**First-Year Project (FYP) Report**

Essentially blocked:   
Can structural factors block the essentialist effects of formal explanations?

Marianna Zhang

**Introduction**

***Categorization***

Universally, people impose structure on the world by grouping individual things into categories. Human beings everywhere from a young age naturally and easily organize their otherwise messy experiences using categories. We group individual things together, treating them as alike in some respect, and together different from other things in the world. We consider an individual thing not just as an individual thing, but as one of potentially indefinitely many instances of a category to which it belongs. Categorization is a universal and powerful cognitive tool that allows us to group various individual things together and treat the individuals as the same in some respect. Our ability to categorize allows us to generalize beyond our immediate experience as to make inductive inferences about other category members we have never experienced. For an example, my representation of cat allows me to make inferences about a new cat that I’ve never encountered: it probably meows, likes catnip, and fights with dogs. Although psychologically powerful and pervasive, categorization can also lead to pernicious consequences, especially in the social domain, where they can form the basis of stereotyping and prejudice.

What is the structure of these rich category representations, and how do children develop this structure? I will contrast two construals of how a given category structured: an essentialist construal and a social-structural construal. This project will use gender categories as a case study, since gender is a rich space to pit these construals against each other.

***Two construals of category structure***

One way we could represent the structure of a given category is to construe a category in essentialist terms. Psychological essentialism is a cognitive bias to believe that categories represent real natural distinctions in the world that carve nature at its joints. Each category is thought to have its own unique essence: for an example, there might be a “male essence” for the category of men, and “female essence” for the category of women. All category members are thought to share this deep underlying internal “essence”, which is stable and unchanging, and provides individual members with their identity. The category essence is thought to cause shared obvious and non-obvious properties among category members. These properties are fixed at birth, such that all category members have the innate potential to exhibit the shared properties, and stable over time. As a result, the category essence produces fundamental similarities among category members, and differences compared to non-category members (Gelman 2003). The category essence could but need not be elaborated or cashed out, eg as biological sex, DNA, chemical formula, etc. One could have a placeholder essence and believe that a category possesses an essence without knowing what the essence is.

In contrast to essentialism, another framework we could use to represent the structure of a category is a social-structural construal. On this view, categories are situated in a rich social context, which influence the very nature of the category. For an example, the gender categories male and female are situated within a social context, which provides a social meaning for what it means to be male and what it means to be female.[[1]](#footnote-1) Different social-structural forcesoperating on each category contribute to differences between categories, such that category membership is confounded with social context. In the case of gender, men and women might be conceptualized as subject to different expectations, standards, socialization, resources, biases, etc. in our society. The nature and structure of a category could be interpreted as contingent or even constructed by the social context that it is situated within.

***The role of social input for category construal***

We can contrast these two different construals of categories, comparing them as different ways that people think about categories. (We will bracket off the metaphysical question of what categories in the world actually are, and only consider the psychologically tractable question of how people conceptualize categories.) When children are developing category representations, how do they choose between these two construals of a given category? I will present these two construals as opposing and distinct construals for illustration, but we can consider these two construals to be two points on a theoretical continuum, with various tenable intermediate representations that vary across categories (eg some categories essentialized, some categories represented in a social-structural way), across properties (eg overhypotheses about some properties likely to be essentialized, some properties likely to be subject to social-structural factors), and across properties within a category (eg some properties of a category considered essentialized, some properties subject to social-structural factors).

Although children and adults across various cultural communities essentialize some set of categories, the categories that are essentialized are only a small subset of all the categories that they represent, and there is substantial cultural variation in what particular social categories are essentialized, which together suggest a role for social input in guiding children to essentialize particular categories and not others (Gelman 2003). Evidence for variation comes from a variety of cultural communities. Children growing up in more politically conservative communities in the US are more likely to essentialize race than children growing up in more politically liberal communities in the US (Rhodes & Gelman 2009). Children growing up in more religious communities in Israel are more likely to essentialize ethnicity than children growing up in more secular communities in Israel (). In India, upper-class adults are more likely to essentialize ethnicity than lower-class adults, and the female category is more likely to be essentialized than the male category (Mahalingam, 2003a). Cultural variation in what categories are essentialized suggests that essentialism is the product of a general cognitive heuristic towards essentialism and social cues about what categories to construe in essentialist ways.[[2]](#footnote-2) In particular, subtle cues in language that guide children to essentialized particular categories have been a topic of much research in the field.

