Running head: CATEGORY STRUCTURE AND INDUCTION

Category structure affects the developmental trajectory of children's inductive inferences for both natural kinds and artifacts

Julia R. Badger*

Department of Experimental Psychology, University of Oxford, Oxford, OX1 3UD

*Corresponding Author: julia.badger@psy.ox.ac.uk; +44 (0) 1865 271417

Laura R. Shapiro

School of Life and Health Sciences, Aston University, Birmingham, B4 7ET, UK

l.r.shapiro@aston.ac.uk; +44 (0) 121 204 4052

Abstract

Inductive reasoning is fundamental to human cognition, yet it remains unclear how we develop this ability and what might influence our inductive choices. We created novel categories in which crucial factors such as domain and category structure were manipulated orthogonally. We trained 403 4- to 9-year-old children to categorize well matched natural kind and artifact stimuli with either featural or relational category structure, followed by induction tasks. This wide age range allowed for the first full exploration of the developmental trajectory of inductive reasoning in both domains. We found a gradual transition from perceptual to categorical induction with age. This pattern was stable across domains, but interestingly, children showed a category bias one year later for relational categories. We hypothesize that the ability to use category information in inductive reasoning develops gradually, but is delayed when children need to process and apply more complex category structures.

Adults can use both perceptual cues and category knowledge when making inductive inferences (Bulloch & Opfer, 2009; Gelman & Markman, 1986). For example, if we were to install a nest box intended for small birds, we could use perceptual cues to judge that a larger bird would not inhabit it. In contrast, our category knowledge would allow us to judge that both would visit a bird feeder. The extent to which young children are able to use both perceptual and categorical cues when making inductive inferences remains under debate (e.g., Badger & Shapiro, 2012; Gelman, 2003; Gelman & Markman, 1986; Gentner, 1988; Medin & Ortony, 1989; Opfer & Bulloch, 2007; Sloutksy & Fisher, 2004; Sloutksy, Kloos & Fisher, 2007). A range of methodologies and stimuli have been used to investigate children's inductive strategies and these have produced contradictory findings. In some studies, young children show an early bias for category induction (e.g., Bulloch & Opfer, 2009; Opfer & Bulloch, 2007; see Gelman, 2003, for an overview) or analogical reasoning (e.g., Goswami, 2001). It has also been shown that young children see category membership as a good basis for generalization (e.g., Hayes, McKinnon & Sweller, 2008). In contrast, in other studies, young children appear biased towards perceptual induction (e.g., Sloutsky et al., 2007) or surface characteristics (e.g., Gentner, 1988), and gradually transition towards category-based reasoning at a later developmental stage (e.g., Badger & Shapiro, 2012; Fisher & Sloutsky, 2005; Gentner, 1988; Gentner & Rattermann, 1991; Rattermann & Gentner, 1998). These contradictory findings have led to a strong debate as to the validity of the methodologies and stimuli used.

The triad paradigm is commonly used to investigate children's induction choices because it directly sets perceptual against categorical preferences (e.g., Badger & Shapiro, 2012; Gelman & Markman, 1986; Sloutsky et al., 2007). A target item is presented alongside two test items, one of which matches the target in category but not overall appearance (the categorical choice); the other matches the target in overall appearance but not category (the perceptual choice). The child is asked to generalize a hidden property of the target onto one of the two test items, or to generalize one of the test items' hidden properties to the target. A target-to-categorical choice generalization is

classified as category induction. A target-to-perceptual choice generalization is classified as perceptual induction. Gelman and Markman (1986) used real kind stimuli in their triads: for example, a blackbird (target), a flamingo (category), and a bat (perceptual). They found that children aged 4-5 years generalized the hidden property between the target and category choice significantly more often. They concluded that as long as the children knew the category membership of the items, they would successfully apply category induction. In contrast, Sloutsky et al. (2007) used artificial biological kinds (Ziblets and Flurps) to gain full control of the information children are using to make inductive inferences. They used a relational category structure, in which category membership was defined by the relationship between two features (relative number of "fingers" to "buttons": an item was a Ziblet if it had more fingers than buttons). Overall appearance was non-predictive of category membership. Children aged 4 to 5 were trained to categorize these stimuli accurately, but made significantly more perceptual induction inferences. This was interpreted as an early perceptual bias, which would develop into a category bias with age. There is an ongoing argument as to whether real or novel stimuli provide a stronger test of category induction. Gelman and Waxman (2007) claim that Sloutsky's novel stimuli were not biologically plausible and the category rules were arbitrary. Therefore, the children would not have responded to the stimuli in the same way as real biological kinds as they would not have believed them to be real. Using novel stimuli with a complex (and arbitrary) category structure could underestimate children's category induction capabilities. Whereas using real kind stimuli could cause us to falsely interpret appearance or familiarity-based choices as category induction (children may choose the more perceptually similar or more familiar item). Unfortunately, the age range in the Sloutsky et al. study is too narrow to know whether the children would have used category induction as they aged, or whether both younger and older children would have used perceptual induction as they would not have accepted the stimuli as natural kinds.

A more rigorous test of children's category induction capabilities is to examine how these preferences change as children develop (as in Badger & Shapiro, 2012). If Sloutsky's perceptual bias was caused by the use of unrealistic stimuli with complex category structure, then a "perceptual bias" is unlikely to change with development (older children would continue to accept the perceptual choice). In contrast, if this were a genuine perceptual bias that transitions into a category bias with experience, then children's preferences should change with age. The current study examines whether children's category induction choices change as a function of children's age (4-9 years old), and the nature of the categories with which children are confronted. Specifically, the factors governing category membership (hereafter category structure; e.g., Kittur, Hummel & Holyoak, 2004; Saiki & Hummell, 1998; see also Barsalou, 1993), and the ontology of the category (hereafter domain, e.g., natural kinds vs. artifacts; Keil, 1989).

