Cognition 204 (2020) 104383

Contents lists available at ScienceDirect

Cognition

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

Structural thinking about social categories: Evidence from formal explanations, generics, and generalization

Nadya Vasilyeva*, Tania Lombrozo

Princeton University, United States of America

ARTICLE INFO

Keywords:
Structural explanation
Kind representation
Generalization
Essentialism
Inherence
Social categorization

ABSTRACT

Many theories of kind representation suggest that people posit internal, essence-like factors that underlie kind membership and explain properties of category members. Across three studies (N = 281), we document the characteristics of an alternative form of construal according to which the properties of social kinds are seen as products of *structural* factors: stable, external constraints that obtain due to the kind's social position. 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 the generalization of category properties to individual category members when kind membership and social position are confounded. Our findings thus challenge these phenomena as signatures of internalist 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 the dispensability of properties for category membership. We discuss the broader implications of these findings for accounts of formal explanation, generic language, and kind representation.

Consider the claim that "immigrants hold poorly-paid jobs," or that a given person holds a poorly-paid job "because she is an immigrant." Such claims could come from someone convinced that immigrants are less capable, less hard working, or less focused on achievement than others – in other words, from someone who attributes the association between the category (being an immigrant) and the property (holding a poorly-paid job) to inherent, deep, and possibly essential characteristics of the group, in isolation from its context. We call this an *internalist* construal. However, the very same claims could come from someone who instead holds that immigrants face unique socio-economic barriers and challenges in contemporary society – in other words, from someone who attributes the association between the category and the property to stable external constraints that act on category members in virtue of their position within a larger structure, and that shape the probability distribution over outcomes available to them, making some category-property combinations more likely than others. We call this a *structural* construal.

Our aim in the current paper is to characterize the psychological signatures of adopting a structural construal, and to identify key similarities and differences between internalist and structural construals. In particular, we investigate whether some of the properties that have been taken to be hallmarks of an internalist construal are also compatible with a structural construal. The three hallmarks that we consider are supporting formal explanations (e.g., "he holds a poorly-paid job because he is an immigrant"), generics ("immigrants hold poorly-paid jobs"), and the generalization of properties within a category (e.g., assuming that because one category member holds a poorly-paid job, others are likely to do so as well). Below we suggest, and subsequently demonstrate experimentally, that all three of these characteristics hold for both kinds of construals. However, we also suggest – and go on to test – important ways in which internalist and structural construals diverge. On a structural construal, properties associated with a category are seen as less defining of a category and as more dispensable or malleable. A structural construal also supports the generalization of properties across kinds that share a social position (such as "immigrant"), whereas generalizations following an internalist construal instead track the kind. Below we review relevant work on internalist and structural construals before providing an overview of the three experiments we go on to report.

1. Prior work on internalist construals

Internalist construals have been widely documented, especially in the domains of social and natural kinds (Atran, Estin, Coley, & Medin, 1997; Bastian and Haslam, 2006; Gelman, 1988, 2003; Keil, 1992; Rangel & Keller, 2011; see also Gelman, 2013, and Kelemen & Carey,

* Corresponding author at: Psychology Department, Peretsman Scully Hall, Princeton University, Princeton, NJ 08544, United States of America.
E-mail address: nadezdav@princeton.edu (N. Vasilyeva).

https://doi.org/10.1016/j.cognition.2020.104383
Received 1 October 2019; Received in revised form 16 June 2020; Accepted 17 June 2020
Available online 06 July 2020

2007, on essentializing artifacts). Numerous theories of kind representation emphasize a tendency to look "within" the kind for deep, causally active, and explanatorily powerful factors that hold categories together, shaping and maintaining the properties of their members. This tendency can emerge from 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 (the inherence heuristic; Cimpian & Salomon, 2014). Internalist thinking 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 (Haslam, Rothschild, & Ernst, 2000). 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 (Cimpian, Mu, & Erickson, 2012; Leslie, Cimpian, Meyer, & Freeland, 2015; US National Academy of Sciences, US National Academy of Engineering, and US Institute of Medicine Committee on Maximizing the Potential of Women in Academic Science and Engineering, 2007). Overemphasizing internal, controllable causes of failure is also associated with low endorsement of governmental support programs (Kluegel, 1990).

Internalist construals have been associated with two particular linguistic forms: formal explanations and generic expressions. Formal explanations are explanations that appeal to category membership to explain a property (e.g., "Priya doesn't like math because she's a girl"). Gelman, Cimpian, and Roberts (2018) argue that such explanations reflect internalist beliefs, with the appeal to the category serving as a placeholder for unspecified inherent features. For example, participants found it more natural to accept an internalist explanation as a way to unpack a formal explanation (e.g., "It flies because it is a bird. More specifically, it flies because of deep internal features.") than the reverse (e.g., "It flies because of deep internal features. More specifically, it flies because it is a bird."). In another task, Gelman et al. compared inherent explanations with environmental explanations (e.g., "It flies because of its environment."), asking participants whether each explanation is "a more specific or a more general version" of a formal explanation. Inherent explanations were rated as "more specific" versions of formal explanations, whereas ratings for environmental explanations were not systematic. The authors conclude that "formal explanations can more easily be elaborated as inherent explanations than environmental explanations" (p. 56).

The second linguistic form, a generic expression, has received particular attention in association with internalist thinking. Generic expressions attribute a property to a category in general (e.g., "immigrants hold poorly-paid jobs"; "girls aren't good at math"), without specifying an individual or quantifying the claim. Even though a generic is sometimes judged appropriate when it describes a mere statistical association (e.g., "barns are red"; Prasada & Dillingham, 2006, 2009; see also Tessler & Goodman, 2019), the dominant view is that the default or paradigmatic interpretation of generics conveys that there is a deep, underlying nature to the kind, and that the attributed property is causally grounded in that nature or essence (e.g., Cimpian, 2010; Cimpian & Markman, 2009, 2011; Haslanger, 2011; Leslie, 2007, 2008, 2012, 2014, 2017). For example, Wodak, Leslie, and Rhodes (2015) write that "generics are by default understood…as being true in virtue of the intrinsic nature of the kind" (p. 627). Rhodes, Leslie, and Tworek (2012) argue that generic language is an important vehicle for transmitting essentialist beliefs about social groups across generations, and offer supporting evidence of a self-reinforcing cycle: learning about novel social groups through generics promotes essentialist beliefs about them, and essentialist beliefs in turn promote generic language. As argued by Bian and Cimpian (2017), generics come with a built-in "explanatory perspective," according to which the cited attributes are "core, non-accidental aspects of what the relevant groups are like deep down" (p. 23).

A third hallmark of internalist thinking is the robust generalization of a property within, but not necessarily across, the boundaries of a kind. If a category is defined by an essence that is causally responsible for a given property, then that property might be expected to extend to nearly all members of the kind, but not necessarily beyond. The reverse inference – inferring psychological essentialism from such patterns of generalization – has characterized much of the literature on essentialism. Haslam et al. (2000) refer to this aspect of essentialism (group homogeneity) as "entitativity," and assess it through the endorsement of statements such as: "Some categories contain members who are very similar to one another; they have many things in common. Members of these categories are relatively uniform." In the context of mental disorders, essentialist beliefs are assessed with items such as: "[the disorder] is a relatively uniform disorder, so that [people with the disorder] are very similar to each other" (Haslam & Ernst, 2002; see also Ahn, Flanagan, Marsch, & Sanislow, 2006). As a final example, Rhodes et al. (2012) include a measure of property generalization from one category member to other members in a composite measure of essentialization, with higher generalization scores leading to higher composite scores. These examples reflect a commitment to the following idea: that "the perception of a strong level of similarity (…) among group members (i.e. group entitativity) suggests the existence of a deep essence that would account for the detected regularities" (Yzerbyt, Corneille, & Estrada, 2001, p. 141). In other words, robust property generalization on the basis of category membership is taken as evidence of an internalist construal.

In the following section, we develop the alternative notion of a structural construal, and we suggest that these three hallmarks of internalist thinking – formal explanations, generic expressions, and generalization – are also consistent with a structural construal.

2. An alternative framework: structural construals

While internalist thinking has been explored extensively in the psychological literature, alternative ways of representing kinds have received much less attention. One important alternative is *structural thinking*, based on the notion of structural explanation developed in the philosophy of social science (Ayala & Vasilyeva, 2015; Ayala-López, 2018; Garfinkel, 1981; Haslanger, 2011, 2015; Langton, Haslanger, & Anderson, 2012; Ritchie, 2019; Vasilyeva & Ayala-Lopez, 2019). A structural construal of a category-property association represents the association as arising 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. Structures are organized wholes consisting of nodes (positions) and their relationships; each social position is defined through its relationship to other elements of the social structure (e.g., a subordinate role subject to particular constraints by virtue of its position within a hierarchy). Positions can be shared in the absence of shared intrinsic essences (Haslanger, 2000, 2015), and positions can differ across social structures and cultures. To illustrate, the generics "women don't drive," or "women are bad at math," could 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 categories and social positions are typically confounded within a given time and place, category-property associations could emerge for internalist and/or structural reasons. We suggest that formal explanations and generic claims imply a non-accidental connection between category membership and the attributed property, but permit both internalist and structural interpretations. Mirroring our introductory 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") while holding an internalist or structural construal. On an internalist construal, the property ("being in prison") is attributed to the category itself (e.g., presumed criminal inclinations). On a structural construal, the same property is instead attributed 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.; see Ritchie, 2019, for a relevant discussion).

