings have important implications for reasoning about social kinds.

Keywords: structural explanation; kind representation; generalization; essentialism; inherence; social categorization

Introduction

Kind representations allow people to organize, store, and use conceptual information efficiently and productively. We rely on our representations of social groups and natural kinds to make sense of the world, generate explanations, and make predictions about the individual category members we encounter. Most theories of kind representation, especially for natural and social kinds, emphasize an internalist bias, a tendency to look "within" the kind for deep, causally active, and explanatorily powerful factors that hold categories together, and that shape and maintain the properties of their members. This internalist bias can take the form of assumptions about internal causal structure (psychological essentialism; Gelman, 2003), or a preference for explanations citing factors that are inherent, as opposed to contextual or extrinsic (inherence heuristic; Cimpian & Salomon, 2014). Different manifestations of internalist bias have been widely documented (Haslam et al., 2010; Gelman, 2003; Rangel & Keller, 2011), and it has been proposed as a conceptual default, with profound – and often negative - consequences for the way we think about and behave towards members of social categories. For example, explaining a dearth of women in mathematics by appeal to their "essential" or inherent nature can discourage girls from pursuing careers in this field (Leslie, Cimpian, Meyer, & Freeland, 2015).

Some linguistic forms have been argued to promote internalist construals, in particular of social kinds. Generic expressions, which attribute a property to a category in general (e.g., "women fail math tests") have received particular attention (e.g., Cimpian, 2010; Cimpian & Markman, 2010; Rhodes, Leslie, & Tworek, 2012). There is also evidence that formal explanations, which appeal to category membership to explain a property (e.g., "Priya doesn't like math because she's a girl"), reflect internalist beliefs (Gelman, Cimpian, & Roberts, 2018; Prasada & Dillingham, 2006).

While internalist modes of thinking have been extensively explored in the psychological literature, alternative ways of representing kinds have received much less attention. One such alternative is *structural thinking*, and in particular, a structural construal of category-property connections (Haslanger, 2015; Vasilyeva, Gopnik, & Lombrozo, 2018). On a structural construal, stable associations between categories and their properties arise from stable external constraints acting on category members. For example, the categories "women," "men," "Blacks," and "Latin@s" occupy relatively stable social positions within a given social structure. These positions can differ across cultures and possess their own properties. To illustrate, the generics "women don't drive," or "women are bad a math," can be true in one social system but false in another. Such culture-dependence is one cue that a property-category association should be attributed to a *social position* rather than to *the category occupying that position.*

Because a social position and the category that occupies it can share the same label (e.g., "women"), we contend that generics and formal explanations can be interpreted in either internalist or structural terms. For example, a person could endorse a formal explanation ("He ended up in prison because he's Black") or a generic ("Black men end up in prison") for different reasons: under an internalist construal, attributing the property ("being in prison") to the category itself (e.g., presumed criminal inclinations), or under a structural construal, attributing the same property to the social position, constituted by a conglomeration of stable constraints acting on members of the category in virtue of occupying that position (e.g., unequal opportunities for Black youth, biased hiring and other barriers to wealth, racial profiling by the police, etc. – all the factors that together constitute the social position "Black" in the US).

In the current research, we test the prediction that adults can construe property-category associations in either

internalist or structural terms, and that both construals support formal explanations (Study 1) and generics (Study 2). However, we also investigate important ways in which internalist and structural construals are expected to differ. Because internalist and structural construals allocate different roles to category membership (vs. a category's social position) in explaining an associated property, we expect internalist and structural construals to result in different intuitions about using the property in category definitions (Study 1), different "mutability" judgments about true category membership when the property is removed (Study 1), and different patterns of property generalization as category membership and/or social position change (Study 2).

Documenting these predicted patterns of similarity and difference across internalist and structural construals is important for a number of reasons. First, alternatives to internalist thinking have rarely been articulated and tested. Documenting a psychologically real alternative can thus enrich our understanding of the mental representations that support our thinking about social (and potentially non-social) kinds. Second, given that internalist construals have been linked with the perpetuation of stereotypes and other negative social effects (Bastian & Haslam, 2004; Cimpian, 2010), an alternative form of construal could identify changes in mindset that would mitigate these effects. A structural construal is especially promising in that it explains (rather than ignores) property-category associations that in fact obtain (such as a low proportion of women in math) while also pointing to structural factors that could be targets of intervention.