*Generic language as a cue for essentialism*

One type of linguistic cue for essentialism is generic language, such as “girls play with dolls” or “a girl plays with dolls”, which describe a category as a whole. There is not a single syntactic marker that reliably indicates that a statement is generic – rather, whether a statement is interpreted generically involves a combination of morphosyntactic cues, background knowledge about the predicate and the category, and the physical and social context (Cimpian & Markman, 2008; Cimpian, Meltzer, & Markman, 2011). Generics are potentially powerful, because our experience is filled with individual things, and generic language is a simple way of communicating a generalization about a whole category. A generic statement may cue essentialism by suggesting that the property is a causally relevant, non-accidental, and conceptually central property for that category (Cimpian & Markman 2009/2011). Indeed, generics have been shown to elicit essentialism in children in both biological and social domains (Gelman et al., 2010; Rhodes, Leslie, Tworek, 2012). Four-year-olds who heard generic language about a novel social category (e.g. “Zarpies sleep in trees” or “A Zarpie sleeps in trees”) showed elevated essentialist beliefs about that social category, compared to a group who heard a control statement. Namely, the children who heard generic language were more likely to explain the presence of properties in category members in terms of intrinsic causes, considered category properties to be innate, and extended properties of a single category member to other category members. In addition, generics are an ecologically valid cue for essentialism: generics are prominent in parents’ speech to children, which causes children to hear a lot of generic language in their environment (). Parents who were induced to hold essentialist beliefs about a novel social category (e.g. told that Zarpies are a distinct kind of people, biologically and culturally different from others) produced more generics when reading a book about that category to their children. These studies suggest that generics are a cue for essentializing the category mentioned.

*Formal explanations as a cue for essentalism*

Recently, formal explanations have been suggested as another linguistic cue for essentialism. Formal explanations explain why an individual has a certain property by reference to the category it belongs to, such as “Suzy plays with dolls because Suzy is a girl” (Prasada & Dillingham 2006/2009). Formal explanations explain that an individual has some property because it is a member of a particular category. Formal explanations directly link an individual and a property to an entire category, such that they could support learning about an entire category from a single individual. As such, formal explanations are potentially powerful cues to category structure. At the same time, formal explanations present a puzzle: formal explanations are potentially powerful cues to category structure, yet they are almost tautological sounding. If we ask someone, “Why does Fido bark?” and they answer “because Fido is a dog”, the fact that Fido is a dog is generally not informative. We are generally aware of the individual’s category membership. Yet we do seem to be gleaning *something* from hearing the explanation, and such explanations sound natural and meaningful to us.

Formal explanations have been a relatively new frontier of research, and what we know about them is largely limited to metalinguistic judgments from adults. Prasada & Dillingham 2006 introduced formal explanations to the literature in the context of deriving various kinds of category properties (eg merely statistical, conceptually central, etc). Native English speakers rate an explanation like “That dog is four-legged because it is a dog” as a more felicitous than “That barn is red because it is a barn”, despite the fact that having four legs and being red are statistically common properties of dogs and barns, respectively (Prasada & Dillingham, 2006, Exp 2a). Prasada & Dillingham have argued that the former explanation is rated as more felicitous than the latter because having four legs is central to what it means to be a dog, while being red is not central to what it means to be a barn. Prasada & Dillingham propose that a formal explanation is felicitous if and only if the property mentioned is conceptually central or an aspect of what it means to be a category member.

Building on the pioneering work of Prasada & Dillingham, recent proposals have gone further and suggested that formal explanations not only pick out properties that are an aspect of what it means to be a category member, but pick out properties that are *inherent* to what it means to be a category member. In particular, formal explanations have been suggested to be placeholders for inherent explanations, such that there is something inherent about being a category member that gives rise to category members possessing that property, e.g. that there is something inherent about being a girl that causes girls to play with dolls (Gelman, Cimpian, Roberts, 2018). As a result, hearing a formal explanation may be a cue to essentialize the category mentioned in the explanation.

This suggestion that formal explanations can be a cue for essentialism has been supported by a recent study examining whether formal explanations increase essentialism in children. Muradoglu et al, 2019 found that formal explanations do indeed boost essentialism for known animal categories, and perhaps for gender as well. Muradoglu et al, 2019 involves two studies, one about animal categories, and the other about gender categories. Five- and six-year-olds either heard a formal explanation (e.g. “Margaret likes to eat gooseberries because she is a girl.”) or a control statement (e.g. “Margaret is a girl. She likes to eat gooseberries.”) for why an animal (Exp 1) or a girl (Exp 2) possessed a novel behavioral property or a novel behavioral preference. After each of 4 trials, participants were measured on 5 canonically-used measures of essentialism: 2 stability over time items, 2 innateness measures, and 1 inductive potential measure. The items were the following, administered in this order with the 2 innateness measures in random order. Participants responses on all 5 items across 4 trials were averaged into an essentialism score ranging from 0 to 1 for each participant.

|  |  |
| --- | --- |
| Stability past | Now, here’s a question for you: Margaret is 8 years old.  Do you think she liked to eat gooseberries when she was 4 years old? (Y/N) |
| Stability future | Do you think she will always like to eat gooseberries? (Y/N) |
| Innateness stop | Would Margaret stop liking to eat gooseberries if her family tried to stop her from eating gooseberries? (Y/N) |
| Innateness switch | Would Margaret like to eat gooseberries if she grew up in a family where no one liked to eat gooseberries? (Y/N) |
| Inductive potential | Do you think “just this girl,” “a few girls,” or “a whole lot of girls” like to eat gooseberries? (3-point) |

*Figure 1.* The 5 canonical essentialism measures used in Muradoglu et al, 2019.