Just as formal explanations and generics can support both internalist and structural construals, we suggest that patterns of within-category property generalization taken to support an internalist construal are also compatible with a structural construal. If category members are subject to the same set of stable structural constraints, they may end up displaying similar properties in the absence of a common inherent predisposition. For example, in a social structure with a gender wage gap, no guaranteed child care, and limited parental leave, women who have children may make similar "choices" in leaving their jobs or switching to part-time work – but this homogeneity of outcomes results from a particular choice architecture rather than from shared preferences (Cudd, 2006; Haslanger, 2015; Okin, 1989). Thus both internalist and structural explanations should support the extension of category properties to individual category members within a shared context, albeit for different reasons. When category and position are unconfounded, a divergence between internalist and structural construals should emerge, with an internalist construal supporting generalization within a kind and across social positions, and a structural construal showing the opposite pattern.

3. Prior work on structural thinking

Research on internalist construals has typically contrasted internalist thinking not with structural thinking, but with more broadly *externalist* thinking, where the external factors are not (all or mostly) structural. For instance, Gelman et al. (2018) employ a general pointer to external factors (e.g., "because of its environment"), which could be structural or not. Tworek and Cimpian (2016) employ explanations that appeal to idiosyncratic historical events (e.g., "Black is associated with funerals because of some historical or contextual reason – maybe because an ancient people originated the practice for some idiosyncratic reason and then spread it to many parts of the world," or brides wear white "just because of something that happened a long time ago": a really important Queen "just decided to wear a white dress to her wedding"; "I guess there is no real reason why brides wear white. It's not like there's anything special about white that makes it go with brides"). Likewise, open-ended responses citing one-time, external events are taken as representative examples of externalist explanations (e.g., "because he drinked that and it went into his bones"; Cimpian & Markman, 2011; [X is good at leeming] "because his dad teached him"; Cimpian & Erickson, 2012). Unlike structural factors, many of these considerations do not apply to the category as a whole, nor do they support within-category generalization. It thus remains a live option that hallmarks of internalist thinking (formal explanation, generics, and generalization) are equally compatible with structural thinking.

In fact, Vasilyeva, Gopnik, and Lombrozo (2018) – who report the first study designed to differentiate structural from internalist construals – find that both forms of construal support formal explanations. In their study, three-to-six-year-old children and adults learned about a novel association between category membership (being a girl) and a property (playing Yellow-Ball during school recess). Additionally, half of the participants received information about stable structural constraints that could explain the association: they learned that girls were assigned to a classroom with physical characteristics that reliably made one outcome (playing Yellow-Ball) more likely than another (playing Green-Ball). This context represented the fixed position occupied by category members, with its own constraints and affordances. The remaining participants learned about classrooms with no such constraints, inviting them to infer underlying preferences to explain the deviation of group choices from chance (Kushnir, Xu, & Wellman, 2010). Across several tasks, older children and adults generated and favored different kinds of explanations for the category-property association, and in the structural condition they rated a girl's game choice as more likely to change if her circumstances changed. Even 3-year-olds showed signs of early structural thinking. Importantly, though, despite this differentiation between internalist and structural construals, all age groups endorsed formal explanations (she plays Yellow-Ball "because she is a girl") equally across the structural and internalist conditions. These findings support our prediction that hallmarks of internalist thinking are compatible with a structural construal, but Vasilyeva, Gopnik, and Lombrozo (2018) did not investigate generic language, nor generalization of novel properties within versus across categories and positions.

Finally, outside of the social domain, research on category-based induction reveals the representational flexibility that structural thinking requires. For example, Shafto, Kemp, Bonawitz, Coley, and Tenenbaum (2008) asked participants how likely it was that a target organism had a specified property given that another organism had the property (e.g., "Carrots have the XD enzyme for reproduction. How likely is it that rabbits also have the XD enzyme for reproduction?"). Importantly, the two organisms could be related to each other taxonomically and/or within a food chain. When the property being generalized was physiological (e.g., have the XD enzyme), participants generalized properties more reliably the shorter the taxonomic distance, consistent with an internalist picture on which taxonomic similarity predicts similarity in essence (perhaps DNA). But when the property being generalized was a disease (e.g., carry the bacteria XD), participants tended to generalize *down* the food chain more than *up* the food chain, suggesting that structural position (in the food chain) informed judgments (see also Coley & Vasilyeva, 2010; Heit & Rubinstein, 1994; Medin, Coley, Storms, & Hayes, 2003; Ross & Murphy, 1999). Additional work has shown that these patterns of generalization are predicted by the explanation for a category-property association that a participant entertains (Vasilyeva & Coley, 2013; see also (Lombrozo and Gwynne, 2014; Sloman, 1994; Vasilyeva, Ruggeri, & Lombrozo, 2018), supporting the link between construal and generalization that we go on to test.

In sum, prior work on internalist construals of social categories has contrasted internalist construals with alternatives, but has not differentiated structural construals from other externalist forms of construal. Vasilyeva, Gopnik, and Lombrozo (2018) already show that at least under some conditions, adults are able to adopt a structural construal, and that this construal supports formal explanations as effectively as an internalist construal, but they do not investigate other hallmarks of internalist thinking. Finally, research outside the social domain suggests flexibility in categorical representations, with explanations for property-category associations tracking flexible patterns of generalization, but without a focus on structural explanations in the social domain. The studies we report below not only extend this prior work to more complex and realistic social categories and properties, but additionally test our predictions concerning all three putative hallmarks of internalist thinking articulated above: formal explanation, generic language, and generalization.

4. Overview of current studies

Across three studies, we test the hypotheses that both internalist and structural construals support formal explanations (Study 1), generics (Studies 2–3), and generalization when category membership and structural position are preserved (Study 3). We additionally test the hypotheses that along other dimensions, internalist and structural construals diverge: we expect different patterns of property generalization once category membership and structural position are allowed to vary independently (Study 3), different intuitions about using the property in category definitions (Study 1), and different judgments about true category membership when the property is removed (Study 1). As we elaborate in the General Discussion, testing these hypotheses not only furthers our understanding of structural thinking in the social domain, but also has implications for theories of formal explanations and generic language, for how we interpret prior results taken to support internalist construals, and for efforts to mitigate the harmful effects of stereotypes.

5. Study 1

In study 1, participants were introduced 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 varied whether the category-property connections were explained with an internalist explanation (e.g., appealing to group identity, as one example of an inherent characteristic that can hold across contexts[1]), were explained with a structural explanation (appealing to social position), or were incidental (the associations just happen to be true). To test whether this manipulation was successful in inducing different construals, we adapted measures originally developed in Prasada and Dillingham (2006, 2009) to differentiate "principled" and "statistical" connections, similar to the measures used 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), property dispensability ratings (i.e., whether an individual missing the property is still a true category member), and formal explanation evaluation (i.e., whether the presence of the property can be explained by appeal to category membership).

This experimental design has two primary aims. First, the study serves as a conceptual replication of Vasilyeva, Gopnik, and Lombrozo (2018), but with novel, experimentally-controlled social categories, and with a wide range of realistic properties matched in statistical association (namely cue and category validity, as explained below). Following Vasilyeva, Gopnik, and Lombrozo (2018), we expected both internalist and structural construals to support formal explanations (e.g., "He holds a poorly paid job because he's a Borunian"). Additionally, we expected the internalist and structural conditions to differ with respect to partial definitions and property dispensability. 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, resulting in higher endorsement of partial category definitions citing the property under the internalist than under the structural explanation. Likewise, removing an internalist feature should produce more damage to category membership than removing a feature contingent on external structure, translating into lower ratings of category membership once the property is removed in the internalist than the structural condition.

Our second aim was to strengthen our interpretation of the similarities and differences between internalist and structural construals through the inclusion of a control condition: the incidental features with incidental "explanations." For these features, we predicted a profile of effects different from either internalist or structural construals. Based on Prasada and Dillingham (2006, 2009), we expected that incidental features would not support definitions and would be seen as dispensable (like structural features), but that they would not support formal explanations (in contrast to both internalist and structural features). This finding would support the claim that *both* internalist and structural construals support formal explanations *more* than the incidental baseline condition, going beyond the predicted null effect of no difference between the internalist and structural construals.

5.1. Method

5.1.1. Participants

Seventy-seven participants (mean age 33, range 18–60; 38 identified as women, 39 identified as men) were recruited on Amazon Mechanical Turk in exchange for $1.50. In this and in subsequent studies, participation was restricted to workers with an IP address within the United States and with an approval rating of 95% or higher from at least 50 previous tasks. An additional 33 participants were excluded for failing an attention check, explained below.

In this and subsequent studies, target sample sizes were determined based on a priori power analyses (using G*Power) to detect a small to medium effect size for the target interactions with 0.95 power, based on the number of experimental conditions in a given study design.

5.1.2. Materials, design, and procedure

Participants read a short vignette introducing the novel social category of "Borunians" - a group of immigrants originally from Bo-Aaruna who settled in a fictional country, Kemi. Borunians were characterized by 18 unique features (see Table 1), all of which were introduced in generic form (e.g., "Borunians hold mostly poorly paid jobs."). These 18 features were presented in three blocks of 6, with each block presenting features of a single type: *internalist* (tying the feature to Borunian tradition and identity), *structural* (tying the feature to the structural constraints acting on Borunians due to their position within Kemi society), or *incidental* (roughly equivalent to Prasada and Dillingham's (2006) "statistical"). An additional norming study[2] verified that the three feature types did not differ in mean cue validity (i.e., probability of belonging to the category given the feature) or category validity (i.e., probability of having the feature given the category membership). Feature type was thus a within-subjects factor, with both blocks and items within blocks presented in a random order.