While structural explanation has received attention within the philosophy of social science (Ayala, 2018; Ayala & Vasilyeva, 2015; Haslanger, 2015; Garfinkel, 1981), there has been little empirical work on the topic to date. In a recent paper, Vasilyeva, Gopnik, and Lombrozo (2018) reported a study investigating structural thinking in adults and children aged 3-6. Using open-ended explanations, category definition tasks, mutability judgments, and measures of formal explanation, they found that even 3-year-olds showed signs of early structural thinking, with greater differentiation between internalist and structural construals in older children and adults. The present work goes beyond Vasilyeva, Gopnik, and Lombrozo (2018) in five important ways: in using more realistic social categories (a group of immigrants); in exploring structural reasoning about novel social groups; in using a wide range of properties matched in terms of property/cue validity (Study 1) or content (Study 2); in the introduction of a control condition (Study 1 and Study 2), and in exploring judgments concerning generics and generalizations under different conditions (Study 2).

Study 1

In Study 1, we introduce participants to a novel social category ("Borunians," an immigrant group in the fictional country of Kemi), along with a suite of associated properties (e.g., holding low-paying jobs). Across properties, we vary whether the category-property connections are explained in a way that is internalist (e.g., appealing to group identity), structural (appealing to social position), or incidental (the associations just happen to be true). To test whether this manipulation is successful in inducing different construals, we adapt measures originally developed in Prasada and Dillingham (2006, 2009) to differentiate "principled" and "statistical" connections, and used also in Vasilyeva, Gopnik, and Lombrozo (2018). These measures include partial definition evaluation (i.e., whether the category can be defined in terms of the property), category mutability ratings (i.e., whether an individual missing the property is a true category member), and formal explanation evaluation (i.e., whether the presence of the property can be explained by appeal to category membership).

First, as explained in the introduction, we expected both internalist and structural construals to support formal explanations (e.g., "He holds a poorly paid job because he's a Borunian"). Second, we expected the internalist and structural conditions to differ with respect to partial definitions and mutability. A definition of a category in terms of an "essential"/inherent feature should be more appropriate than a definition citing a feature that holds only in virtue of a category's position in a social structure. Likewise, removing an internal feature should produce more damage to category membership than removing a feature acquired through a social position, and therefore contingent on external structure. These predictions found support in Vasilyeva, Gopnik, and Lombrozo (2018); we test them here with a more realistic social kind, a broader range of features, and a modified mutability measure.

Finally, and going beyond Vasilyeva, Gopnik, and Lombrozo (2018), we included features with an "incidental" explanation for which we predicted a profile of effects different from either internalist or structural thinking, based on Prasada and Dillingham (2006, 2009). We expected that incidental features would not support definitions and would be seen as easily mutable (like structural features), but that they would not support formal explanations (in contrast to both internalist and structural features).

Method

Participants Seventy-seven participants (38 women, 39 men; mean age 33) were recruited on Amazon MTurk in exchange for $1.50; in this and subsequent studies participation was restricted to workers with an IP address within the United States and with a HIT approval rating of 95% or higher from at least 50 previous HITs. An additional 33 participants were excluded for failing a memory check.

Materials, Design, and Procedure Participants read a short vignette introducing the novel social category of "Borunians" - a group of immigrants settled in a fictional country, *Kemi*, who originally immigrated from Bo-Aaruna. Borunians were characterized by 18 unique features, with 6 of each type: *Internalist* (tying the feature to Borunians' tradition and identity), *incidental* (roughly equivalent to Prasada and Dillingham's (2006) "statistical"), and

Table 1. Examples of features used in Study 1.