In the first study about animal categories, there was the hypothesized main effect of explanation, such that 5 and 6-year-olds who heard the formal explanation had higher measures of essentialism than those who heard the control statement (Figure 2). In the second study about gender, there was no main effect of explanation, contrary to predictions (Figure 2). There was an unanticipated interaction between explanation and age, such that the 6-year-olds who heard the formal explanation had higher essentialism measures than the 6-year-olds who heard the control statement, and the 5-year-olds had the same measure of essentialism for both explanations. Muradoglu et al conclude that formal explanations may be a cue for essentialism at age 5 for animal categories and starting at age 6 for gender.

**A close up of a map

Description automatically generatedFigure 2.** The results from Muradoglu et al, 2019 for left, animal categories, and right, gender.

Previous work on formal explanations have focused on linking formal explanations to an essentialist construal of a category, perhaps because formal explanations suggest that there is something inherent about being a category member that gives rise to the property (Gelman, Cimpian, Roberts, 2018).

*Two interpretations of formal explanations*

Contrary to previous work, I will argue here that formal explanations actually present an ambiguity, such that they can be subject to an essentialist interpretation or a social-structural interpretation. In particular, when appropriate information about the social-structural context is available, we may produce and interpret formal explanations in social-structural terms.

Intuitively, we can consider formal explanations that sound natural and felicitious, yet do not seem to yield an essentialist interpretation. Consider the following formal explanations:

Spot can’t enter the cafeteria because Spot is a dog.

I walk in groups late at night because I’m a woman.

Rosa Parks couldn’t sit in the front of the bus because she was black.

These formal explanations that sound natural and felicitous, yet we do not infer that there is something intrinsic about being a dog that prevents Spot from physically entering a cafeteria, or something intrinsic about being a woman that causes the speaker to walk in groups late at night, or something intrinsic about being black that prevents Rosa Parks from sitting at the front of the bus. Rather, these examples of formal explanations point to *structural factors* at play. Structural factors are stable extrinsic factors that act on a category by virtue of where the category is situated within a larger structure (Haslanger, 2016). For an example, dogs are situated within a human society that would prefer that nonhuman animals stay separate from human eating spaces, women are situated within a social context of gender-based violence, and African-Americans like Rosa Parks were situated within a historical context of Jim Crow laws. These formal explanations highlight stable social forces that act on a category, due to the position that that category occupies in a larger social structure, rather than something inherent or essential about the category itself. To return to Prasada & Dillingham’s proposal for what makes for a felicitous formal explanation, these formal explanations may sound felicitous to us because the social experience of being a member of a category can be part of what it means to be a category member. For an example, the social experience of being a woman can be part of what it means to be a woman. More broadly, the social context that is confounded with category membership can become integrated into what it means to be a category member. From these explanations, it is plausible that formal explanations could be interpreted as an expression of structural factors acting on a category, rather than something inherent or essential about a category. In other words, formal explanations could have a viable social-structural, rather than essentialist, interpretation.

Can the presence of structural factors in the context block an essentialist interpretation of a formal explanation, and promote a social-structural interpretation instead? A study that could address this question would manipulate the presence of structural factors in the situation and examine whether that modulates the effect of formal explanations on essentialism. In other words, the study would be a 2x2 study with two independent variables: the presence or absence of structural factors, and a formal explanation or a control statement, and one dependent variable: a measure of essentialism, such as mutability. If structural factors block the essentialist effects of a formal explanation, the essentialism-boosting effects of hearing a formal explanation should be observed in the non-structural condition, but that boost should be decreased, eliminated, or even reversed in the structural condition.

A study that comes close to but does not directly address the target question is Vasilyeva, Gopnik, Lombrozo, 2018. Vasilyeva et al, 2018 found that 3 to 6-year-old children and adults can reason about structural factors, and that they judge a formal explanation to be equally apt in a situation with structural factors versus in a neutral situation without structural factors where an essentialist interpretation is more likely. The paradigm they used involved a fictional school with gender-segregated classrooms, where students decide which of two novel games – Yellow-Ball or Green-Ball – to play at recess by tossing a pebble into one of two buckets in their classroom. In the structural condition, the bucket sizes are skewed between the boys’ classroom and the girls’ classroom, e.g. the bucket for Yellow-Ball is very big in the girls’ classroom, and the bucket for Green-Ball is very big in the boys’ classroom. This manipulation is meant to suggest that different genders are subject to differential social-structural factors. In the nonstructural condition, the bucket sizes are equal in each classroom, suggesting that different genders are subject to the same social-structural factors. Participants in both conditions then saw that most girls play Yellow-Ball, and most boys play Green-Ball. For participants in the structural condition, the structural context could explain why different genders play different games, eliminating the need for an essentialist construal. For participants in the nonstructural condition, equivalent bucket sizes should suggest that different genders are equally likely to play either game, so different genders playing different games might be surprising. Previous work suggests that given an unrepresentative skewed sample, like the game-playing behavior observed here, children will attribute a preference, so children in nonstructural context may think that different genders prefer different games (Kushnir, Xu, Wellman, 2010). Consequently, the structural context might decrease essentialism, compared to a nonstructural context.