Category Structure

The structure of a category is critical when making generalizations about an unknown item. For example, since knives have a common feature (a blade), the presence of this feature on an unknown item suggests that the item is likely to share other properties common to the knife category (e.g., has the property of "cutting"). Categories can be deterministic – whereby all items of a category share at least one parameter that is exclusive to that category – or probabilistic – whereby a percentage of items of a category share at least one parameter but no feature is shared by all (e.g., see Kittur et al., 2004). We will focus on two key deterministic structures: featural and relational. As discussed by Kittur et al. most category learning experiments use exemplars which can be distinguished on the basis of particular features. Although these are usually artificial categories, a real world example would be the category of knives, which share a unique feature (featural structure; e.g., all knives have a blade). In fact, in the real world, categories are often better described by the relationship between features (relational structure; e.g., the relationship between legs and seat defines a chair). Within the adult category learning literature, the effects of featural

versus relational structures have been investigated systematically (e.g., Tomlinson & Love, 2010). However, until now no study has investigated the influence of these types of structure (featural vs. relational) on the development of induction. This is surprising considering that the natural default for young children's induction may be partly explained by the structure of the stimuli seen. This could explain the conflicting findings described above; for example, Gelman and Markman's (1986) triads could generally be distinguished by focusing on featural information: individual, albeit non-obvious, features. In contrast, the novel stimuli used in Sloutksy et al. (2007) were relational and explicitly designed so that they could only be distinguished by *comparing* two features.

Reasoning about relational category structures is likely to place greater demands on a child's cognitive processing. It is well known that the number of relations that can be processed in parallel increases with age (e.g., Andrews & Halford, 2002; Halford, 1993), as does the ability to apply complex rules to a reasoning task (e.g., Zelazo, Frye & Rapus, 1996; Zelazo & Müller, 2002). Despite this knowledge, there is no study that has systematically and directly compared the influence of the category structure of items on inductive preference. Therefore, it is not yet valid to directly compare results from studies with different category structures.

Children's understanding of natural kind and artifact domains

There is evidence that children as young as 3 years old are domain sensitive when making assumptions about the properties of an object (e.g., Backscheider, Shatz & Gelman, 1993; Carey, 1985; Inagaki & Hatano, 1996; Jipson & Gelman, 2007; Mandler & McDonough, 1996). For example, Jipson and Gelman (2007) found that 3-5 year old children clearly distinguished between artifact and natural kinds when shown video clips of unfamiliar items and asked to generalize biological properties such as 'does this one eat?'. In addition, children have been shown to successfully associate living traits such as breathing, survival and growth to natural kinds, and nonliving functional traits to artifacts (e.g., Gelman, 1988; Greif, Kemler Nelson, Keil & Gutierrez, 2006; Inagaki

& Hatano, 1996; Keil, 1992; Kemler Nelson, Egan & Holt, 2004). Thus children are clearly able to make category distinctions from a general understanding of the conceptual difference between artifacts and natural kinds (Booth & Waxman, 2002; Booth, Waxman & Huang, 2005). Children also make more subtle distinctions between natural kinds versus artifacts when generalizing properties or behavior. Specifically, when confronted with natural kinds, they focus on the item's 'insides' (e.g., Gelman, 2003; Gottfried & Gelman, 2005); will be more likely to assign category membership based on internal properties (e.g., Gelman & Wellman, 1991), and appear to be sensitive to underlying causal mechanisms explaining behavior (e.g., Springer & Keil, 1991).

There is an ongoing debate as to whether this sensitivity to domain difference reflects a cognitive bias to assume that animal and artifact concepts are fundamentally different in nature (e.g., Brandone & Gelman, 2009; Gelman, 2003; Goswami, 2001; Keil, 1989; Rattermann & Gentner, 1998; Wellman & Gelman, 1998; see Medin, Lynch & Solomon, 2000 for an overview). According to Gelman (2003), children instinctively recognize that natural kinds are rich in 'essence' below the surface, whereas artifacts are "unlikely to have a rich cluster of nonobvious properties that are intrinsic to the objects being classified" (Gelman, 2003, p. 49). If these different assumptions reflect a fundamental, early-learned distinction between natural kinds and artifacts, this should also be reflected in different induction preferences. Specifically, if 'insides' are assumed to be more important for natural kinds, and children as young as 3 years old can distinguish between properties of natural kinds and artifacts (see Gelman, 2003), then a stronger bias to focus on nonobvious category information should be observed for this domain. Although many studies have examined developmental changes in the way children view natural kinds and artifacts, there are few studies that include both types of stimuli in induction tasks, and the focus of the research was not to directly compare the two domains (for example, Graham, Welder, Merrifield & Berman, (2010); Noles & Gelman, 2012). Finally, none of these studies have looked at whether the developmental trajectory of inductive reasoning for natural kinds and artifacts is the same.