Internalist: Borunian traditions are extremely important to them, and form part of their identity: Borunians have a special tattoo on one arm.	*Incidental:* Here are some statements about Borunians that are true, but there's nothing about these features that ties them to Borunian culture, tradition, personality or anything about their place in Kemi society: Borunians barbeque in their back-yards all year round, so they buy a lot of barbequing coal all year round.	*Structural:* Here are a few characteristics that Borunians have due to their position in the Kemi society and governmental policies applying to Borunians: Borunians are *not* allowed to take any job with an income over 20,000 Kemi dollars per year (approximately 20,000 USD) if other appli-cants for the same job include Kemi citizens who are equally or more qualified. Due to this regu-lation, Borunians hold mostly poorly paid jobs.

structural (tying the feature to the structural constraints acting on Borunians due to their position within Kemi society). Sample features are shown in Table 1. All features were presented in generic form. A norming study with a separate group of 23 participants verified that the three feature types did not differ in mean cue and category validity.

After learning the features, each participant performed one of three judgments – formal explanation (e.g., Question: Why does he hold a poorly paid job? Answer: Because he is a Borunian. How good is this explanation? 1 not good at all - 7 very good), *partial definition* (e.g., Question: What is a Borunian? Answer: A Borunian is a person who holds a poorly paid job. How good is this answer? 1 not good at all – 7 very good), or *mutability* (e.g., Imagine an alternative world where people we call Borunians do not hold mostly poorly paid jobs. From your perspective, would you call them really and truly Borunians? (1 definitely no - 7 definitely yes). Each judgment involved 18 ratings, one about each feature. Prior to the main set of ratings, participants practiced the judgment type they were assigned on two practice trials that involved rating a feature of a dog ("has four legs") and of a barn ("is red").

In sum, the study implemented a 3 (judgment type: formal explanation, partial definition, mutability; between subjects) by 3 (feature type: internalist, structural, incidental; within subjects) design.

Results and Discussion

Participants' ratings were analyzed in an ANOVA with feature type as a within-subjects factor and judgment as a between-subjects factor, followed by planned *t*-tests. The main effect of judgment was significant, $F(2,74) = 5.70$, $p = .005$, $\eta_p^2 = .133$, and the main effect of feature type was marginal, $F(2,148)=2.99$, $p=.053$, $\eta_p^2=.039$. However, of most theoretical importance was the significant interaction between judgment and feature type, $F(4,148)=31.54$, $p<.001$, $\eta_p^2=.460$. As shown in Figure 1, each feature type had a unique "profile" across the three judgments. As predicted, and replicating Prasada and Dillingham's (2006, 2009) findings, internalist features (relative to incidental features) better supported formal explanations ($p=.003$) and definitions ($p<.001$), and were judged less mutable ($p< .001$). Also as predicted, structural features (relative to internalist features) supported definitions less strongly, and mutability judgments more strongly ($ps<.001$). However, they supported formal explanations to the same extent ($p=.327$), and more strongly than incidental features did ($p<.001$).

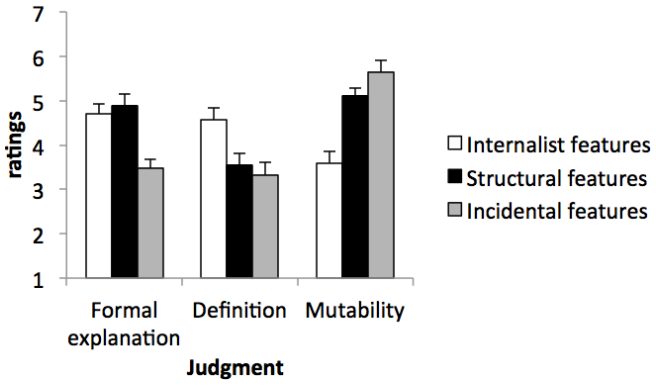

Figure 1: Participants' ratings as a function of feature type and judgment in Study 1. Error bars represent 1 SEM.