After learning the features, each participant made one of the following three judgments, illustrated for the feature "holds a poorly paid job":

[*Partial definition*] Question: What is a Borunian? Answer: A

[1] While many internalist explanations cite genetic or other biological factors, the relevant theoretical frameworks (e.g., psychological essentialism, the inherence heuristic) are not committed to exclusively biological explanations. Instead they emphasize the general tendency to look "within" the kinds for inherent, causally active, and explanatorily powerful factors. For biological kinds, these are likely to be biological; for social kinds, these may be biological or social. In everyday speech people use expressions such as "hard-core Republicans," "card-carrying humanists," "diehard Red Sox fans," "true hippies," and "Catholics to the bone" to suggest a deeply ingrained group identity without necessarily implying that it is genetic or biological. The internalist explanations in this study appealed to this group identity, while subsequent studies included a mix of biological and social internalist factors.

[2] For the norming study, 23 participants were recruited from Amazon Mechanical Turk as in Study 1; an additional 15 participants were excluded for failing an attention check like that in Study 1. Each participant read the same general backstory about Borunians, and saw 36 features comprising candidate internalist, structural, and incidental properties. After a short practice session, they completed either the category validity or the cue validity judgments for each of the 36 features. The category validity questions measured the perceived probability of a property given category membership, and were of the following form: "We randomly picked a hundred Borunians in Kemi. How many out of the 100 [have property P]? Please enter your best guess as a number from 0 to 100." The cue validity questions measured the perceived probability of the category membership given a property, and took the following form: "We randomly picked a hundred people in Kemi who [have property P]. How many out of the 100 are Borunians? Please enter your best guess as a number from 0 to 100." Based on these ratings we identified six features of each type that did not differ in mean category validity ($M_{\text{int}} = 83.60$, $M_{\text{str}} = 84.04$, $M_{\text{inc}} = 82.29$, linear mixed model $F(2,22) = 0.19$, $p = .832$) or in mean cue validity ($M_{\text{int}} = 66.30$, $M_{\text{str}} = 69.03$, $M_{\text{inc}} = 66.88$, $F(2,20) = 0.25$, $p = .782$). Additional analyses showed that the internalist and structural features did not differ in average word length ($M_{\text{int}} = 28.17$, $SD = 11.16$; $M_{\text{str}} = 31.67$, $SD = 15.17$; $p = .268$), but both the internalist and structural features were longer than the incidental features, $M_{\text{inc}} = 10.50$, $SD = 6.28$, p's $< .001$. The internalist and structural features also did not differ in mean length in the subsequent studies; the relevant analyses are available through the OSF.

Table 1
Examples of features used in Study 1, accompanied by the corresponding overview statements used to introduce each feature type. The complete list is available through OSF, https://osf.io/xzb74/?view_only=a5fb74c00275418f828a68aaa0af2ec0.

Feature type	Example
Internalist	Borunian traditions are extremely important to them, and form part of their identity: Borunians have a special tattoo on one arm.
Structural	Here are a few characteristics that Borunians have due to their position in 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 applicants for the same job include Kemi citizens who are equally or more qualified. Due to this regulation, Borunians hold mostly poorly paid jobs.
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.

> Borunian is a person who holds a poorly paid job. How good is this answer? (1 not good at all – 7 very good)
> [*Property dispensability*] 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)
> [*Formal explanation*] 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).

Judgment type was thus a between-subjects manipulation, with each participant making the same judgment for all 18 features, presented in random order. 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").

At the end of the study, participants completed an attention check: they verified a mix of nine features of Borunians as "True" or "False." Participants failing to answer all of these questions correctly were excluded from the final dataset.

5.2. 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 type was significant, $F(2,74) = 5.70, p = .005, \eta_p^2 = 0.133$, and the main effect of feature type was marginal, $F(2,148) = 2.99, p = .053, \eta_p^2 = 0.039$. However, of most theoretical importance was the significant interaction between judgment and feature type, $F(4,148) = 31.54, p < .001, \eta_p^2 = 0.460$.

As shown in Fig. 1, each feature type had a unique "profile" across the three judgments. As predicted, structural features supported definitions less strongly, and dispensability judgments more strongly ($ps < .001$), than did internalist features; however, structural and internalist features did not differ in the extent to which they supported formal explanations ($p = .327$). Also as predicted, incidental features did not support formal explanations as well as internalist features ($p = .003$) or structural features ($p < .001$). Incidental features also received lower definition ratings and higher dispensability ratings than internalist features ($ps < .001$); however, incidental features did not differ from structural features on the definition ratings ($p = .265$), although they were rated more dispensable ($p = .023$).

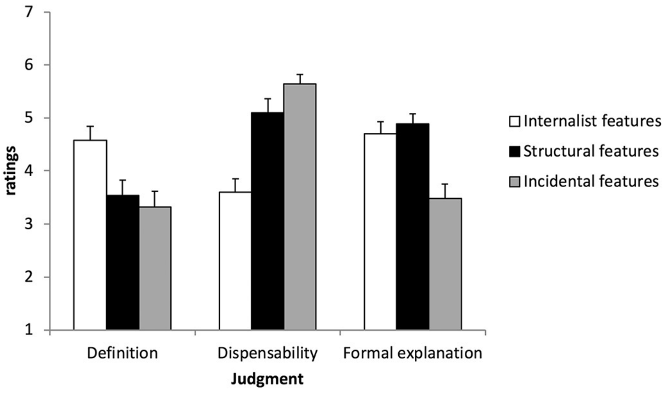

Fig. 1. Study 1: Partial definition, dispensability, and formal explanation ratings as a function of feature type. Error bars represent 1 SEM.

In sum, Study 1 succeeded in eliciting a different profile of judgments across feature types. Most crucially, structural features were differentiated from internalist features in being less defining and more mutable, but supported formal explanations equally well, and more successfully than incidental features.

Two additional features of the results from Study 1 are especially noteworthy. First, a structural pattern of responses was successfully elicited despite introducing category-property associations in the form of generic claims – the hallmark that we explore in Study 2. This was accomplished by offering appropriate cues in the feature description, but did not require explicit guidance or training in structural reasoning, suggesting that we tapped in to a form of reasoning that our participants found natural and appropriate. Second, it's 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 or to the social position it occupied. This is consistent with the idea that construals are closely related to the way in which a category-property association is explained, where these explanations guide generalization (see Gwynne & Lombrozo, 2014; Sloman, 1994; Vasilyeva & Coley, 2013; Vasilyeva, Ruggeri, & Lombrozo, 2018) – the hallmark we explore more systematically in Study 3.

6. Study 2

Study 1 succeeded in replicating key results from Vasilyeva, Gopnik, and Lombrozo (2018) with more realistic materials and with properties matched for cue and category validity. Importantly, we found that both internalist and structural construals support formal explanations, and that they do so to a greater extent than incidental relationships. Study 1 also provided indirect support for the idea that generics can support both internalist and structural interpretations: even though all participants learned about category properties through generic language (of the form "Cs have property P"), different explanations successfully induced different construals of the property-category associations.

In Study 2, our primary aim was to provide a more direct test of the claim that generic language supports a structural interpretation. To reiterate, we argue that generic statements are compatible with multiple construals: a given statement, such as "women have trouble getting tenure in mathematics," can be construed in internalist terms (women have this property inherently, by virtue of being women) or in structural terms (women have this property by virtue of their stable position in our society). In the former case, the property is attached to the social kind "woman"; in the latter case, it is attached to the stable social position that the kind occupies.

This study addressed two questions: First, do participants recognize both internalist and structural paraphrases of a given generic claim as

Table 2
Sample features and explanations used in Study 2.

Feature	Internalist explanation	Structural explanation
Hold poorly paid jobs	Reason: Borunians' inherent reluctance to take on hard and demanding jobs that may pay well, but also require over-time.	Reason: in order to hire a Borunian for a well-paid job, employers in Kemi are required to file complicated government paperwork
Quit jobs after having children	Reason: a nurturing nature and a predisposition to value family and time spent caring for children over career advancement.	Reason: employers are not required to provide paid parental leave or sponsor childcare for these immigrants.
Have very white teeth	Reason: a genetic predisposition to retain high calcium levels in teeth throughout the lifespan.	Reason: subsidized dental insurance plans that make maintenance and whitening procedures affordable to them.
Maintain healthy body weight	Reason: an effective metabolic system that burns fat fast, and efficiently removes toxins from the body.	Reason: special access to municipal subsidies to purchase organic vegetables directly from local farmers.
Have babies with low birthweight	Reason: a heritable predisposition for developing a small skeletal frame and low muscle mass.	Reason: ineligible for standard health insurance plans that cover adequate prenatal care.

acceptable interpretations of that generic? Second, can internalist and structural explanations successfully induce the corresponding interpretations? As an additional aim, we measured chronic or default preferences regarding generic interpretation – that is, how participants interpreted a generic in the absence of an explanation for the property-category association.