The current study

We aimed to systematically compare the developmental effects of children's inductive reasoning for items that vary in terms of their domain (natural versus artifact) and category structure (featural versus relational). It is difficult to manipulate category structure and domain within existing items since the factors determining category membership vary between domains. For example, within natural kinds, the relationships between category members (e.g., parent / child) or subtle differences in the configuration of parts (e.g., consider alligator vs. crocodile; jaguar vs. leopard) are often critical. In contrast, artifacts are more likely to be distinguished by one or more isolated features (e.g., a knife must have a blade; Levin, Takarae, Miner & Keil, 2001). Novel stimuli allow us full control over category membership information and perceptual similarity, and enables us to manipulate category structure (featural vs. relational) and domain (natural kind vs. artifact) orthogonally. We created two sets of novel natural kind and artifact stimuli. Category membership in one set could only be determined by a featural rule. Category membership in the other set could only be determined by a relational rule. Bearing in mind the concerns raised by Gelman and Waxman (2007), we created our artificial stimuli in line with those designed by Badger and Shapiro (2012). Their novel items restricted the level of children's prior knowledge, but the stimuli had more accessible and realistic membership features and the category rules were more "biologically grounded" (Gelman & Davidson, 2013, p. 330). We used plausible featural and relational rules and pre-tested our stimuli with children to ensure that they viewed them as believable examples of natural kinds and artifacts.

Our four-fold paradigm (1) category learning, 2) initial categorization, 3) category induction, 4) final categorization) followed that developed by Sloutsky et al. (2007). However, unlike Sloutsky et al. we used a developmental manipulation to create items that were dissimilar in appearance, yet from the same category (as in Badger & Shapiro, 2012). The stability of category membership over development was demonstrated and reinforced through an animation in which a juvenile insect

transformed into an adult insect or an incomplete vehicle transformed into a complete vehicle. This

animation was used to demonstrate two different categories of insect (rocky bugs vs. desert bugs) or

vehicle (town trudges vs. country trudges).

Although young children view many properties of natural kinds as stable across

development (e.g., a person's language), there are certain properties that young children view as

unstable across development (e.g., race; Kinzler & Dautel, 2012). Therefore, it is possible that an

apparent perceptual choice may actually reflect a decision to match on "age-category". However,

this is unlikely to occur in our study since Badger and Shapiro (2012) showed – using the same

paradigm and properties with similar stimuli – that children as young as 3 years old did not base

their induction choices on the age of the items. Specifically, young children were no more likely to

select the category choice when this was a dissimilar looking adult (same age as the target) than

when this was a dissimilar looking juvenile (different age to the target). Instead, they consistently

either made choices based on appearance information, or category membership regardless of

stimuli 'age'. This confirmed that children were able to see beyond the age of the stimuli when

considering its properties. Finally, in order to examine the developmental effects of inductive

reasoning for domain and category structure, we tested children aged 4 to 9 years old.

Method

Participants

Four hundred and three primary school children participated: 105 Reception (4-5 years,

mean = 5.04, range = 4.10-5.09 years), 109 Year 1 (5-6 years, mean = 6.04, range = 5.10-6.11 years);

102 Year 2 (6-7 years, mean = 7.04, range = 6.10-7.09 years); 41 Year 3 (7-8 years, mean = 7.08,

range = 7.02-8.01 years), and 46 Year 4 (8-9 years, mean = 8.07, range = 8.02-9.01 years); 214 males

and 189 females. Each child was randomly assigned to either the featural or relational condition, and

completed both a natural kind task and an artifact task. All tests were counterbalanced to avoid order effects between the natural kind and artifact tasks.

An additional 30 children participated in a stimulus similarity pre-test (10 Reception, 10 Year 1 and 10 Year 2; range = 4.02-7.01; 15 males and 15 females), and a further 32 children participated in a domain identification and stimulus naming pre-test (12 Reception, 10 Year 1 and 10 Year 2; range 4.02-7.00; 16 males and 16 females).

Stimuli

There were two domains of stimuli: natural kinds and artifacts. The natural kinds were newly designed novel insects (rocky bug vs. desert bug) and the artifacts were newly designed novel vehicles (town trudge vs. country trudge). Stimuli were either featural or relational in their category membership. From these, four novel groups were created: natural featural, natural relational, artifact featural and artifact relational. In the natural featural condition the two categories of insects differed on head shape. In the natural relational condition the two categories of insects differed on the relationship between the size of mandible versus sting. In the artifact featural condition the two categories of vehicles differed on wheel type. In the artifact relational condition the two categories of vehicle differed on the relationship between the size of the front loader versus the backhoe.

Figures 1a and 1b show the smallest and largest sizes of each feature in the relational condition. However, there were 3 possible sizes (small, medium and large) so that focusing on one feature would not have allowed accurate categorization; the relative size of the two features was critical (e.g., medium front loader & small backhoe = country trudge; medium front loader & large backhoe = town trudge; rules explained in procedure, below).

[Figures 1a and 1b here]

Simple animations were used to demonstrate a transition from juvenile to adult bug for the natural kinds (see Figure 2a) and from incomplete to complete trudge for the artifact kinds (Figure 2b). Each stimulus differed from every other stimulus on at least one dimension (see Appendix A).

[Figures 2a and 2b here]

Similarity pre-tests were conducted to check that the target and distractor (perceptual choice) were significantly more similar in appearance than the target and category choice for every induction triad. Children saw 48 triads in total: 12 featural natural, 12 featural artifact, 12 relational natural and 12 relational artifact. For each triad, they were asked which of the two test items looked most like the target. We analyzed the total number of choices (out of 12) for each condition. One-sample t-tests showed that for all conditions the perceptual distractor was chosen significantly more often than chance (see Table 1). Identification and naming pre-tests were conducted to check that the stimuli were considered to be biologically plausible. Children were shown four images from each category (two juvenile/ incomplete and two adult/ complete) one at a time, in random order. They were asked to state whether the item was a living (natural kind) or a non-living (artifact) thing, and what they believed it to be. Insects were labeled as living 95% of the time, with common responses being 'bug', 'ant' or 'caterpillar'. Vehicles were labeled as non-living 87% of the time, with common responses being 'digger', 'scales', 'shapes' or 'plates'. There was no significant difference in the percentage of correct classifications for insects versus vehicles: t (31) = 1.42; p = .165.