In sum, we find the predicted profile of effects for category-property associations introduced with a structural explanation. These associations behaved like internalist features in supporting formal explanations, but like incidental features in terms of partial definitions and mutability. It is worth noting that this structural pattern of responses was elicited by offering appropriate cues in the feature description, but did not require any explicit guidance or training in structural reasoning. This suggests that this mode of thinking may occur naturally in adults' cognitive lives when appropriate cues are present. It's also notable that the cues took the form of explanations, which presumably fed into causal-explanatory models that supported a representation that attached the property to the category versus the social position it occupied.

Study 2

In Study 2, participants received information about the prevalence of a property in the Borunian population, as well as an internalist, structural, or no explanation for the category-property association. Participants then rated their endorsement of a corresponding generic claim ("Borunians [have property]"), and generalized the properties in question to individual targets that varied both in category membership (same or different) and in social position (same or different).

This design allowed us to test two predictions. First, we predicted that internalist and structural construals would similarly support generic claims, with higher endorsement the greater the prevalence. Prior work has already shown that an internalist construal is not *necessary* for a generic to be endorsed; even statistical connections can support generics (Prasada & Dillingham, 2006, 2009; Tessler & Goodman, 2016). However, a structural construal additionally supports an interpretation of a generic claim whereby the category label refers to the social position that the category occupies.

Second, we predicted different patterns of generalization for properties explained internally vs. structurally as the generalization target varied in category membership and social position. In most real-life contexts, social category and social position are confounded, meaning that both internalist and structural explanations support the extension of category properties to individual category members (albeit for different reasons). De-confounding the category and position in our task, however, allows a predicted divergence between internalist and structural consturals to emerge. We expected a structural explanation to support greater generalization on the basis of shared position (relative to internalist), and an internalist explanation to support greater generalization on the basis of shared category (relative to structural). Additionally, we expected the effect of prevalence on generalization to be moderated by explanation, such that for participants who received a structural explanation (vs. internalist), prevalence effects would be weaker when social position was not preserved, and for participants who received an internalist explanation (vs. structural), prevalence effects would be weaker when category membership was not preserved. Comparisons to the control condition allowed us to assess the extent to which these effects were driven by a structural construal, an internalist construal, or both.

Method

Participants One-hundred-and-fifty-seven adults (76 women, 80 men, 1 agender; mean age 37) participated online in exchange for $1.50. An additional 30 participants were excluded for failing memory and attention checks.

Materials, Design, and Procedure We developed a new set of twelve features describing a fictional immigrant category, Borunians, introduced as in Study 1, and an internalist explanation and a structural explanation for each feature (see Table 2 for sample features and explanations). For the internalist condition, we intentionally chose a range of explanations spanning from more biological to those citing group preferences, values, and traditions (see further comments on this in the General Discussion). We also took care to keep the internalist and structural explanations of similar average length.

Each participant was assigned to one explanation condition (internalist, structural, or control), and completed two blocks of measures: generic truth ratings, and individual generalizations. In the generic truth rating block, participants saw the 12 features of Borunians, one at a time, in a random order, each accompanied by prevalence information (e.g., "Percentage of Borunians who hold poorly paid jobs: 48%"). For participants in the internalist or structural conditions, this was also accompanied by an explanation of the corresponding type (e.g., "Reason: in order to hire a Borunian for a well-paid job, employers in Kemi are required to file complicated government paperwork"). The feature prevalence (i.e., the percentage of Borunians with the feature) was drawn from a pool of 12 unique values, binned into Low (M=25%, range 20-29), Medium (M=50%, range 46-55), and High (M=75%, range 71-80). Below the prevalence information and explanation (if presented), participants read a generic statement attributing the feature to the category (e.g., "Borunians hold poorly paid jobs"), and were asked to classify it as "True" or "False."

In the individual generalization block, participants were asked to generalize a property from the kind (Borunians) to an individual. Participants were asked to rate their confidence that one of the properties previously attributed to Borunians (e.g., "holds a poorly paid job") held for that individual on a 9-point scale ranging from -4 (I'm confident it's false) to +4 (I'm confident it's true). Crucially, we manipulated both the category membership and the social position of the target individual: same vs. different category membership, and same vs. different social position. The resulting four scenarios are described in Table 3.