Vasilyeva et al, 2018 administered a number of dependent measures to participants, including a mutability task, which is a measure of essentialism (e.g. Suzy’s parents move her to the boys’ classroom – which game do you think Suzy will play?), and a formal explanation rating task, in which children and adults rated how good of an explanation a formal explanation was (e.g. “Suzy plays Yellow-Ball because Suzy is a girl” Is this a good or a bad explanation?). For the mutability task, children and adults demonstrated higher mutability, that is, lower essentialism, in the structural condition than in the nonstructural condition, which suggests that the presence of structural factors in the situation blocked them from otherwise forming essentialist beliefs about gender. For the formal explanation rating task, children and adults rated the formal explanation as equally apt regardless of which context they had seen, despite the fact that participants are adjusting their essentialism relative to the context. This study thus hints at the possibility that formal explanations may be amenable to a structural rather than essentialist explanation.

Vasilyeva et al, 2018 provides an excellent structural versus nonstructural paradigm for addressing the target question, but does not directly address the issue of whether formal explanations have two available interpretations for two reasons. Firstly, Vasilyeva et al, 2018 used a formal explanation *rating* task, where children and adults were asked to judge a puppet’s formal explanation as a good explanation or not. In addition to the uncertainty about how much to trust a 3-year-old’s metalinguistic judgments, the rating task only provides a measure of how apt the formal explanation is, and does not probe how the formal explanation is being interpreted. Secondly, Vasilyeva et al 2018 was designed to separately address the questions of how structural situations impact the aptness of a formal explanation, and how structural situations impact essentialism, but does not provide evidence as to how structural situations interact with a formal explanation as to impact essentialism. This study takes Vasilyeva et al, 2018 as a starting point, but aims to directly address whether structural situations can block the essentialist effects of a formal explanation.

This study adapts the structural versus nonstructural paradigm from Vasilyeva et al, 2018 and the formal explanation versus control statement paradigm from Muradoglu et al, 2019 in order to address whether structural situations can block the essentialism that would typically result from hearing a formal explanation. In this study, 5 and 6-year-olds were introduced to the same gender-segregated classroom paradigm as from Vasilyeva et al 2018, hear a formal explanation (e.g. “She plays Yellow-Ball *because* she is a girl”) or control statement (e.g. “She is a girl. She plays Yellow-Ball.”) adapted from Muradoglu et al to the gender-segregated paradigm, and then will complete a few measures of essentialism. If structural contexts do block the essentialist effects of a formal explanation, there should be an interaction between situation and explanation, such that formal explanations relative to the control statement should boost essentialism in the nonstructural condition, but that boost should be reduced, eliminated, or reversed in the structural condition. If structural contexts fail to block the essentialist effects of a formal explanation, there should only be a main effect of explanation, such that formal explanations relative to the control statement boost essentialism in both structural and nonstructural conditions.

This project's [repository](https://github.com/mariannazhang/structuralblocking) and [preregistration](https://osf.io/7ja86/) can be found online.

**Methods**

*Participants.* 5 and 6-year-old children (n=38, study in progress) were recruited from Stanford’s Bing Nursery School, and the Tech Museum in San Jose. Adults (n=183, additional n=17 excluded for failing a comprehension check) were recruited from Amazon Mechanical Turk. Adult participants were required to be located within the United States, and to have a HIT acceptance rate of 80% or above. MTurk participants were paid 73 cents each.

We chose a target sample size of n=192 5-6yo and n=184 adults, which is n=48 and n=46 respectively in each cell of this 2x2 experimental design. We chose these sample sizes because n=46 per cell was the maximum sample size used in prior studies, n=48 was within our resource constraints, and n=48 suited the counterbalancing scheme we used with children. By running our target analysis on simulated data based on results from prior studies and our own predictions, we found that n=184 should yield 68% power.

*Procedure.* Participants were shown a storybook adapted from Vasilyeva, Lombrozo, & Gopnik, 2018, presented in PowerPoint slideshow format on a laptop display. Two between-subjects variables were be manipulated, resulting in a 2x2 design: context (structural vs non-structural) and explanation (formal explanation vs control statement).

*Context manipulation.* The storybook describes a fictional school with gender-segregated classrooms, where children decide what game to play at recess by tossing a pebble into one of two buckets in their classroom: a yellow bucket and a green bucket. If a child tosses a pebble into the yellow bucket, they will play Yellow-Ball. If a child tosses a pebble into the green bucket, they will play Green-Ball. In the *structural condition,* the buckets are of unequal sizes in each classroom. In the girls’ classroom, the yellow bucket is much larger than the green bucket, and in the boys’ classroom, the green bucket is much larger than the yellow bucket. In the *non-structural condition,* the buckets are of equal sizes in each classroom.