To accomplish these aims we presented participants with generic claims about novel groups of fictional immigrants, accompanied by either an internalist explanation, a structural explanation, or no explanation. Participants were then asked to evaluate the plausibility of different interpretations of the statements, described to participants as different "meanings" (we return to the question of generic "meanings" vs. interpretations in the General Discussion). These included an internalist meaning (implying that the property is an inherent characteristic of the social kind), a structural meaning (implying that the property is a product of stable structural constraints acting on a category), and an accidental-statistical meaning (simply stating that there's "no particular reason" why the category-property relationship holds).

We predicted that depending on the provided explanation, participants would rate different meanings of a given generic as most plausible. Specifically, we predicted that relative to other conditions, internalist explanations would promote the endorsement of internalist meanings, and that structural explanations would promote the endorsement of structural meanings. We also expected that a provided explanation would suppress the perceived plausibility of alternative meanings (e.g., providing an internalist explanation would suppress a structural meaning). The experiment and these predictions were preregistered on aspredicted.org, https://aspredicted.org/as8k4.pdf.

Finally, Study 2 also introduced a methodological modification relative to Study 1: rather than using different features matched in category and cue validity for each explanation type, the very same features were (across participants) presented with different explanations. These features were also more heterogenous, spanning from more behavioral (e.g., sell artisan souvenirs) to more biological (e.g., sunburn easily). These modifications helped us isolate effects of internalist versus structural explanations without confounding explanations with features, and also ensured that our effects were not restricted to features of a particular type.

6.1. Method

6.1.1. Participants

Forty-seven participants (age $M = 36$, range 20–66, 18 identified as women, 28 identified as men, 1 identified as non-binary gender) were recruited through Amazon Mechanical Turk and completed the study online in exchange for $1.50. An additional thirteen participants were excluded for failing the attention and/or memory check described below.

6.1.2. Materials, design and procedure

Participants first read a brief description introducing two novel social groups, Borunians and Aluns, who were forced by a military conflict to flee their respective countries, settling in the neighboring countries of Kemi and Oam (complete stimuli available through OSF). The order in which the two groups were introduced was counterbalanced.

Participants were familiarized with the meaning plausibility rating scale that they would use in the subsequent task with several examples. For instance, they read that for the statement "Borunians like orange," a plausible meaning is that "Borunians like the color orange," while an implausible meaning is that "Borunians like to climb orange trees." The training purposefully avoided the internalist vs. structural distinction relevant to the main task.

After this background and training, participants completed the primary task, which consisted of rating meaning plausibility for twelve generic statements attributing a property to Borunians or Aluns. For the sake of generality, properties ranged from the more social (e.g., "hold poorly paid jobs," "get college degrees from prestigious schools") to the more biological (e.g., "have babies with low birthweight," "have very white teeth"), and included items with both negative and positive valence (see Table 2 for sample features, and consult OSF for the full list). We used two social groups in this study to accommodate this range of properties, as some might appear contradictory if attributed to the same group (e.g., "getting college degrees from prestigious schools" and "holding poorly paid jobs").

Four of these generics were accompanied by an internalist explanation ("internalist explanation" condition), four were accompanied by a structural explanation ("structural explanation" condition), and four were not accompanied by any explanation ("explanation absent" condition; see Table 2 for sample explanations). The trials were blocked by condition, with the explanation absent block always presented first, and the internalist and structural blocks subsequently presented in a random order. Each participant saw a given feature in one explanation condition only, but across participants the features rotated through the conditions so that each feature appeared in all three explanation conditions.

On each trial, participants were presented with a single generic statement, with or without an explanation (depending on condition). Then they saw a schematic drawing of a person with a speech bubble containing that generic (e.g., "Now, someone says: Borunians have babies with low birthweight."). Underneath, they saw three potential meanings, illustrated below for the feature "low birth weight":

> Internalist meaning: "Borunians, by virtue of being Borunians, have babies with low birthweight";
> Structural meaning: "Borunians, by virtue of their position in Kemi society, have babies with low birthweight";
> "No reason" meaning: "Borunians, for no particular reason, have babies with low birthweight".

Participants were asked to rate each candidate meaning on a rating scale ranging from 1 (not a plausible meaning) to 7 (a very plausible meaning). The order of the structural and internalist meanings was counterbalanced across participants, with the "no reason" meaning always presented last.

After the main task participants completed a memory check (sorting 16 features and explanations based on whether they were mentioned in the survey) and an attention check (complying with a request to select a specific response option). Only participants who passed the attention check and correctly sorted at least twelve features/explanations were included in the final sample.

At the end, participants completed gender stereotyping and political orientation measures, included as exploratory measures to inform future studies (the analyses involving these measures are available on the OSF project page[3]), and they indicated their age and gender identification.

6.2. Results

A repeated measures ANOVA on meaning plausibility ratings as a function of explanation condition (internalist, structural, absent) and meaning type (internalist, structural, no reason)[4] revealed several significant effects. First, there was a main effect of meaning type, $F(2, 92) = 91.74$, $p < .001$, $\eta_p^2 = 0.666$: on average, the internalist meaning, $M = 5.17$, was rated as more plausible than the structural meaning, $M = 4.40$, and the "no reason" meaning was rated as the least plausible, $M = 2.89$, all $ps < .001$. Second, there was a main effect of explanation condition, $F(2, 92) = 16.77$, $p < .001$, $\eta_p^2 = 0.267$: averaging across meaning types, participants were more generous in their plausibility ratings in the explanation absent condition, $M = 4.50$, than in the structural explanation condition, $M = 4.09$, $p < .001$, or the internalist explanation condition, $M = 3.87$, $p = .001$ (the latter two also differed significantly, $p = .027$). Most important, however, was the significant interaction between explanation condition and meaning type, $F(4,184) = 40.79$, $p < .001$, $\eta_p^2 = 0.470$, illustrated in Fig. 2. To examine the interaction, we conducted a series of planned comparisons.

We first tested whether the internalist and structural meanings were the favored meanings under their corresponding construals. When an internalist explanation was presented, the internalist meaning was rated as more plausible than the structural meaning, $p < .001$, or than "no reason," $p < .001$. When a structural explanation was provided, the structural meaning was rated as more plausible than the internalist meaning, $p < .001$, or than "no reason," $p < .001$. Comparing across explanations, the internalist meaning was less plausible under the structural explanation than under the internalist explanation, while the structural meaning showed the opposite pattern, $ps < .001$. These findings support our key prediction that generic claims can support both internalist and structural construals, and that internalist and structural explanations can induce corresponding construals.

Comparisons to the explanation absent condition additionally revealed that both types of explanations boosted explanation-congruent meanings and suppressed explanation-incongruent meanings. The internalist explanation (relative to explanation absent) boosted the plausibility of an internalist meaning, $p = .002$, and suppressed the plausibility of a structural meaning, $p < .001$. The structural explanation (relative to explanation absent) boosted the plausibility of a structural meaning, $p < .001$, and suppressed the plausibility of an internalist meaning, $p = .002$.

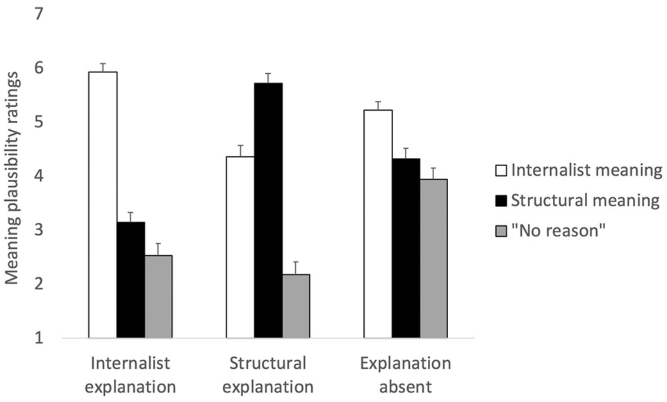

Fig. 2. Study 2: Plausibility ratings of different meanings of generic statements, as a function of explanation and meaning type. Error bars represent 1 SEM.

Comparisons to the "no reason" meaning also provide some interesting benchmarks. Not surprisingly, receiving an explanation of any type (i.e., a "reason" for the association) made the "no reason" meaning seem less plausible, relative to receiving no explanation, $ps < .001$. Of more theoretical interest is whether each explanation type suppressed the explanation-incongruent meaning to "no reason" levels. The answer is that they did not: under the internalist explanation the structural meaning was still judged more plausible than "no reason" ($p = .004$), and under the structural explanation, the internalist meaning was still judged more plausible than "no reason" ($p < .001$).

Finally, we examined whether in the absence of a provided explanation, some meanings are chronically more salient, and therefore judged more plausible. Comparisons of the three meanings in the explanation absent condition revealed that the internalist meaning was rated higher than the structural meaning ($p = .002$) and the "no reason" meaning ($p < .001$); the difference between the latter two did not reach significance, $p = .208$.

6.3. Discussion

The main goal of Study 2 was to test our proposal that generic claims can be interpreted in line with both internalist and structural construals. The results showed that generics can indeed accommodate both forms of construal, in particular when participants have access to the corresponding explanations.