To ensure that participants still remembered the prevalence level and the explanation of each feature, the generalization rating block was split into three sets of four questions each. Each set of four questions was preceded by a reminder display with four features along with their prevalence levels and explanations (repeating the information from the first block). Further, to reduce memory load for prevalence levels, all four features in a set were pulled from the same prevalence bin (e.g., all had High prevalence). Following the reminder, participants saw

Table 2. Sample features and explanations used in Study 2. Each explanation was presented within the frame "Reason: [explanation]."

Feature	**Internalist Explanation** [Reason: ….]	**Structural explanation** [Reason: ….]
Follow a largely vegetarian diet	… a deficiency in digestive enzymes required for digesting meat	…special access to municipal subsidies to purchase vegetables directly from local farmers
Sell artisan souvenirs	…a natural affinity for design and great facility with fine-motor tasks	....special subsidies from the Kemi government to Borunians to obtain vendor permits for artisan booths
Get sunburn easily	…a genetic variation which makes Borunian skin very vulnerable to the effects of sunlight	…a high proportion of contaminants and skin irritants in the neighborhoods where Borunians live; these substances make their skin vulnerable to the effects of sunlight
Participate in donkey races	…agility and inherent skill with animals	…not allowed to participate in horse or car races
Live with their parents through adulthood	…a special value attached to family and elders, as well as living in tight-knit communities	… inability to afford the cost of maintaining independent residences
Hold poorly paid jobs	… strong preference to work regular hours; avoidance of demanding jobs that may require over-time	…in order to hire a Borunian for a well-paid job, employers in Kemi are required to file complicated government paperwork
Have poor credit ratings	…Borunians' reliance on a peculiar calendar with a different month length results in frequent late payments	….government banks imposed an additional step to verify every transaction for new immigrants, resulting in frequent late payments

Table 3. Descriptions of generalization targets produced by crossing same/different social category with same/different social position (note: social position is not the same as geographic location; non-Borunians in Kemi occupy a different social position from Borunians).

Scenario	Category	Position	Description
ALL SAME	Same	Same	Azz is a Borunian, and lives in Kemi.
MOVED	Same	Different	Nuvo is a Borunian who moved from Kemi a long time ago, and now lives in a completely different country, with an entirely different social system and regulations.
ADOPTED	Different	Same	Pau is a NON-Borunian by birth, who was adopted into a Borunian family in Kemi at a very young age, a long time ago, in a secret adoption (meaning the fact of adoption was never revealed, nobody except the parents knew that the child was adopted, and the child was brought up as a Borunian).
ALL DIFFERENT	Different	Different	Eken is a NON-Borunian who lives in Kemi.

the four generalization questions (one from each row of Table 3), in random order. The assignment of features to prevalence levels and question types, as well as the order of question sets, were counterbalanced across participants.

At the end of the survey, participants responded to a series of memory and comprehension checks (e.g., asking them to classify a list of characteristics and explanations as mentioned vs. not mentioned in the survey), as well as individual difference measures that are not analyzed here.

Results

Generic Truth Ratings Data were analyzed in a mixed effects logistic regression, predicting generic truth ratings from numerical prevalence, explanation, and their interaction (allowing for random intercepts for participants and items). To compare all three explanation conditions in this and the following regression models, the model was fit with the control condition as the reference group, and then re-fit with the structural condition as the reference group. Prevalence was the only significant predictor ($p<.001$): the odds of a "true" judgment increased 1.10 times per unit of increase in prevalence. Binning the prevalence predictor into three levels, the mean proportions of "true" responses were .25 (Low), .74 (Medium), and .94 (High). All other predictors were not significant, $ps \geq .244$, indicating that explanation condition did not affect overall generic endorsement, nor moderate the effect of feature prevalence.