*Statistical evidence.* All participants then see statistical evidence that in the girls’ classroom, the majority of girls throw their pebble into yellow bucket, and thus play Yellow-Ball, and that in the boys’ classroom, the majority of boys throw their pebble into the green bucket, and thus play Green-Ball. A series of memory checks confirm that the participants remember the gender-segregation of the classrooms, the size of the buckets in each classroom, and the statistical evidence for what games girls and boys play.

*Explanation manipulation.* The explanation manipulation was adapted from Muradoglu et al, submitted. All participants are introduced to a girl called Suzy. In the formal explanation condition, participants are given a formal explanation for why Suzy plays Yellow-Ball (“(Shows picture of Suzy) Look here. This is Suzy. (shows picture of a Yellow-Ball) Suzy plays Yellow-Ball. And do you know why Suzy plays Yellow-Ball? *Because* she is a girl. So, just to remind you, Suzy plays Yellow-Ball because she’s a girl.”). In the control statement condition, participants are given a control statement that does not constitute a formal explanation (“(Shows picture of Suzy) Look here. This is Suzy. She’s a girl, and do you know what else? (shows picture of a Yellow-Ball) Suzy plays Yellow-Ball. Suzy plays Yellow-Ball. So, just to remind you, Suzy is a girl and she plays Yellow-Ball”). Note that this manipulation differs from the formal explanation evaluation task in Vasilyeva et al, 2018, in that this manipulation directly provides an explanation to the participant, whereas the evaluation task presents a possible explanation to be evaluated as a good explanation or not by the participant.

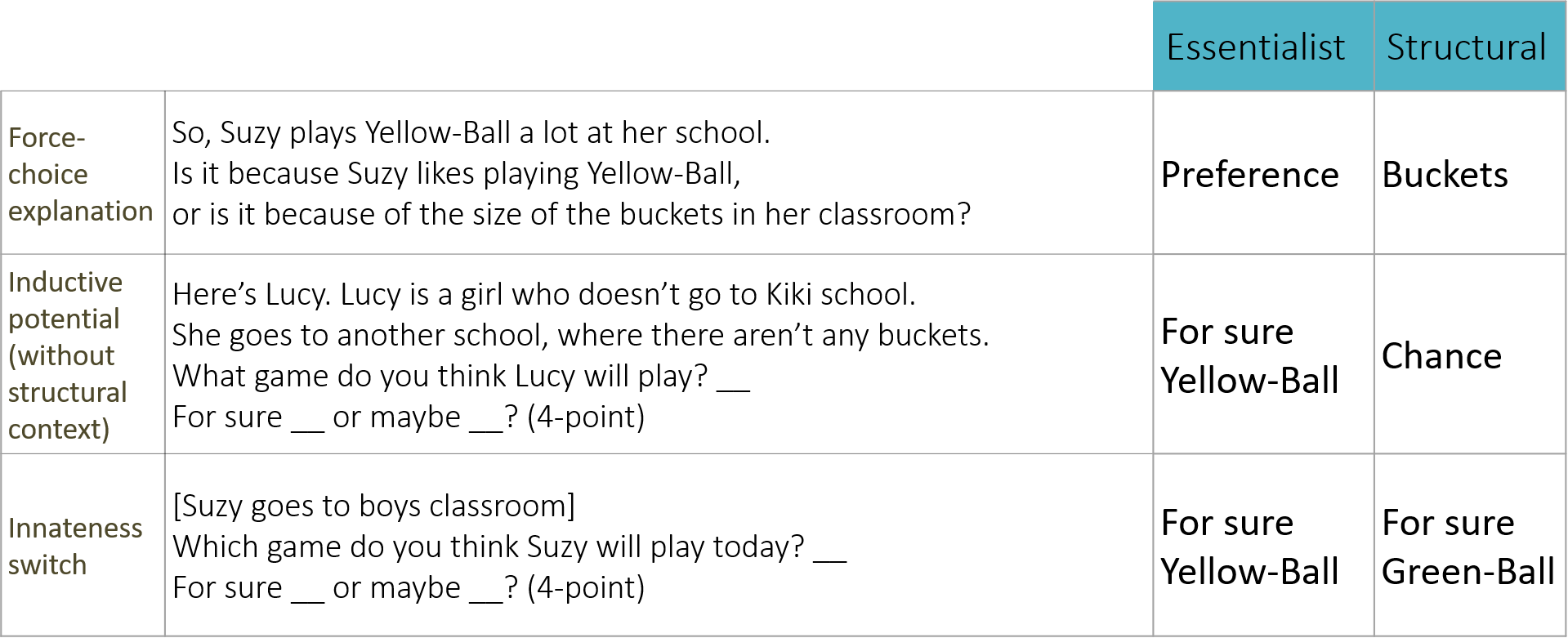
*Essentialism measures*. All participants then will complete 3 measures of essentialism (Figure \_).