[Table 1 here]

Design

Each child was randomly assigned to either the featural or relational condition and completed both a natural kind task and an artifact task. Category structure (featural vs. relational) and year group (Reception (4-5 years), Year 1 (5-6 years), 2 (6-7 years), 3 (7-8 years) and 4 (8-9 years) were between subjects and domain (natural vs. artifact) was within subjects.

Procedure

Children completed the natural kind task and the artifact task in counterbalanced order, in two sessions. A four-fold procedure was followed for both conditions: category learning, initial categorization, category induction, and final categorization (as in Badger & Shapiro, 2012; Sloutsky et al., 2007).

Category learning

Children took part in two conditions (one natural and one artifact), but were only taught the rules of the condition they were about to take part in; they did not learn both natural and artifact rules at the same time. Rules for differentiating between the two categories in a given task were explained to the children whilst they watched an example of each category transform (see Figures 2a and 2b for examples). Children were told that each item comes in different colors and shapes, and we could only tell the difference based on the identified features (the rules). This ensured children understood that these extra dimensions (color and shape) were irrelevant.

In the featural condition, children were taught that they could differentiate items based on a single feature: head shape of desert bug versus rocky bug (natural featural condition); wheel type of country trudge versus town trudge (artifact featural condition). So children did not see these features as arbitrary, the function of this critical feature was explained in the following ways: *"desert bugs live in the sand and have round heads for making soft burrows"; "rocky bugs live in rocks and have sharp pointy heads for digging"; "country trudges are used in fields and have tracked wheels for muddy surfaces"; "town trudges are used in towns and have round wheels to work on paved surfaces".*

In the relational condition, children were taught that they could differentiate items based on the relationship between two features: size of sting versus size of mandibles for desert bug versus rocky bug (natural relational condition); size of front loader versus size of backhoe for country

trudge versus town trudge (artifact relational condition). The function of this critical feature combination was explained in the following way: "*desert bugs live in the sand and have larger mandibles to catch prey walking above on the sand, and a smaller sting to 'knock-out' prey*"; "*rocky bugs live in rocks and have smaller mandibles to hold onto prey in between rocks, and a larger sting to kill prey*'; "*country trudges are used in fields and have a larger front loader to move soil, and a smaller backhoe to move stones*", "*town trudges are used in towns and have a smaller front loader to carry materials, and a larger backhoe to break concrete*".

For all items, the experimenter pointed to the features on the screen as they gave these explanations, and it was not necessary for the child to know the terms used to label each feature. Nevertheless, terms were explained to the child if they asked (e.g., if a child asked what the mandibles were, it was explained that mandibles were the bug's jaws/ mouthparts used to catch and eat food).

Following the initial introduction to the two categories in one task condition, eight randomized categorization trials were shown and children had to identify the category at the beginning and at the end of each transformation. So for example, a child had to correctly identify an incomplete trudge, as well as the complete trudge at the end of the animation transformation, for that trial to be scored as correct. Corrective feedback was given, and the rule was reiterated, for example *"Yes well done, it is a country trudge because it is used in fields and has tracked wheels for muddy surfaces" (incomplete trudge); "Yes well done, it is still a country trudge and it is still used in fields and has tracked wheels for muddy surfaces" (complete trudge)*. This was to remind children of the functional importance of the features and the stability of the category.

Initial Categorization

Twelve further categorization trials were presented in random order, six of each category within the task condition. Children were not given feedback and the rule was not reiterated. This section tested whether the children could correctly identify the categories without help.

Category Induction

Twelve induction trials were presented in random order, with six targets from each category. Each trial began with a start to finish transformation (e.g., juvenile to adult) of the target category, followed by a triad of the target and two test items, one perceptual distractor and one category choice (see Figures 3a and 3b). Each target item was given a hidden property relating to its function or behavior, for example, *'can be used at night'* or *'sleeps during the day'* (based on types of hidden properties used in Badger & Shapiro, 2012; see Appendix B). The child was instructed to point to the test item which shared this property. If the child chose the perceptual distractor test item, this was coded as a perceptual choice; choosing the same-category test item was coded as a categorical choice. Due to ongoing debate as to the influence of labels on induction (see Noles & Gelman, 2012), no labels were used: items were referred to as 'this one'. The task structure and hidden properties provided were the same for both the featural and relational conditions. The only difference between the two condition tasks was the rule differentiating the stimuli categories (either featural or relational rules).

Final Categorization task

This final phase followed the same procedure as the Initial Categorization task but with different stimuli, and was included to check whether children could still correctly identify the categories after the induction task.

[Figures 3a and 3b here]

Results

Categorization Performance

In order to be sure that our analyses of induction choices only included children who were proficient at using the categorization rule, we used a very stringent cut-off based on their categorization performance. Specifically, children had to score 10/12 or above in both the initial and final categorization tasks to be included in any further analyses (scores of 10/12, 83%, are significantly above chance; binomial test proportion = 0.5, *p* = .04). Fifty-five children did not meet this criterion for the categorization tasks in either domain (43 Reception (4-5 years); 9 Year 1 (5-6 years), and 3 Year 2 (6-7 years); 37 children did not meet criterion for one of the two domain tasks (17 Reception; 15 Year 1; 4 Year 2, and 1 Year 4). In order to be absolutely confident that the children included in our final analysis were fully competent at using the categorization rule, these children's data were removed. This resulted in a final total of 311 children: 45 Reception: 42/63 featural and 3/42 relational; 85 Year 1: 51/55 featural and 34/54 relational; 95 Year 2: 52/52 featural and 43/50 relational; 41 Year 3: 21/21 featural and 20/20 relational, and 45 Year 4: 25/25 featural and 20/21 relational.