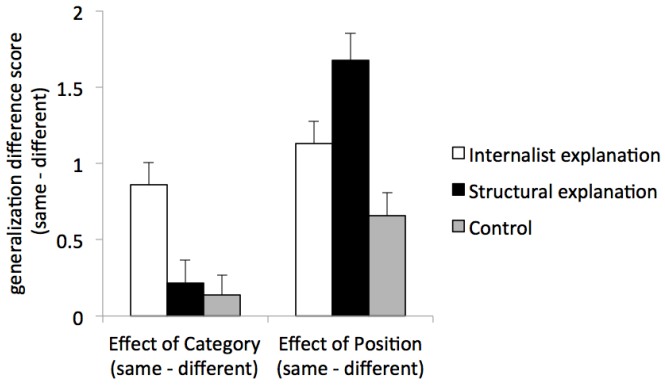

Figure 2: To represent the interactions between explanation condition, shared category membership, and shared social position, we created "generalization difference scores" (mean difference in generalization to individual in same vs. different category, and same vs. different social position). Error bars represent 1 SEM.

Individual Generalization A hierarchical linear model predicting individual generalization from centered numerical prevalence, explanation condition, shared category (yes or no), and shared social position (yes or no), with random intercepts across participants, revealed a four-way interaction, $p=.001$. To investigate this interaction further we ran additional analyses. First, to evaluate the prediction that an internalist explanation elevates the importance of shared category membership as a basis for generalization (relative to structural or control), we dropped prevalence and shared social position from the model, and predicted individual generalization from condition and shared social category. As expected, we observed significant interactions between regressors. The effect of shared category membership was stronger in the internalist condition relative to structural, $p=.006$, and to control, $p=.002$, which did not differ from each other, $p=.734$ (Figure 2). Second, to evaluate the prediction that a structural explanation elevates the importance of shared social position as a basis for generalization (relative to internalist or control), we predicted individual generalization from condition and shared social position. Again, we observed the expected interactions between regressors (see Figure 2), revealing a stronger effect of shared social position in the structural condition than either the internalist, $p=.014$, or control condition, $p<.001$. The internalist condition also heightened the relevance of social position relative to the control condition, $p=.036$, which suggests that our internalist explanations (perhaps by appealing to culture) also involved some social / structural elements.

Next, we addressed the prediction that the effect of prevalence on generalization would be moderated by explanation type. Given that the prevalence estimates that were offered corresponded to Borunians in Kemi (and not necessarily to non-Borunians or Borunians in other social positions), we expected the effect of prevalence to weaken with distance from the "ALL SAME" generalization target. However, we also expected that a change in category membership would attenuate the effect of prevalence more strongly in the internalist than in the structural condition, and that a change in social position would attenuate the effect of prevalence more strongly in the structural than in the internalist condition. To address this prediction, we considered the two cells that crossed category membership and social position ("ADOPTED" and "MOVED"; see Table 3), and ran separate models predicting individual generalization from prevalence and explanation condition (see Figure 3).

In the "MOVED" scenario, prevalence positively predicted generalization, β=.41, p<.001. Mirroring the results presented in Figure 2, participants were also *less* likely to generalize in the structural condition than in the internalist condition, B=-.45, p<.001, or in the control, B=-30, p=.008 (the latter two did not differ, B=.15, p=.181). Most crucially, however, we also observed interactions, such that the effect of feature prevalence was *weakened* in the structural condition relative to the internalist condition (B=-.31, p=.002) and control (B=-.33, p=.001); the effect of prevalence did not vary across the latter two explanation conditions (B=.018, p=.865).

In the "ADOPTED" scenario, prevalence positively predicted generalization, β=.67, p<.001. However, the predicted interaction between prevalence and explanation, with an attenuated effect of prevalence in the internalist condition (relative to control) was only marginal, B=-.19, p=.0502. Moreover, contrary to our expectations, the effect of prevalence was significantly attenuated in the structural condition (relative to control), B=-.33, p=.004. The extent to which the prevalence effect was attenuated, relative to control, did not differ across the two explanation condition, p=.117.