Figure. The 3 measures of essentialism used in this study.

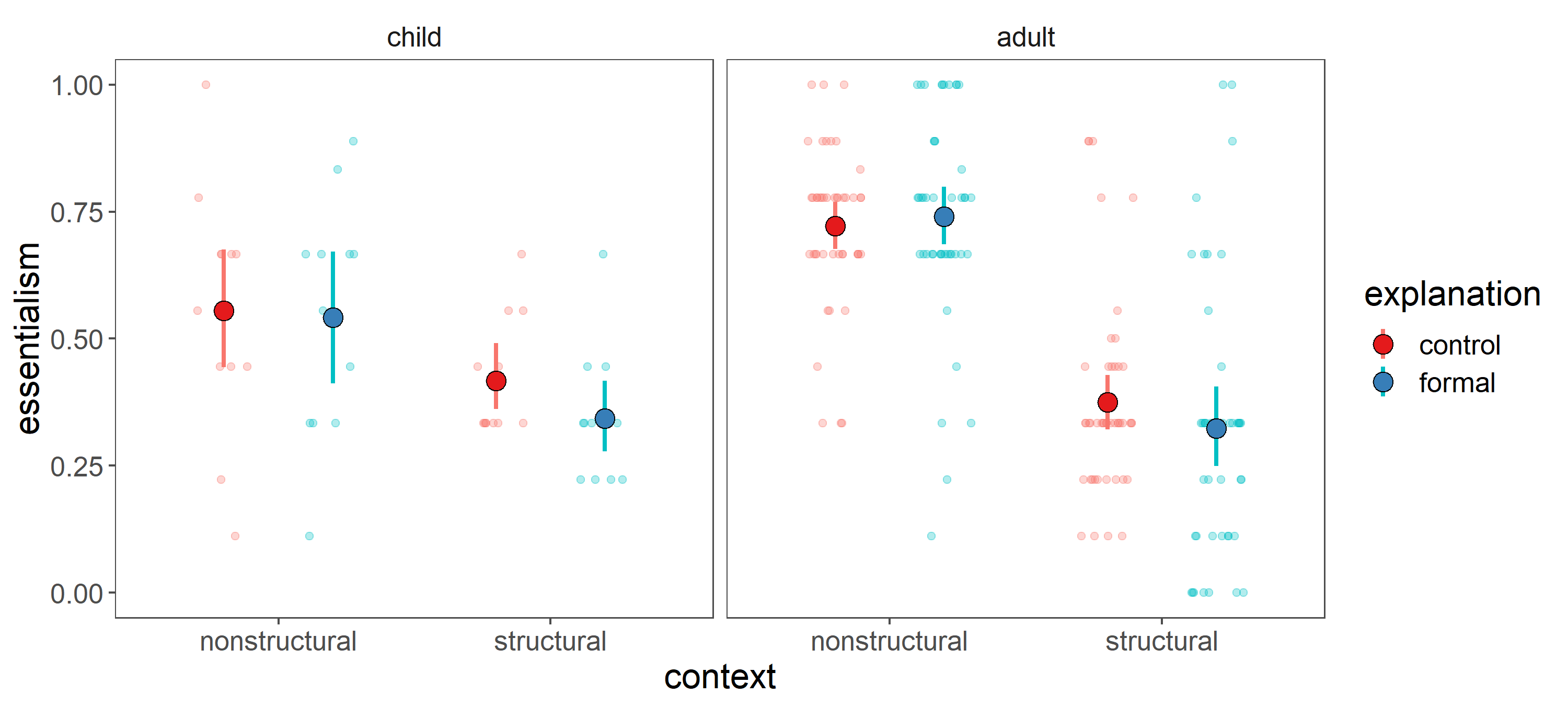
The first measure was a force-choice explanation measure. Participants were told: “Suzy plays Yellow-Ball a lot at her school. Is it because Suzy likes playing Yellow-Ball, or is it because of the size of the buckets in her classroom?”. Participants could then cite the preference explanation or the context explanation. By pitting preference against context, this item teases apart an essentialist construal, which may be more likely to cite an internal feature, from a structural construal, which may be more likely to cite the context.

The second measure was an inductive potential measure without the structural context. Participants were introduced to a new category member who is not situated within the previous context: “Here’s Lucy. Lucy is a girl who doesn’t go to Kiki school. She goes to another school, where there aren’t any buckets. What game do you think Lucy will play?”. Participants were asked whether they think Lucy will play Yellow-Ball or Green-Ball, and whether they think that Lucy will maybe play that game or for sure play that game. The two-step prompt produces 4 possible responses: 1=for sure Yellow-Ball, 2=maybe Yellow-Ball, 3=maybe Green-Ball, and 4=definitely Green-Ball. This measure provides purchase on the distinction between an essentialist construal and a structural construal because an essentialist construal would predict that a new category member should also exhibit the property regardless of context, since category membership is what matters, but a structural construal would predict that whether a new category member exhibits the property depends on whether they too are subject to the same structural context, since the context is what matters.