Additionally, we found that internalist and structural explanations not only boost congruent interpretations, but also suppress alternative interpretations below the explanation absent baseline. However, they do not entirely rule out these alternative explanations (to the level of the "no reason" meaning). One explanation for these results is that there could be some level of compatibility between internalist and structural interpretations, consistent with prior literature on compatibility vs. discounting among different types of explanations (see Heussen, 2010; Lombrozo & Gwynne, 2014; Sloman, 1994). In the present case, internalist and structural considerations could potentially be integrated through a hybrid causal story – for instance, if participants posit internalist reasons to explain a structural association (e.g., that government subsidies focus on artisan products because Borunians are inherently skilled). Additionally, moderate ratings for explanation-incongruent interpretations could reflect uncertainty about which single explanation actually holds. We return to this point in the General Discussion.

Finally, we also observed that in the absence of a provided explanation, internalist meanings were rated more plausible than either the structural or "no reason" alternatives. This finding fits well with prior literature documenting internalist interpretations of generic language, and suggests that while structural interpretations can be induced with minimal cues, they are not the default or habitual interpretation of

[3] Neither gender stereotyping nor political affiliation predicted any of the response variables in Study 2 or Study 3, and did not participate in interactions with other predictors.

[4] Additional analyses showed that neither the order of meaning plausibility questions, nor the order of explanation blocks significantly affected the ratings, and did not interact with other variables, so all reported analyses collapse across these counterbalancing variables.

Table 3
Sample features and explanations used in Study 3. 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

generic claims attributing properties to social groups, at least under the conditions assessed by our task.

7. Study 3

Studies 1 and 2 showed that formal explanations and generic language, which have been previously proposed as signatures and/or promoters of internalist reasoning, are in fact compatible with a structural construal. The primary aim of Study 3 was to investigate the relationship between construal and *generalization*, another phenomenon that has been proposed as a hallmark of internalist (and in particular essentialist) thinking. Specifically, we tested the prediction that *both* internalist and structural construals support category-based generalization to individuals who share both the same category membership and the same structural position. However, we also predicted that because internalist and structural construals allocate different roles to category membership vs. social position in explaining an associated property, we should find different patterns of property generalization once category membership and structural position are allowed to vary independently. Under an internalist construal, properties should be generalized to individuals who share the same category membership (regardless of social position). Under a structural construal, properties should be generalized to individuals who share the same social position (regardless of category membership).

To test these predictions, we presented participants with information about the prevalence of a property in the Borunian population, as well as either an internalist explanation, a structural explanation, or no explanation for the category-property association. Participants then rated their endorsement of a corresponding generic claim ("Borunians [have property P]"), and generalized the properties in question to individual targets who varied both in category membership (same or different) and in social position (same or different), as shown in Table 4.

The property prevalence manipulation was included for two reasons. The first concerned the evaluation of generic claims. When a property is highly prevalent within a category, that alone can be sufficient to support generics, even in the absence of a clear internalist or structural construal (e.g., "barns are red"); in contrast, lower-prevalence associations support generic claims under more limited conditions (e.g., "mosquitoes carry the West Nile virus"; Leslie, 2008). By varying both explanation types and prevalence levels, and examining how these two factors jointly influence endorsement of generics, we increase the odds of identifying conditions under which the two construals diverge. In particular, if a structural construal does not support generics as effectively as an internalist construal at low-to-moderate levels of prevalence, our design would allow us to discover that this is the case. This provides a more conservative test of our hypothesis that internalist and structural construals behave similarly with respect to generic language.

The second reason for manipulating property prevalence concerned the property generalization task. Our key prediction was that an internalist construal would support generalization based on shared *category*, whereas a structural construal would support generalization based on shared social *position*. We thus predicted that in addition to general effects of prevalence (with greater generalization at higher levels of prevalence), the effect of prevalence would be moderated by explanation: under the internalist explanations, we predict a dampened effect of prevalence when generalization targets do not share category membership with the premise; in contrast, under the structural explanation, the effect of prevalence should be weakened when generalization targets occupy a different social position than the premise.

7.1. Method

7.1.1. Participants

One-hundred-and-fifty-seven adults (mean age 37, range 18–68; 76 identified as women, 80 identified men, 1 identified as agender) participated online in exchange for $1.50. An additional 30 participants were excluded for failing memory and attention checks.

7.1.2. Materials, design, and procedure

We developed a new set of twelve features describing a fictional immigrant category, Borunians, introduced as in Studies 1 and 2, and an internalist explanation and a structural explanation for each feature (see Table 3 for sample features and explanations; the complete set is available on OSF). For the sake of generality, we intentionally chose a range of internalist explanations spanning from more biological to those citing group preferences, values, and traditions (see further comments on this in the General Discussion).

Each participant was assigned to one explanation condition[5] (internalist, structural, or control), and completed two blocks of measures: generic truth ratings, and individual generalizations, in that order. 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),

[5] We switched to a between-subject manipulation of explanation in this study to reduce memory demands on participants, who were already tasked with remembering the prevalence level for each feature. This also serves to show that each construal can be induced without contrasting it with alternative construals.

Table 4
Study 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.

and High (M = 75%, range 71–80) levels. 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 illustrated in Table 4.

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 the four generalization questions (one from each row of Table 4), 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 gender stereotyping and political affiliation measures, as in Study 2; the analyses involving the latter two measures are reported through OSF.

7.2. Results

7.2.1. 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). 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 (χ^2 (1) = 426.10, $p < .001$): 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 0.25 (Low), 0.74 (Medium), and 0.94 (High). The effect of explanation condition was not significant, χ^2 (2) = 2.77, p = .250: the mean proportions of "true" responses across the internalist, structural, and no explanation conditions were, respectively, 0.66, 0.65, and 0.61. Finally, the interaction term between prevalence and explanation conditions was not significant, χ^2 (2) = 0.44, p = .805, indicating that explanation did not moderate the effect of feature prevalence.

7.2.2. Individual generalization

A hierarchical linear model predicting generalization to an individual 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 a significant interaction between regressors (model likelihood ratio = 11.34, p = .003). 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 (see Fig. 3). 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 (model likelihood ratio = 20.79, $p < .001$), revealing a stronger effect of shared social position in the structural condition than either the internalist, p = .014, or control condition, $p < .001$ (see Fig. 3). 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. Specifically, we predicted that a change in category membership would be more disruptive to the effect of prevalence in the internalist than in the structural condition, and that a change in social position would be more

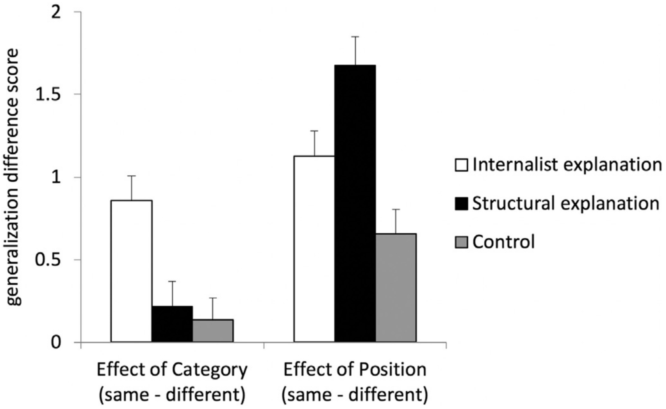

Fig. 3. Study 3: To represent the interactions between explanation condition, shared category membership, and shared social position, we created "generalization difference scores" (mean difference in generalization to individual in same vs. different category, and same vs. different social position). Error bars represent 1 SEM.