These children scored highly and above chance on the initial and final categorization tasks: **Natural:** Initial *M* = 97% (*SD* = 6.82), *t* (310) = 123.87; *p* < .001; Final *M* = 95% (*SD* = 7.99), *t* (310) = 100.36; *p* < .001. **Artifact:** Initial *M* = 97% (*SD* = 7.26), *t* (310) = 113.16; *p* < .001; Final *M* = 97% (*SD* = 7.18), *t* (310) = 115.59; *p* < .001. **Featural:** Initial *M* = 98% (*SD* = 3.94), *t* (381) = 238.93; *p* < .001; Final *M* = 98% (*SD* = 4.43), *t* (381) = 211.20; *p* < .001. **Relational:** Initial *M* = 96% (*SD* = 10.07), *t* (239) = 70.41; *p* < .001; Final *M* = 94% (*SD* = 10.42), *t* (239) = 64.83; *p* < .001. However, since so few Reception children passed the strict inclusion rule, only Years 1-4 were considered in the analyses.

Induction Performance

A mixed GLM was conducted with children from Year 1-4 (5-9 years; 266 in total) to examine whether the percentage of category choices made in the induction task differed by year group,

domain and category structure. Although children completed the natural and artifact conditions in counterbalanced order, it is possible that they transferred a strategy from one domain to the other. We therefore included the order in which the tasks were completed as an additional factor (natural then artifact or artifact then natural). A significant effect of year group was found, F (3, 250) = 10.72; p < .001, η_p^2 = .11, with the percentage of category choices increasing with year group (Year 1 (5-6 years): $M\%$ = 39.30, SD = 42.45; Year 2 (6-7 years): $M\%$ = 50.48, SD = 44.29; Year 3 (7-8 years): $M\%$ = 70.24, SD = 39.14; Year 4 (8-9 years): $M\%$ = 71.50, SD = 39.99). No significant effect of domain was found (natural vs. artifact), F (1, 250) = 2.55; p = .11, η_p^2 = .01. A significant effect of category structure was found (featural vs. relational), F (1, 250) = 21.09; p < .001, η_p^2 = .08, with children making fewer category choices in the relational condition (featural: $M\%$ = 64.56, SD = 41.40; relational: $M\%$ = 41.50, SD = 45.35). There was a non-significant effect of order, F (1, 250) = .25; p = .62, η_p^2 = .001, showing that the counterbalancing of tasks adequately controlled for potential carryover effects. All interactions were non-significant.

In order to examine age-related changes in children's induction preference in more detail, we calculated the percentage of children showing each type of bias (perceptual or categorical) at each age / year group. Since the GLM indicated that the pattern varied by category structure but not domain, we calculated these percentages separately for the featural and relational conditions, collapsed across the two domains. Children making at least 10/12 category choices in the induction task were considered to have a category bias (% of category choices significantly above chance, binominal test proportion 0.5, p = .04); children making 2/12 or fewer category choices were considered to have a perceptual bias (% of category choices significantly below chance, binominal test proportion 0.5, p = .04). Children making between 3 and 9 category choices had no significant bias. Figure 4a shows that for the featural condition, we observed a gradual transition from only 38% of children showing a categorical bias in Year 1 through to the majority of children (84%) showing a categorical bias in Year 4. Figure 4b shows the same trend for the relational condition, however the

percentage of children showing a category bias is delayed by a year in the relational versus featural condition.

[Figure 4a and 4b here]

Discussion

Previous research into inductive reasoning has produced contradictory results (e.g., Gelman & Markman, 1986; Sloutsky et al., 2007). We present the first systematic investigation of the developmental trajectory of category induction when both category structure and domain were explicitly manipulated.

Our data yielded three main findings. Firstly, we found clear developmental effects with a preference for perceptual induction in the younger age groups, which transitions to a preference for category induction in the older age groups. Secondly, children's inductive preference was affected by the category structure of the items. Although the developmental pattern was similar for both category structures, it took a further year for the majority of children to show a category bias in the relational condition, compared with the featural condition. Finally, we found no significant difference in inductive preference between the domains of natural kind and artifact.

Developmental Differences

We examined the development of category induction in children from age 4 to 9 years (UK school years Reception through to Year 4). In all conditions, we observed a gradual transition from a preference for perceptual induction in the youngest year groups through to a preference for category induction in the older year groups. Importantly, we only included data from children who passed our strict inclusion criteria in both initial and final categorization tasks. In the featural condition, the majority of children in all year groups met criterion on our categorization tasks. In the

relational condition, very few Reception children met criterion on our categorization tasks. We

therefore didn't include this year group in our inferential analyses.

Our exclusion rate in the Reception year group for the relational condition was 93%. This

was surprising given that the relational condition was adapted from Sloutsky et al. (2007), who had a

much higher success and inclusion rate (exclusion rate of 20%). This difference may partly be

explained by our more stringent inclusion criteria (10/12 – 83% – on both initial and final

categorization tasks, rather than 6/8 – 75%), but there may also be more general differences in our

samples (e.g., general attainment levels and socio-economic status). Interestingly, our Year 1

children in the relational condition showed a very similar pattern of performance to the sample of

Sloutsky et al. Even though these children were a year older than Sloutsky's sample, the majority

were biased towards perceptual induction, with only 25% of children showing a category induction

bias. The pattern was similar in Year 2, and it wasn't until Years 3 and 4 that the majority of children

showed a bias towards category induction.