The last measure is the mutability measure – here called innateness switch – drawn from Vasilyeva, Lombrozo, & Gopnik, 2018. The innateness switch measure is a canonical measure of essentialism (Gelman & Wellman 1991). Participants were told that a new policy has been instituted at the school, such that any child of any gender can now attend any classroom, and that Suzy’s parents decided to transfer Suzy to the boys’ classroom, since they know the teacher there. The motivation for the transfer was specified to be Suzy’s parents knowing the teacher in the boys’ classroom so that participants are not given reason to believe that Suzy herself wanted to transfer to the boys’ classroom. The change of classroom constitutes a change of environment analogous to a switched-at-birth task. Participants were asked whether they think Suzy will play Yellow-Ball or Green-Ball, using the same two-step procedure detailed previously. A response of For sure Yellow-Ball suggests that playing Yellow-Ball is an immutable innate property that cannot be changed by a change in context, while a response of For sure Green-Ball suggests that playing Yellow-Ball is a mutable property that can shift with a change in context.

*Exploratory measure.* A measure of group normativity was also collected as an exploratory measure, since essentialism has been linked to normativity, and formal explanations have induced an expectation that category members should possess the property mentioned, and a sense of normative violation when category members do not.

*Analysis plan*. Child participants will be excluded for experimenter error, insufficient English language skills, parental interference, or failing to complete the task. Adult participants were excluded for failing the game comprehension check (n=17), or the attention check (n=0). The analysis of interest will be an ANOVA on subjects' essentialism measures with an interaction term for context and explanation.

**Results**

*Figure* \_\_\_. Essentialism per context and explanation condition, with child participants on the left facet and adult participants on the right facet.

*Children.* Data collection is still in progress for the children, with about a fifth of the target sample size collected, so it is difficult to interpret the incomplete data so far. The current data trends suggest that there may not be an essentialism boost from formal explanations in the nonstructural condition, and that there might be an essentialism reversal from formal explanations in the structural condition.

*Adults.* Rather than predicted context and explanation interaction, there was a strong main effect of context (F(1, 179) = 152.251, p < 0.001), with no main effect of explanation (F(1, 179) = 0.283, p = 0.60) and no interaction between context and explanation (F(1, 179) = .234, p = 0.27).

The adult data is in line with the main effect of context that was found among adults in Vasilyeva et al, 2018. The strong main effect of context is remarkable, given that adults might be thought to already have a significant amount of contextual knowledge about the structural context of gender, a familiar category, even if the classroom paradigm and the game-playing property are novel.

It is interesting that the adults aren’t affected at all by hearing the formal explanation. We did not have a strong a priori prediction about the adults, since the link between formal explanations and essentialism not been studied in adults before. A possible interpretation of the lack of a main effect for explanation is the following. Participants hear about the context first, and then receive an explanation later. Given just the contextual information, adults may establish a construal that then guides their interpretation of the formal explanation, such that the formal explanation falls in line with whatever construal was established by the context.

**Discussion**

Language is a powerful source of information as children are constructing category representations. Whether we conceive of gender and other social categories in essentialist terms or social-structural terms is socially consequential, e.g .form the basis for stereotyping and prejudice, form the basis for differential policies, and justification for inequities in social structure. Previous work has suggested that formal explanations, like generics, serve as a linguistic cue for what categories to essentialize. Here, I’ve argued that formal explanations may be more ambiguous than previously thought. Formal explanations may not necessarily always yield an essentialist interpretation. In cases where there is a structural context present, formal explanations might yield a social-structural interpretation rather than an essentialist one. Another way of wording this is that the structural context “blocks” the essentialist reading of the formal explanation, and allows one to take a social-structural construal. In this report, I have presented a theoretical argument and some preliminary data for the plausibility of formal explanations having this alternative social-structural interpretation, but this is an ongoing project. This project serves as an initial foray into the question of how children integrate linguistic and contextual information when developing category representations.

*Structural blocking of essentialism*

If the presence of structural factors in the context blocks the essentialist effects of formal explanations, the results would demonstrate the power of contextual information to overcome linguistic cues, and suggest a method for blocking the transmission of potentially pernicious attitudes about social groups. Even though children and adults rate a formal explanation as apt in a variety of situations (Vasilyeva, Lombrozo, Gopnik, 2018), this set of results would suggest that children do not necessarily infer an essentialist interpretation from a well-formed formal explanation. This set of results would demonstrate that awareness of structural factors could block the formation of potentially pernicious essentialist attitudes about gender. While much of the literature has focused on how essentialism is induced and transmitted, this set of results would demonstrate a novel mechanism to block essentialism. In particular, the structural manipulation in this study parallels structural inequalities in present society. In this study, the bucket sizes in their classroom biases girls to engage in certain types of activities over others. Similarly, in present society, resource inequalities, differential standards, and implicit bias, among other structural factors, cause women to engage in particular activities (e.g. engage in certain types of play, dress in certain ways), and reach divergent educational and social outcomes compared to men (e.g. academic achievement, underrepresentation in various fields, management, and leadership). As a result, this set of results could have practical implications for how to combat the early origins of stereotyping and prejudice by raising awareness of structural factors.