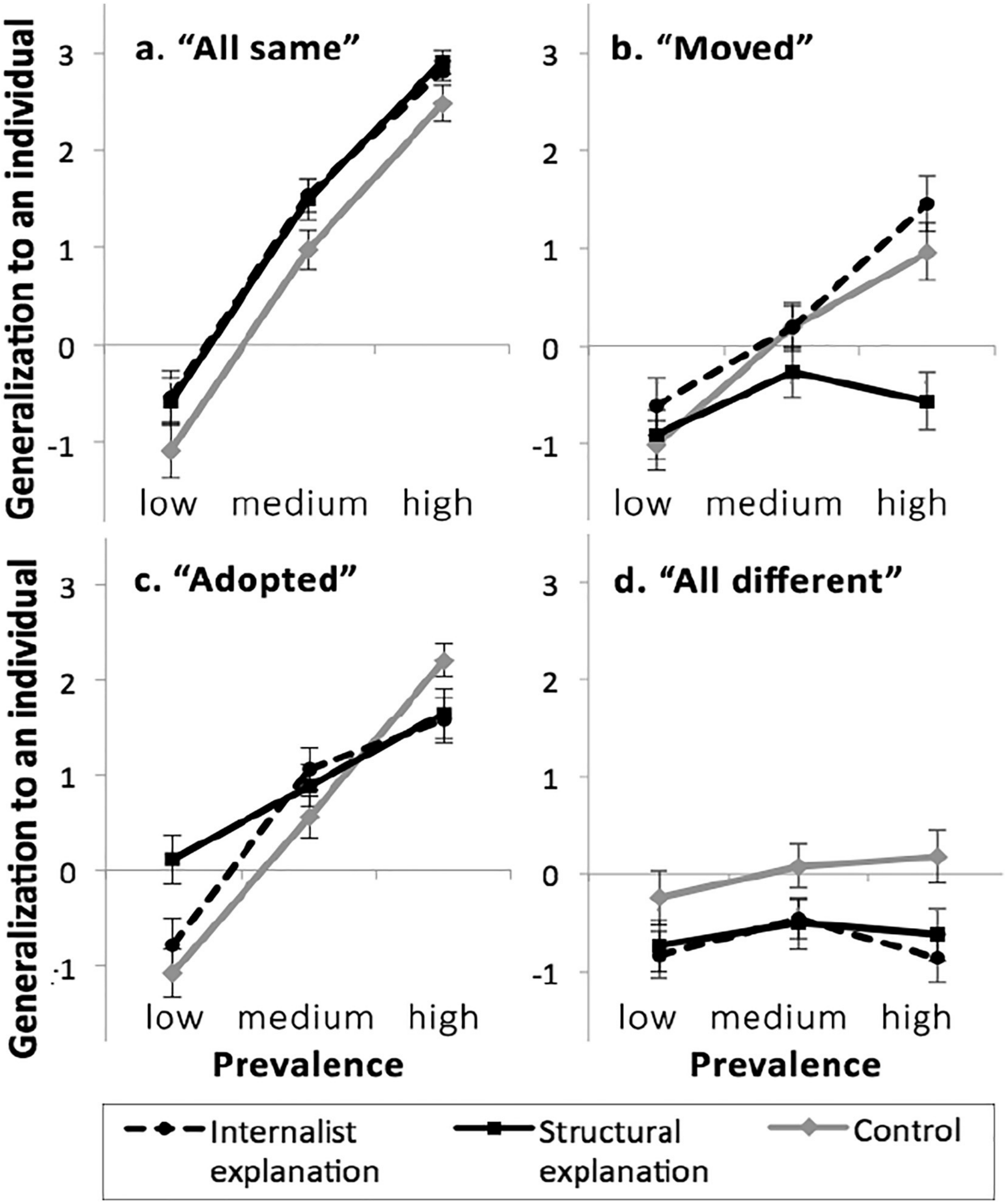

Fig. 4. Study 3: Mean individual generalization ratings as a function of within-category feature prevalence (binned into low, medium, and high ranges for presentation) and explanation type, split by the scenario (same or different category, and same or different social position). Error bars represent 1 SEM.

disruptive to the effect of prevalence in the structural than in the internalist condition. This made the two cells that crossed category membership and social position (“ADOPTED” and “MOVED”; see Table 4) the targets of analysis. We ran separate models for each cell, predicting individual generalization from prevalence and explanation condition (see Fig. 4).

In the “MOVED” scenario (Fig. 4, panel b), prevalence positively predicted generalization, β = 0.41, $p < .001$. Mirroring the results presented in Fig. 3, participants were also *less* likely to generalize in the structural condition than in the internalist condition, β = −0.45, $p < .001$, or in the control condition, β = −30, $p = .008$ (the latter two did not differ, β =0.15, $p = .181$). 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 (β = −0.31, $p = .002$) and control (β = −0.33, $p = .001$); the effect of prevalence did not vary across the latter two explanation conditions (β = 0.018, $p = .865$).

In the “ADOPTED” scenario (Fig. 4, panel c), prevalence also positively predicted generalization, β = 0.67, $p < .001$. However, the predicted interaction between prevalence and explanation, with an attenuated effect of prevalence in the internalist condition (relative to control), was only marginal, β = −0.19, $p = .0502$ – although, as predicted, at the low prevalence level, participants who received an internalist explanation were less likely to generalize than those in the structural condition, $t(103) = -2.34$, $p = .021$. Yet, contrary to our expectations, the effect of prevalence was also attenuated in the structural condition (relative to control), β = −0.33, $p = .004$. The extent to which the prevalence effect was attenuated, relative to control, did not differ across the two explanation conditions, $p = .117$.

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 the “ALL SAME” scenario

(Fig. 4, panel a), prevalence was a positive predictor of generalization, $\beta = 0.70$, $p < .001$, and both internalist and structural explanations boosted generalization relative to control ($\beta_{\text{Int}} = 0.20$, $p = .031$; $\beta_{\text{Str}} = 0.19$, $p = .036$); the internalist and structural explanations did not differ, $p = .945$. As predicted, there were no significant interactions, $ps \geq .780$.

In the "ALL DIFFERENT" scenario (Fig. 4, panel d), feature prevalence did not predict generalization, $\beta = 0.10$, $p = .162$. Participants were less likely to generalize in either explanation condition, relative to control ($\beta_{\text{Int vs. control}} = -0.39$, $p = .005$; $\beta_{\text{Str vs. control}} = -0.34$, $p = .013$); the two explanations did not differ, $p = .708$. As predicted, there were no significant interactions, $ps > .238$.

7.3. Discussion

Study 3 documented three important respects in which internalist and structural construals overlap. First, conceptually replicating Study 2 with a different task, Study 3 found that both internalist and structural construals support generics. Second, going beyond Study 2, Study 3 found that generic endorsement varied with the prevalence of the association, but that this effect was not moderated by explanation type. Third, Study 3 found that when generalizing to an individual with the same category and social position, internalist and structural explanations supported generalization to the same extent.

Importantly, the similarities between internalist and structural construals did not emerge because the explanation manipulation was ignored or otherwise ineffective. Once category and social position were unconfounded in the generalization task, the predicted differences emerged. 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. Overall, however, across scenarios, the patterns were consistent with our predictions, such that – for example – a structural construal made participants' generalizations less responsive to whether or not category membership was preserved.

We speculate that 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 (Fig. 3). Plausibly, the internalist condition could have been made even "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 primary goal in this study was instead to characterize the psychological profile of structural thinking as distinct from internalist thinking, we opted for greater experimental control over maximally representative features.

8. General discussion

Across three studies, we characterize structural thinking about social categories. Structural thinking offers a way to make sense of non-accidental associations between category members and their properties without attributing the properties to the inherent or essential nature of the category – that is, without adopting an internalist construal. Instead, people can adopt a structural construal, which accounts for observed correlations between properties and categories by citing stable external constraints. We show that adults can adopt a structural or internalist construal with minimal cuing, and that each construal has a characteristic profile of similarities and differences across our tasks. We begin by reviewing the similarities.

First, with regard to formal explanations, we found that internalist and structural explanations supported such explanations to the same extent, and to a greater extent than mere statistical associations. Second, and in tension with prominent views of generic language (Cimpian & Markman, 2011; Leslie, 2007; Rhodes et al., 2012; Wodak et al., 2015), generic language does not necessarily convey nor induce essentialist beliefs. In Study 1, the generic language that introduced a category-property association did not prevent an alternative construal. In Study 2, participants endorsed internalist and structural interpretations flexibly, depending on the provided explanation. And in Studies 2 and 3, both construals supported the endorsement of generic claims. Third, we found that both internalist and structural construals support generalization, producing identical patterns of generalization when a social category and a social position are confounded.

Despite these similarities across internalist and structural construals, we predicted (and found) important differences across construals as well. First, we found that properties were seen as less defining of a category under a structural construal relative to an internalist construal (Study 1). This finding is consistent with the results of Vasilyeva, Gopnik, and Lombrozo (2018). Second, going beyond prior work, we found that properties were regarded as less dispensable to category membership under an internalist construal than a structural construal (Study 1). Third, we found that when the typical confound between category and social position was disrupted, each construal led to a unique pattern of generalization: an internalist construal promoted generalization based on shared category membership, while a structural construal promoted generalization based on shared social position (Study 3).

Our studies challenge the widespread assumptions that formal explanations, generic expressions, and within-category generalization are unique hallmarks of internalist thinking. While these hallmarks may successfully differentiate an internalist construal from some externalist construals (e.g., appealing to idiosyncratic or unstable external factors), they do not differentiate an internalist construal from a structural construal. Our findings thus have important methodologic implications – namely, they warn against using these judgments as measures of essentialism. More striking, however, are the potential implications for our understanding of generic language and its role in perpetuating essentialist beliefs (e.g., Haslanger, 2011; Leslie, 2017; Rhodes et al., 2012; Roberts, Ho, & Gelman, 2017; Wodak et al., 2015). While there have been calls to reign in essentialist thinking by limiting generic language (e.g., Leslie, 2017), the prospects for achieving this linguistic change are rather grim. Replacing generics with overt quantified statements ("all", "most", "some") or ambiguous statements ("*they* are good at X"), for example, is unlikely to work since these are often misinterpreted and mis-remembered as generics by children and adults (Hollander, Gelman, & Star, 2002; Leslie & Gelman, 2012; Mannheim, Gelman, Escalante, Huayhua, & Puma, 2011).[6] The alternative approach of negating a generic ("women do NOT quit their jobs after having children") does not question the presupposition that "the relevant kind has a distinctive nature or ideal" (Wodak et al., 2015, p. 631). Moreover, a straight negation could have the appearance of going against the data (when the objective property prevalence is in fact relatively high within the category), discrediting the speaker.

In contrast to these approaches, adopting a structural construal of a

[6] A related strategy might be to draw attention to the statistics, such as the precise level of feature prevalence. This strategy could be effective for a generic involving a striking property (a la "sharks attack people") rather than a characteristic property (a la "sharks have fins"). For example, Wodak et al. (2015) suggest a possible reaction to the generic "Muslims are terrorists": asking "What percentage of Muslims commit terrorist acts?". But when the property prevalence is in fact fairly high, relative to other categories, and warrants a characteristic generic, appealing to the statistics could backfire if it prompts the listener to construct an internalist explanation for the high prevalence.

generic successfully explains a real association between a property and a category, but without promoting essentialist beliefs. Could such a construal mitigate the effects of generic language on essentialist thinking? In the reported studies, a simple verbal explanation manipulation proved effective, suggesting that when we encounter generics in speech, offering a structural explanation – or, perhaps, merely suggesting that structural factors may be relevant and should be considered – could help block internalist inferences, potentially even in young children (Vasilyeva, Gopnik, & Lombrozo, 2018; Vasilyeva & Lombrozo, 2020).