These data are consistent with perceptual to categorical transition accounts of inductive

reasoning (Badger & Shapiro, 2012) and analogical reasoning (Gentner, 1988; Gentner &

Rattermann, 1991; Rattermann & Gentner, 1998). However, for the first time in the inductive

reasoning literature, this study also considered the developmental effects when category structure

and domain of the stimuli were manipulated systematically. These two novel findings are discussed

below.

Category structure

This is the first study to show that the category structure of items used in inductive

reasoning tasks (featural or relational) influences children's preference for using perceptual or

category induction. The progression from a perceptual to categorical preference was delayed a year

for the relational category structure with the turning point at Year 3 (7-8 years old) rather than Year

2 (6-7 years old; as for the featural category structure). This finding goes some way to uncovering

why previous work has produced contradictory findings: studies using stimuli that can largely be

categorized based on individual features (e.g., Gelman & Markman, 1986) have observed a category-

bias in children at a younger age than studies using categories that were defined by relational rules

(e.g., Sloutsky et al., 2007). There are three possible explanations for the delay we observed for

relational categories. Firstly, the categorization rule in the relational condition may have been too

complicated for children to follow. More children failed to meet our categorization criteria in the

relational condition suggesting that learning the categorization rule was harder. However, the

children included in our analyses scored 10/12 or above in both the initial and final categorization

tasks, showing their proficiency in remembering and applying the relational categorization rules over

the entire testing session. Secondly, it could be argued that children struggled more with the

relational category structures because these types of rules are less realistic than featural rules.

However, several researchers have argued that in the real world, categories are more often defined

in terms of relational rather than featural structures (e.g., Kittur et al., 2004; Saiki & Hummell, 1998;

see also Barsalou, 1993). Thus, it is unlikely that our relational categories were intrinsically less

realistic. Finally, it is possible that another developmental change occurring at this time is influential

when making induction decisions. Specifically, this effect is consistent with theories of cognitive

control and level of task demand (Andrews & Halford, 2002; Halford, Wilson & Phillips, 1998; Zelazo

& Frye, 1998; Zelazo & Müller, 2002). For example, according to Andrews and Halford (2002), the

"Relational Complexity" of a task is defined as the number of relations that must be considered

jointly, and the ability to process more complex relations increases with age. It is not possible to

make a direct comparison between our task and the age estimates provided by Andrews and

Halford, since our tasks are so different. Nevertheless, our relational condition is clearly more

complex than our featural condition. In our relational condition, two features must be identified and

compared in order to identify category, compared to identifying one feature in the featural

condition. This poses an additional processing step before making the decision as to which item

shares hidden properties with the target, compared to our featural condition. Although younger

children may have struggled more with the complex task due to their less developed executive

function and inability to hold and process multiple forms of information, we can not be certain of

the cause of the delay without further investigation. What we can say though, is that future

induction studies need to carefully consider the category structure of their items before they can

directly compare their findings against other studies.

Domain

We found no evidence of domain differences in inductive reasoning (counter to predictions

derived from Goswami, 2001; Keil, 1989; Rattermann & Gentner, 1998; Wellman & Gelman, 1998).

In particular, young children were no more likely to make decisions based on category knowledge for

natural kind items. This is inconsistent with the notion that there is an early bias to assume that

hidden properties are more diagnostic for natural kinds (e.g., Gelman, 2003). Children did not

appear to instinctively recognize an 'essence' in the natural kinds as being more important for

induction decisions. This suggests that domain differences observed in other tasks do not extend to

induction. In our study, the key features determining category membership were equally nonobvious

for natural kinds and artifacts, and equally associated with the function or behavior of an item. We

consider three explanations for our null effect. Firstly, it is possible that children did not differentiate

between domains of the stimuli as they did not recognize the natural kinds as biologically plausible

and therefore treated all items in the same way. However, the domain identification and stimulus

naming pre-tests revealed that children spontaneously assigned domain-appropriate names to the

items almost all of the time, making this explanation unlikely. In fact, the pre-test was a very harsh

test of the plausibility of our stimuli since children were only shown the images briefly, had never

seen them before, and were given no background or context. In contrast, in the main experiment,

children were given a domain-plausible description which provided some context about where you

might find the item and an insight into its function or behavior. Secondly, it is possible that causal assumptions are not elicited by novel stimuli, as children do not believe the items to be real natural kinds. However, our stimuli were no less realistic than pictures children are frequently exposed to in books and on TV, and children readily engage with such stories appropriately (e.g., Ganea, Ma & DeLoache, 2011). In fact, during the experiment, children often spontaneously commented on the stimuli as if they were pictures depicting real things (e.g., *"I've seen a rocky bug in my garden"* or *"I know that's a town trudge because they go down my road sometimes").* This suggests that the images were seen as plausible and real, making this second explanation unlikely. Finally, the developmental effects observed in both domains suggest that over time, children gain an understanding of the importance of category membership in both natural and artifact categories. This is the first investigation of this developmental trajectory in a study with well-matched stimuli. Perhaps our findings show that instead of placing more importance of category membership onto natural kinds, children learn to increasingly acknowledge the importance of category membership for induction decisions, in both natural kind and artifact domains. Thus, children show a similar development of inductive reasoning in both domains. We cannot make firm conclusions based on a null effect, and our stimuli are 'artificial' to allow matching across the domains. However, these findings are consistent with a domain-general transition from a perceptual to a categorical bias in inductive reasoning.