Finally, we considered the remaining two generalization targets, "ALL SAME" and "ALL DIFFERENT," for which we did not predict differential effects of explanation type. In

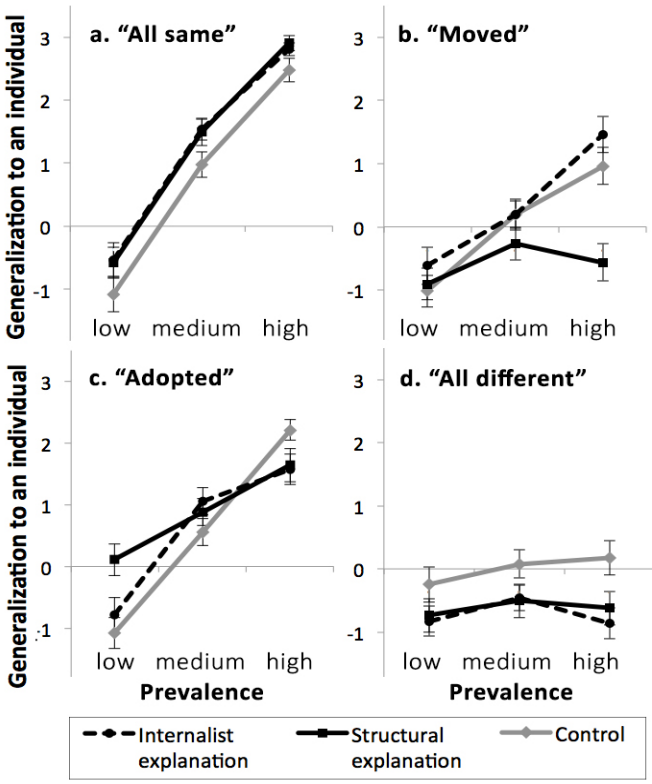

Figure 3: Mean individual generalization ratings as a function of within-category feature prevalence (binned into low, medium, and high ranges for presentation) and explanation type, split by the scenario (same or different category, and same or different social position). Error bars represent 1 SEM.

predictor of generalization, β = 70, p<.001, and both the "ALL SAME" scenario, prevalence was a positive internalist and structural explanations boosted generalization relative to control (B_{Int}=.20, p=.031; B_{Str}=.19, p=.036); the internalist and structural explanations did not differ, p=.945. As predicted, there were no significant interactions, ps ≥.780.

In the "ALL DIFFERENT" scenario, feature prevalence did not predict generalization, β=.10, p=.162. Participants were less likely to generalize in either explanation condition, relative to control ($B_{Int\ vs.\ control}$ =-.39, p=.005; $B_{Str\ vs.\ control}$ = -.34, p=.013); the two explanations did not differ, p=.708. As predicted, there were no significant interactions, ps >.238.

Discussion

Study 2 identified important respects in which internalist and structural construals overlap: both support generics, and both are equally sensitive to within-category/position statistics when it comes to endorsing generics or drawing generalizations to individuals within the same category/position. On the other hand, internalist and structural construals diverge when it comes to generalizations that break the typical confounds between categories and social positions: an internalist construal favors generalization (and reliance on within category/position statistics) across changes in social position; a structural construal is less sensitive to the preservation of category membership. These patterns emerged clearly in the "MOVED" scenario; the "ADOPTED" scenario (which was also the most unusual) was less clear.

In real life, the divergence between internalist and structural construals might be even more pronounced than that observed here. For experimental purposes, we used the same features across explanation conditions; as a result, many invoked culture and group identity, possibly downplaying more internalist factors. Indeed, shared social position was more influential overall than shared category, and shared position boosted the generalization of internalist features relative to control (Figure 2). Plausibly, the internalist condition could have been made "more internalist" by using different feature sets across conditions and citing exclusively biological factors in internalist explanations, as is common within the abundant literature documenting essentialist (or more broadly internalist) reasoning. Given that our goal was instead to document the reality of a structural construal as distinct from an internalist construal, we opted for greater experimental control over maximally representative features.

General Discussion

Across two studies we document underappreciated flexibility in people's construal of social kinds: in addition to adopting an internalist construal (familiar from prior research), people are capable of adopting a structural construal, which makes sense of observed correlations between properties and categories without tying them to the inherent nature of the category. Given the dangers of internalist construals in the social domain, an alternative that

makes sense of observed correlations without perpetuating them could be of no small social value.