Our findings also contribute a new position to philosophical debates on the role of social generics in sustaining oppression and injustice. Mirroring the proposals reviewed above, Haslanger (2011) and Langton et al. (2012) have argued that we ought to reject or negate the truth of racial and gender generics when we hear them (because they misrepresent social artifacts as essences), and convey information about corresponding statistical associations using quantified statements instead. However, Ritchie (2019) offers an insightful argument against a general prohibition on racial and gender generics in favor of quantified statements, since generic generalizations convey systematic, lawlike regularities (Carlson, 1995; Nickel, 2017), and are thus particularly well-suited for describing systematic patterns of structural oppression. Overt quantified statements, instead, can convey a misleading idea that the association could be due to accidental factors, masking the systematic nature of oppression. Ritchie's argument defending – and in some cases, prescribing – the use of social generics is also based on her questioning whether generics necessarily attribute intrinsic features to a group. Ritchie identifies examples of generics (e.g., “Blacks face economic, legal, and social discrimination”; “Women are expected to want children”) that attribute clearly structural properties to a category, and thus avoid the essentialist worry (see also Haslanger, 2000, on shared social positions in the absence of shared essences, and related acknowledgments of the role of external factors in shaping associations that generics convey in Rhodes & Mandalaywala, 2017; Nickel, 2017).

We are sympathetic with Ritchie's arguments, but we add yet another position to the debate. While Ritchie focuses on the content of some properties to make the case that not all generics attributing properties to social groups essentialize, we make a broader claim that when a social category and a social position are confounded (as they are in most real-life contexts), the source of an association between a property and a category/social position is under-determined. Specifically, the association is compatible with at least two causal-explanatory models, attributing the properties either to the inherent nature of the social group, or to the stable social position it occupies. The practical and political implications of our account are thus different from those of Leslie (2008), Haslanger (2011), Langton et al. (2012), or Ritchie (2019). Instead of recommending that we abandon social generics, negate them, replace them with quantified statements, or focus on promoting the use of select generics to highlight systemic oppression, we draw attention to the importance of (re)explaining generics in order to disambiguate them and convey an appropriate structural meaning (see Vasilyeva & Ayala-Lopez, 2019, for a more detailed outline of this position).

Our focus on the crucial role of explanation also sets our position apart from Noyes and Keil (2019), who make the related claim that generics do not necessarily signal essentialism. However, their proposal is that generics convey the information that the category in question is a genuine kind rather than a shallow collection, such as “white things.” They operationalize “kind-ness” using measures of group homogeneity across superficial differences/similarities, the uniformity of properties across members, and formal explanation (but see Nickel, 2017 for a discussion of true generics concerning “non-kind collections” rather than kinds, p. 440). Our proposal is not incompatible with Noyes and Keil's, but it requires a different notion of kinds, focusing on causal-explanatory regularities sustaining the observed category-property associations, rather than within-kind homogeneity. In fact, a structural construal can be powerful precisely because it explains observed correlations without making the homogeneity assumption: it reveals how the generics “women are expected to want children” (Ritchie, 2019; Saul, 2017) and “women choose part-time work after having children” can hold true even in the face of extreme variability in the desires, interests, preferences, and priorities of individual women.

Our generalization results also have potentially important theoretical and social implications. First, 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. This reinforces the insight from prior work on flexible inductive inference, showing that generalization tracks a broad range of dimensions that go beyond category-based similarity, suggesting that generalization is not a reliable indicator of essentialist representations (Coley & Vasilyeva, 2010; Heit & Rubinstein, 1994; Medin, Coley, Storms & Hayes, 2003; Ross & Murphy, 1999).

Second, from a practical standpoint, getting a fuller picture of how people generalize from categories to individuals has important real-life implications. For example, when a woman applies for a position where statistics suggest that women are likely to fail or drop out, the hiring committee might realize that the property of dropping out is a product of the social position rather than of being a woman per se. If they create a favorable social environment for the woman in their organization, the property of dropping out may not generalize to her. Although we have not offered direct evidence that a structural construal is capable of shifting such real-life decisions, our findings suggest that this is a promising direction to pursue.

Our findings raise a number of important questions for future research. First, although we show that both internalist and structural construals support generic claims, it is unclear whether the generics are *ambiguous* with respect to construal, or merely underspecified.[7] On the former view, generics are ambiguous because category labels (e.g., “women”) can refer either to an (essentialized) category or to a social position. Indeed, there are live debates about how to define terms such as “woman,” with some focusing on internal characteristics, but others, such as Haslanger (2000), defining a woman as someone who is systematically socially subordinated on the basis of presumed female sex (that is, in terms of social position). Prior accounts of generic meaning have identified related forms of ambiguity: for example, one can endorse both “boys cry” and “boys don't cry,” treating the term “boys” in either a descriptive sense (boys in fact do cry) or a normative sense (“real” boys shouldn't cry; Knobe, Prasada, & Newman, 2013; Leslie, 2015).

Another possibility, however, is that generics themselves are not ambiguous, but are rather underspecified with respect to an internalist or structural construal. That is, generics could have a single meaning – for instance, indicating a non-accidental connection between kind membership and having the property, or expressing a certain type of generalization – which is amenable to multiple causal-explanatory construals. To illustrate, the fact that wood burns can be explained by appeal to phlogiston, or by appeal to modern chemistry, but the fact that multiple candidate explanations are available hardly makes the generic “wood burns” ambiguous.[8]

Establishing whether the meaning of a generic claim is ambiguous or merely underspecified is a question for future research, but on either account, explanations are a natural mechanism for imposing a structural or internalist construal. When asking “why do women dedicate a lot of time to childcare?”, for example, the proffered explanation will constrain the interpretation of the generic explanandum “women

[7] We are grateful to Sandeep Prasada and Michael Strevens for encouraging us to clarify these points.

[8] We owe this example to Sandeep Prasada.

dedicate a lot of time to childcare" such that it is understood as a claim about the category or its position. Structural explanations, in particular, have been argued to shift the explanandum in a number of ways (Haslanger, 2011; Garfinkel, 1981; Skow, 2018). If a person strictly commits to the internalist explanandum - "why do women, an essentialized kind with an inherent underlying nature, have this property?" – the structural explanation will not answer their intended question. It will instead answer a different question, such as "what is it about the social position occupied by women that explains the association with this property?" (swapping a question about categories with a question about positions), or "why is being a woman related to having this property?" (swapping a request for a "triggering cause" with a request for a "structuring cause," which makes it the case that some relationship exists; (Dretske, 1988; Skow, 2018). If this is right, then generic explananda that support internalist vs. structural construals ("why do women dedicate a lot of time to childcare?") are importantly different from more canonical generic explananda that support multiple explanations ("why does wood burn?"). In the former case, different explanations involve subtly different explananda, whereas in the latter case, the candidate explanations are more naturally understood as different answers responding to the same explanandum.

Another open question concerns the relative status of structural and internalist construals. Even if the former can be readily cued, is the latter a privileged default? The results of Study 2 suggest this may be so: in the absence of a provided explanation, internalist meanings of generics were rated as more plausible than structural and purely accidental ("no reason") interpretations. Likewise, in Vasilyeva, Gopnik, and Lombrozo (2018), internalist explanations were somewhat more robust across development than structural explanations (although adults generated both types of explanations flexibly and at similar rates when cued with appropriate contexts). This raises important questions about the cognitive and environmental factors that shape the adoption of a particular construal in a given case. Are internalist construals less cognitively demanding, or thought to be more likely? What determines the probability that a given construal will be adopted in a given case? And are there additional construals that ought to be considered (see work on "principled connections," Prasada & Dillingham, 2006, 2009, for a candidate)?

Another important direction for further research is examining how internalist and structural construals interact in everyday thought. In our studies we specifically aimed to document "clear cases" of structural reasoning and contrast it with internalist reasoning. In reality, the stories people tell are likely to be more complicated, interweaving internalist and structural factors. For example, internalist factors could be seen as driving structural factors (e.g., viewing the current stable social positions as perfect "niches" matching the inherent, internalist predispositions of category members), or the other way around (e.g., viewing early social environment as the source of deep, inherent properties; see Rangel & Keller, 2011). Such causal stories might also include complex feedback loops sustaining the status quo. Examining the ways such hybrid causal-explanatory stories work, along with their social consequences, should be prioritized in future work.

A fourth open question is whether and how category and/or property type could moderate the effects we report here. In particular, "cultural" properties of social groups, such as food or religious customs, may have a default structural interpretation. However, many aspects of culture, including preferences, values, and attitudes, can be understood in internalist terms, where cultural properties reflect shared internal characteristics. Consistent with this, in Study 2 we were able to manipulate the structural vs. internalist interpretation of cultural properties, using the corresponding explanations of properties. When multiple construals can be induced for a given category-property association, the consequences of each construal may still vary across categories. For instance, "biogenetic essentialism" has mixed effects on attitudes to mental illness, reducing personal responsibility and blame, but increasing pessimism about recovery and increasing social distance (Kvaale, Haslam, & Gottdiener, 2013; Phelan, 2005). It will be important to identify whether structural thinking similarly licenses negative inferences, and how its positive and negative consequences interact with the specific domain.