Conclusion

Using a paradigm that enables careful separation of perceptual versus categorical responses, our results support theories of a gradual developmental transition from a perceptual to a categorical bias with age. This trajectory appears to be domain-general for natural and artifact kinds. This would suggest that as children begin to appreciate the importance of category membership through increased knowledge, they understand and apply this in induction tasks. Our first novel finding is the impact of category structure on children's development of inductive reasoning. We found that the

development of a category induction bias is delayed by approximately a year for the more complex

relational category structures. So, although children are already beginning to appreciate the

importance of category membership in induction, their ability to apply this is impeded by more

complex (relational) category membership structures. This is important when considering previous

research methodologies. Whereas Gelman and Waxman (2007) argue that Sloutsky et al (2007)

underestimated children's category induction capabilities because their stimuli were novel and

arbitrary, we argue that the category structure of the items presented is more crucial. Specifically, in

previous research, a preference for perceptual induction was present when more complex relational

category structures were used (as in Sloutsky et al., 2007), and a category induction preference was

present when the categories could be distinguished based on individual features (as in Gelman &

Markman, 1986). Importantly, we found that although children showed a category bias at a later age

for relational structures, there was a gradual increase in category induction in both conditions, which

is not consistent with an early bias for category induction (Gelman, 2003). For the first time, our

results have shown that category structure of items should be taken into consideration when

interpreting developmental differences in category induction. Our second novel finding is the similar

developmental pattern of inductive reasoning for both novel natural kinds and artifact kinds. We

found that although children could name the natural kinds as natural, and the artifact kinds as

artifacts, they applied the same importance of category membership to both. This category bias

increased with age and suggests that once children are able to acknowledge the importance of

category membership, they do so equally for natural kinds and artifact kinds.

References

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive

development. *Cognitive Psychology, 45,* 153-219.

Badger, J. R., & Shapiro, L. R. (2012). Evidence of a transition from perceptual to category induction

in three to nine year old children. *Journal of Experimental Child Psychology, 113,* 131-146.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a

compositional system of perceptual symbols. In A. F. Collins, S. E. Gathercole, M. A. Conway,

& P. E. Morris (Eds.), *Theories of Memory* (pp. 29-101). Hillsdale, NJ: Erlbaum.

Backscheider, A., Shatz, M., & Gelman, S. A. (1993). Preschoolers' ability to distinguish living kinds as

a function of regrowth. *Child Development, 64,* 1242-1257.

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for

infants. *Developmental Psychology, 38,* 948-957.

Booth, A. E., Waxman, S., & Huang, Y. T. (2005). Conceptual information permeates word learning in

infancy. *Developmental Psychology, 41, 491-505.*

Brandone, A. C., & Gelman, S. A. (2009). Differences in preschoolers' and adults' use of generics

about animals and artifacts: A window onto a conceptual divide. *Cognition, 110,* 1-22.

Bulloch, M. J., & Opfer, J. E. (2009). What makes relational reasoning smart? Revisiting the

perceptual-to-relational shift in the development of generalization. *Developmental Science,*

12, 114-122.

Carey, S. (1985). *Conceptual development in childhood.* Cambridge, MA: MIT Press.

Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual

    transition from similarity-based to category-based induction. *Child Development, 76,* 583-

    597.

Ganea, P. A., Ma., L., & DeLoache, J. S. (2011). Young children's learning and transfer of biological

    information from picture books to real animals. *Child Development, 82,* 1421-1433.

Gelman, S. A. (1988). The development of induction within natural kinds and artifact categories.

    Cognitive Psychology, 20, 65-95.

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

    University Press.

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction.

    Cognitive Psychology, 66, 327-353.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*,

    183-209.

Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks: Comments on Sloutsky, Kloos, and

    Fisher (2007). *Psychological Science, 18(6),* 554-555.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-

    obvious. *Cognition, 38,* 213-244.

Gentner, D. (1988). Metaphor as structure mapping: the relational shift. *Child Development, 59,* 47-

    59.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J.

    P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development (pp.*

    225-277). London: Cambridge University Press.

Goswami, U. (2001). Analogical reasoning in children. In D. Gentner, K. J. Holyoak & B. N. Kokinov

(Eds.), *The analogical mind: Perspectives from cognitive science* (p. 437-470). Cambridge,

MA: MIT Press.

Gottfried, G. M., & Gelman, S. A. (2005). Developing domain-specific causal-explanatory frameworks:

The role of insides and immanence. *Cognitive Development, 20,* 137-158.

Graham, S. A., Welder, A. N., Merrifield, B. A., & Berman, J. M. J. (2010). Preschoolers' extension of

novel words to animals and artifacts. *Journal of Child Language, 37,* 913-927.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know

about animals and artifacts? Domain-specific requests for information. *Psychological

Science, 17,* 455-459.

Halford, G. S. (1993). *Children's understanding: the development of mental models*. Hillsdale, NJ:

Erlbaum.

Halford, G. S., Wilson W. H., & Phillips, S. (1998). Processing capacity defined by relational

complexity: Implications for comparative, developmental and cognitive psychology.

Behavioral and Brain Sciences, 21, 803-865.

Hayes, B. K., McKinnon, R., & Sweller, N. (2008). The development of category-based induction:

Reexamining conclusions from the Induction Then Recognition (ITR) paradigm.

Developmental Psychology, 44(5), 1430-1441.

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and

plants. *Child Development, 67,* 2823-2840.

Jipson, J., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and

nonliving kinds. *Child Development, 78,* 1675-1688.