It is an open question whether the presence of structural factors blocks the effects of other linguistic cues as well, such as generics (Rhodes, Leslie, & Tworek, 2012) and directional comparisons (Chestnut & Markman, 2016). If awareness of structural factors blocks formal explanations from cueing essentialism, structural factors could be a powerful cue-general method of blocking essentialism and other socially pernicious beliefs about social categories.

*The nature of formal explanations*

A few open questions suggested by this project concern the nature of formal explanations. As mentioned earlier, they present an interesting puzzle: they sound tautological and uninformative, yet we find them natural and meaningful. In this report, I’ve suggested that their interpretation might shift in really contextually dependent and interesting ways. There are a few open questions raised by formal explanations. First, given this slipperiness about how to interpret a formal explanation, why do we produce formal explanations at all? Why don’t we produce alternative explanations that are clear and unambiguously essentialist or social-structural? Instead of “I walk in groups late at night because I’m a woman”, I could say “I walk in groups late at night because the likelihood of women being targeted for gender-based harassment and violence increases during the night hours”. Grice’s principle of brevity may be at play here, since the unambiguous explanation is much longer than the formal explanation (Grice, 1975). In addition, the speaker may assume that the relevant structural factors are in the common ground, such that the speaker may implicitly gesture towards the relevant structural factors using a formal explanation, and trust that the listener can recover that reference, rather than explicitly stating them (Clark, 1996). The relevant structural factors may also be taboo to mention or state, which may cause the speaker to prefer the implicit formal explanation rather than the explicit unambiguous explanation. Further work could investigate the frequency and conditions under which formal explanations are naturally produced by speakers. The literature has largely focused on aptness ratings for artificially constructed formal explanations, or the spontaneous production of formal explanations in an extremely limited setting. A corpus study could shed light on the reasons for which we produce formal explanations, and what meaning they provide to the listener.

Second, how variations on formal explanations shift their interpretation? For an example, “because I’m a woman” seems to suggest something different from “simply because I’m a woman” or “just because I’m a woman”. The subtle addition of “simply” or “just” seems to change the interpretation of the formal explanation, although anecdotally, people seem to have different intuitions about whether the addition makes the explanation more essentialist or more social-structural.

Despite their interpretative slipperiness, formal explanations can be powerful cues to category structure. Formal explanations directly link an individual and a property to an entire category, such that they could support learning about an entire category from an individual. As such, untangling the puzzle of formal explanations could help us understand how language shapes the development of category representations.

References

Chestnut, E. K., & Markman, E. M. (2016). Are Horses Like Zebras, or Vice Versa? Children’s Sensitivity to the Asymmetries of Directional Comparisons. *Child Development*, *87*(2), 568–582. https://doi.org/10.1111/cdev.12476

Clark, 1996

Cimpian, A., & Markman, E. M. (2008). Preschool children’s use of cues to generic meaning. *Cognition*, 107, 19–53.

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,* 14–25. <http://dx.doi.org/10.1016/j.cognition.2009.07.004>

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development, 82,* 471–492. http://dx.doi.org/10.1111/j.1467-8624.2010.01525.x

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195154061.001.0001>

Gelman, S. A. (2009). Learning from others: Children’s construction of concepts. *Annual Review of Psychology*, 60, 115–140.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, 38, 213–244.

Grice, 1975

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology*, 41:87–100.

Haslanger, S. (2016). What is a (social) structural explanation? *Philosophical Studies, 173,* 113-130.

Kushnir, Xu, Wellman, 2010

(Mahalingam, 2003a)

Muradoglu, M., Marchak, K., Cimpian, A., & Gelman, S. (poster submitted to SRCD). Do formal explanations promote essentialist beliefs in children?

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99,* 73–112. <http://dx.doi.org/10.1016/j.cognition.2005.01.003>

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

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology, 59,* 244–274. <http://dx.doi.org/10.1016/j.cogpsych.2009.05.001>

Rhodes, M., Leslie, S. J., Tworek, C. (2012). Cultural transmission of social essentialism. *PNAS,* 109(34), 13526-13531. <https://doi.org/10.1073/pnas.1208951109>

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The Development of Structural Thinking About Social Categories. *Developmental Psychology,* 54(9), 1735-1744.

1. Here I use the gender binary for ease of illustration, but the social-structural view is not committed to the gender binary, except insofar as it is a product of social context. Note that the essentialist view is committed to such a binary, since categories are thought to carve out distinctly different natural kinds. [↑](#footnote-ref-1)
2. Here, I will not directly address whether input overrides a default essentialist construal, or whether input is required to determine which construal to use in representing a category. The literature has been mixed on these two possibilities. In this report, I will run with the latter possibility, that input shapes which construal to use, although certain types of language might have default modes of interpretation. [↑](#footnote-ref-2)