Finally, structural thinking should be examined beyond the social domain. It is relevant in any domain involving complex systems composed of interacting elements, where the overall state of the system imposes constraints on the states of its elements.[9] Such complex systems are common in science, economics, and other disciplines. Underdeveloped structural reasoning skills might present obstacles to mastering these domains. In fact, psychological essentialism has been shown to interfere with learning a number of biological concepts, such as "species" and "evolution" (Coley & Tanner, 2012; Leslie, 2013; Lombrozo, Thanukos, & Weisberg, 2008; Shtulman & Schulz, 2008); it is important to examine whether structural thinking can generally promote understanding of complex patterns across a range of domains.

In sum, across three 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 position within a larger structure. Both kinds of representations can support formal explanations, generic claims, and generalization, effectively tracking environmental statistics. However, they work differently, with differences that manifest in category definitions, property mutability, and generalization across categories vs. positions. In practice, both internalist and structural construals are likely to be useful, since each captures a real aspect of the environment. Effective agents should thus track both internalist and structural relationships, and as psychologists, we should study both kinds of construal as well.

CRediT authorship contribution statement

Nadya Vasilyeva:Conceptualization, Methodology, Formal analysis, Funding acquisition, Writing - original draft, Writing - review & editing.**Tania Lombrozo**:Conceptualization, Methodology, Funding acquisition, Supervision, Writing - review & editing.

Acknowledgements

This work was supported by the National Science Foundation (Grant SMA-1730660). We thank Sandeep Prasada and two anonymous referees for their useful suggestions.

References

Ahn, W.-K., Flanagan, E. H., Marsch, J. K., & Sanislow, C. A. (2006). Beliefs about essences and the reality of mental disorders. *Psychological Science, 17*(9), 759–766.

Atran, S., Estin, P., Coley, J., & Medin, D. (1997). Generic species and basic levels: Essence and appearance in folk biology. *Journal of Ethnobiology, 17*(1), 17–43.

Ayala, S., & Vasilyeva, N. (2015). Explaining injustice in speech: Individualistic or structural explanations. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.). *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Cognitive Science Society: Austin, TX.

Ayala-López, S. (2018). A structural explanation of injustice in conversations: It's about norms. *Pacific Philosophical Quarterly, 99*(4), 726–748.

Bastian, B. & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. Journal of Experimental Social Psychology, 42, 228–235.

Bian, L., & Cimpian, A. (2017). Are stereotypes accurate? A perspective from the cognitive science of concepts. *Behavioral and Brain Sciences, 40*, Article E3. https://doi.org/10.1017/S0140525X15002307.

Carlson, G. (1995). Truth conditions of generic sentences: Two contrasting views. In G. Carlson, & F. J. Pelletier (Eds.). *The generic book* (pp. 224–237). Chicago: Chicago University Press.

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

[9] Some such systems may also display a relatively stable assignment of elements to their structural positions, thus creating a confound between the nature of the element and its role within the system, and making such systems more similar to the social systems examined so far, and opening the door to multiple construals of generics, etc.

Cimpian, A., & Erickson, L. C. (2012). The effect of generic statements on children's causal attributions: Questions of mechanism. *Developmental Psychology, 48*(1), 159.

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition, 113*(1), 14–25.

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., Mu, Y., & Erickson, L. C. (2012). Who is good at this game? Linking an activity to a social category undermines children's achievement. *Psychological Science, 23*(5), 533–541.

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.

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: Cognitive principles and the development of biology thinking. *CBE—Life Sciences Education, 11*(3), 209–215.

Coley, J. D., & Vasilyeva, N. Y. (2010). Generating inductive inferences: Premise relations and property effects. *Psychology of learning and motivation. Vol. 53. Psychology of learning and motivation* (pp. 183–226). Academic Press.

Cudd, A. (2006). *Analyzing oppression.* Oxford University Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

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

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*(1), 65–95.

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

Gelman, S. A. (2013). Artifacts and essentialism. *Review of Philosophy and Psychology, 4*(3), 449–463. https://doi.org/10.1007/s13164-013-0142-7.

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., & Ernst, D. (2002). Essentialist beliefs about mental disorders. *Journal of Social and Clinical Psychology, 21*(6), 628–644.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*, 113–127.

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs, 34*(1), 31–55.

Haslanger, S. (2011). Ideology, generics, and common ground. In C. Witt (Ed.). *Feminist metaphysics: Explorations in the ontology of sex, gender and the self* (pp. 179–208). Dordrecht: Springer.

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

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 411.

Heussen, D. (2010). When functions and causes compete. *Thinking & Reasoning, 16*(3), 233–250.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology, 38*(6), 883.

Keil, F. C. (1992). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. *In Creations of the mind: Theories of artifacts and their representation,* 212–230.

Kluegel, J. R. (1990). Trends in Whites' explanations of the Black-White gap in socioeconomic status, 1977–1989. *American Sociological Review, 55*(4), 512–525.

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition, 127*(2), 242–257.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*(8), 1134–1140.

Kvaale, E. P., Haslam, N., & Gottdiener, W. H. (2013). The "side effects" of medicalization: A meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review, 33*(6), 782–794.

Langton, R., Haslanger, S., & Anderson, L. (2012). Language and race. In G. Russell, & D. G. Fara (Eds.). *Routledge companion to the philosophy of language* (pp. 753–767). New York: Routledge.

Leslie, S. J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford Studies in Epistemology, 4*(108), 9.

Leslie, S. J. (2015). "Hillary Clinton is the only man in the Obama administration": Dual character concepts, generics, and gender. *Analytic Philosophy, 56*(2), 111–141.

Leslie, S. J. (2017). The original sin of cognition: Fear, prejudice and generalization. *The Journal of Philosophy, 114*, 393–421.

Leslie, S.-J. (2007). Generis and the structure of the mind. *Philosophical Perspectives, 21*(1), 375–405.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *The Philosophical Review, 117*(1), 1–49.

Leslie, S.-J. (2012). Generics. In G. Russell, & D. G. Fara (Eds.). *Routledge companion to philosophy of language (pp. 355–367)*. Routledge.

Leslie, S.-J. (2014). Carving up the social world with generics. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.). *Oxford studies in experimental philosophy* (pp. 208–232). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198718765.003.0009.

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

Leslie, S. J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology, 64*(3), 186–214.

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

Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for accepting evolution. *Evolution: Education and Outreach, 1*(3), 290.

Mannheim, B., Gelman, S. A., Escalante, C., Huayhua, M., & Puma, R. (2011). A developmental analysis of generic nouns in Southern Peruvian Quechua. *Language Learning and Development, 7*, 1–23.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. L. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*(3), 517–532.

Nickel, B. (2017). Generics. In B. Hale, C. Wright, & A. Miller (Eds.). *A companion to the philosophy of language* (pp. 437–462). Malden, MA: Wiley Blackwell.

Noyes, A., & Keil, F. C. (2019). *Generics designate kinds but not always essences: Implications for social categories and kinds.* (under review).

Okin, S. (1989). *Justice, gender and the family.* NY: Basic Books.

Phelan, J. C. (2005). Geneticization of deviant behavior and consequences for stigma: The case of mental illness. *Journal of Health and Social Behavior, 46*(4), 307–322.

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

Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science, 33*(3), 401–448. https://doi.org/10.1111/j.1551-6709.2009.01018.x.

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.

Rhodes, M., & Mandalaywala, T. (2017). *The development and developmental consequences of social essentialism. Wiley Interdisciplinary Reviews. Cognitive Science*Article e1437. https://doi.org/10.1002/wcs.1437.

Ritchie, K. (2019). Should we use racial and gender generics? *Thought,* 1–9. https://doi.org/10.1002/tht3.402.

Roberts, S. O., Ho, A. K., & Gelman, S. A. (2017). Group presence, category labels, and generic statements influence children to treat descriptive group regularities as prescriptive. *Journal of Experimental Child Psychology, 158*, 19–31. https://doi.org/10.1016/j.jecp.2016.11.013.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*(4), 495–553.

Saul, J. (2017). Are generics especially pernicious? *Inquiry,* 1–18.

Shafto, P., Kemp, C., Bonawitz, E. B., Coley, J. D., & Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition, 109*(2), 175–192.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32*(6), 1049–1062. https://doi.org/10.1080/03640210801897864.

Skow, B. (2018). *Causation, Explanation, and the Metaphysics of Aspect*. Oxford University Press.

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

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review, 126*(3), 395.

Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science, 27*(8), 1109–1122.

National Academy of Sciences (US), National Academy of Engineering (US), and Institute of Medicine (US) Committee on Maximizing the Potential of Women in Academic Science and Engineering (2007). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering.* Washington, DC: National Academies Press (US).

Vasilyeva, N., & Ayala-Lopez, S. (2019). Structural thinking and epistemic injustice. In R. Sherman, & S. Goguen (Eds.). *Overcoming epistemic injustice: Social and psychological perspectives* (Rowman & Littlefield International).

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*. Cognitive Science Society: Austin, TX.

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

Vasilyeva, N., & Lombrozo, T. (2020). When generic language does not promote psychological essentialism. Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society.

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.

Wodak, D., Leslie, S.-J., & Rhodes, M. (2015). What a loaded generalization. *Generics and social cognition, 10*(9), 625–635. https://doi.org/10.1111/phc3.12250.

Yzerbyt, V., Corneille, O., & Estrada, C. (2001). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review, 5*(2), 141–155.