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

Keil, F. C. (1992). The origins of an autonomous biology. In M. R. Gunnar & M. Maratsos (Eds.), *the Minnesota symposia on child psychology: Vo.l 25. Modularity and constraints in language and cognition* (p. 103-137). Hillsdale, NJ: Erlbaum.

Kemler Nelson, D. G., Egan L. C., & Holt, M. B. (2004). When children ask "What is it?" what do they want to know about artifacts? *Psychological Science, 15,* 384-389.

Kinzler, K. D., & Dautel, J. B. (2012). Children's essentialist reasoning about language and race. *Developmental Science, 15(1),* 131-138.

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2004). Feature- vs. relation-defined categories: Probab(alistic)ly not the same. *Proceedings of the Twenty Six Annual Conference of the Cognitive Science Society,* 696-701.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception and Psychophysics, 63,* 676-697.

Mandler, J.M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infants. *Cognition, 59,* 307-335.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51,* 121-147.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical learning* (p. 179-195). New York: Cambridge University Press.

Noles, N. S., & Gelman, S. A. (2012). Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches. *Developmental Psychology, 48,* 890-896.

Opfer, J. E., & Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition, 105,* 206-217.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: children's performance on a causal-mapping task. *Cognitive Development, 13,* 453 - 478.

Saiki, J., & Hummel, J. E., (1998). Connectedness and part-relation integration in shape category learning. *Memory & Cognition, 26* (6), 1138-1156

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General, 133,* 166-188.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18,* 179-185.

Springer, K., & Keil, F.C. (1991). Early differentiation of causal mechanisms appropriate to biological and non-biological kinds. *Child Development, 62,* 767-781.

Tomlinson, M.T., & Love, B.C. (2010). When Learning to Classify by Relations Is Easier Than by Features. *Thinking & Reasoning, 16,* 372-401.

Wellman, H. M., & Gelman, S.A. (1998). Knowledge acquisition. In D. Kuhn and R. Siegler (Eds.), *Handbook of child psychology,* 4th ed*.,Cognitive Development* (p. 523-573). New York: Wiley.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11,* 37-63.

Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: the development of executive function. *Current Directions in Psychological Science, 7,* 121-126.

Zelazo, P. D., & Müller, U. (2002). The balance beam in the balance: Reflections on rules, relational

complexity, and developmental process. *Journal of Experimental Child Psychology, 81,* 458-465.

Appendix

A: The dimensions on which each stimulus differed, for natural kinds and artifact kinds.

NATURAL KINDS

Head	Color	Body	Eye color	Mandible-sting relation
Angled	Green	Round	White	Large mandible, small sting
Round	Yellow	Triangle	Black	Large sting, small mandible
	Purple			

ARTIFACT KINDS

Wheels	Color	Cabin	Rim color	Front-back loader relation
Tracked	Green	Round	White	Large front, small back
Round	Yellow	Triangle	Black	Large back, small front
	Purple			

B: The nonobvious properties used in both conditions.

NATURAL KINDS

Has thick blood
Eats flies
Comes from America
Has cold blood
Has soft bones
Moves slowly
Sleeps during the day
Has a good sized heart
Likes to be warm
Doesn't eat in winter
Lays eggs
Has a diminutive stomach

ARTIFACT KINDS

Uses thick oil
Made in America
Has a small engine
Has lots of buttons inside
Moves slowly
Can be used at night
Needs warm oil
Has a good sized engine
Doesn't work in winter
Moves quickly
Has a chair that moves
Has to be kept inside at night

Table 1. *Frequency with which children paired items designed to be higher in perceptual similarity in each condition.*

Category structure	Domain	Test item	Chosen as most perceptually similar (%)	One-sample t-tests
Featural	Natural kind	Perceptual distractor	99%	$t(29) = -481.06; p < .001$
Featural	Natural kind	Category choice	1%	$t(29) = -629.08; p < .001$
Featural	Artifact kind	Perceptual distractor	91%	$t(29) = -83.51; p < .001$
Featural	Artifact kind	Category choice	9%	$t(29) = -104.45; p < .001$
Relational	Natural kind	Perceptual distractor	100%	$t(29) = -1141.00; p < .001$
Relational	Natural kind	Category choice	0%	-
Relational	Artifact kind	Perceptual distractor	96%	$t(29) = -138.22; p < .001$
Relational	Artifact kind	Category choice	4%	$t(29) = -177.71; p < .001$

Figure Captions

Figure 1(a). Examples of each type of bug for the featural condition (left = rocky bugs, right = desert bugs) and relational condition (top = desert bugs, bottom = rocky bugs).

Figure 1(b). Examples of each type of trudge for the featural condition (left = town trudges, right = country trudges) and relational condition (top = town trudges, bottom = country trudges).

Figure 2(a). Example of a natural kind transformation used in the Categorization task (the child is asked to name the infant and then the adult).

Figure 2(b). Example of an artifact transformation used in the Categorization task (the child is asked to name the incomplete object and then the complete object).

Figure 3(a). Example of an artifact induction trial. The child is shown the target transform from incomplete to complete artifact. The complete target remains on-screen whilst the two test items appear, creating the induction triad (target, category choice, perceptual choice).

Figure 3(b). Example of a natural kind induction trial. The child is shown the target transform from infant to adult. The complete target remains on-screen whilst the two test items appear, creating the induction triad (target, category choice, perceptual choice).

Figure 4(a). The percentage of children in each year group (Year 1 – Year 4) showing each type of induction bias for the featural condition.

Figure 4(b). The percentage of children in each year group (Year 1 – Year 4) showing each type of induction bias for the relational condition.