Contrary to the dominant view, generic language does not necessarily convey or induce essentialist beliefs: in Study 1, the generic language that introduced a category-property association did not prevent an alternative construal, and in Study 2, both construals supported the endorsement of generic claims. This calls for refining numerous claims about generics and formal explanations as ways of inducing essentialism or signaling which kinds should be essentialized (e.g., Rhodes, Leslie, & Tworek, 2012). At the same time, it remains a possibility that internalist construals are the default, or cognitively less demanding.

Our generalization results also have important implications: from a theoretical standpoint, they offer yet another illustration of how explanation shapes generalization (Lombrozo & Gwynne, 2014; Sloman, 1994; Vasilyeva & Coley, 2013; Vasilyeva, Ruggeri, & Lombrozo, 2018), directing it along the dimensions of shared category and/or position. From a methodological standpoint, they offer a cautionary note about interpreting generalization measures as indices of essentialism; willingness to generalize a category's property can signal either an internalist *or* a structural construal. Finally, from a practical standpoint, getting a fuller picture of how people generalize from categories to individual has important real-life implications (e.g., a manager deciding whether to hire a woman, based in part on an inference about whether she'll take a parental leave).

One interesting question that deserves future attention is how people represent and reason about the "cultural" properties of social groups, such as food or religious customs. Where do such features fall on the internalist-structural continuum? One possibility is to identify "cultural" with "structural." However, many aspects of culture, including preferencves, values, and attitudes, can be understood in internalist terms, where cultural properties reflect shared internal characteristics. Consistent with these dual interpretations, the very same cultural properties in Study 2 were treated as internalist (through explanations citing "a special value attached to family and elders" or "a strong preference to work regular hours") or as structural (by attributing them to stable external constraints). However, the question of how people reason about "cultural" features in more naturalistic contexts remains open.

In sum, across two studies, we show that internalist and structural construals elicit different representations of categories: in the former case a property is attached to the category, in the latter case to its social position. While both kinds of representations can effectively track environmental statistics and support inferences, they work differently, in ways that could have tangible social consequences.

References

Bastian, B. & Haslam, N. (2005). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228–235

Cimpian, A. (2010). The impact of generic language about ability on children's achievement motivation. *Developmental Psychology*, *46*(5), 1333–1340.

Cimpian, A., Brandone, A.C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive science*, *34*(8), 1452-1482.

Cimpian, A., & Markman, E.M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development*, *82*(2), 471–492.

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences, 37*(5), 461–480.

Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. New Haven: Yale University Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press.

Gelman, S. A., Cimpian, A., & Roberts, S. O. (2018). How deep do we dig? Formal explanations as placeholders for inherent explanations. *Cognitive psychology*, *106*, 43-59.

Haslam, N., Rothschild, L., & Ernst, D., (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*, 113–27.

Haslanger, S. (2015). What is a (social) structural explanation? *Philosophical Studies*, *173*(1), 113-130.

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347(6219)*, 262-265.

Lombrozo, T., & Gwynne, N.Z. (2014). Explanation and inference: mechanistic and functional explanations guide property generalization. *Frontiers in Human Neuroscience, 8*(September), 700. doi; 10.3389/fnhum.2014.00700

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*(1), 73–112.

Rangel, U., & Keller, J. (2011). Essentialism goes social: Belief in social determinism as a component of psychological essentialism. *Journal of Personality and Social Psychology, 100*(6), 1056-1079.

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *PNAS*, *109*, 13526–13531.

Sloman, S. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition, 52*, 1-21.

Vasilyeva, N., & Coley, J.C. (2013). Evaluating two mechanisms of flexible induction: Selective memory retrieval and evidence explanation. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology, 54*(9), 1735-1744.

Vasilyeva, N., Ruggeri, A., & Lombrozo, T. (2018). When and how children use explanations to guide generalizations. In T.T. Rogers, M. Rau, J. Zhu, & C.W. Kalish, (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. (pp. 2609-2614). Austin, TX: Cognitive Science Society.