DOCTORAL THESIS

An investigation into children's inductive reasoning strategies

what drives the development of category induction?

Julia Badger

An investigation into children's inductive reasoning strategies: What drives the development of category induction?

By

Julia Rachel Badger

Doctor of Philosophy

Aston University

July 2011

Aston University

An investigation into children’s inductive reasoning strategies: What drives the development of category induction?

Julia Rachel Badger

Doctor of Philosophy 2011

Thesis Summary

In a series of studies, I investigated the developmental changes in children’s inductive reasoning strategy, methodological manipulations affecting the trajectory, and driving mechanisms behind the development of category induction. I systematically controlled the nature of the stimuli used, and employed a triad paradigm in which perceptual cues were directly pitted against category membership, to explore under which circumstances children used perceptual or category induction.

My induction tasks were designed for children aged 3-9 years old using biologically plausible novel items. In Study 1, I tested 264 children. Using a wide age range allowed me to systematically investigate the developmental trajectory of induction. I also created two degrees of perceptual distractor – high and low – and explored whether the degree of perceptual similarity between target and test items altered children’s strategy preference. A further 52 children were tested in Study 2, to examine whether children showing a perceptual-bias were in fact basing their choice on maturation categories. A gradual transition was observed from perceptual to category induction. However, this transition could not be due to the inability to inhibit high perceptual distractors as children of all ages were equally distracted. Children were also not basing their strategy choices on maturation categories. In Study 3, I investigated category structure (featural vs. relational category rules) and domain (natural vs. artefact) on inductive preference. I tested 403 children. Each child was assigned to either the featural or relational condition, and completed both a natural kind and an artefact task. A further 98 children were tested in Study 4, on the effect of using stimuli labels during the tasks. I observed the same gradual transition from perceptual to category induction preference in Studies 3 and 4. This pattern was stable across domains, but children developed a category-bias one year later for relational categories, arguably due to the greater demands on executive function (EF) posed by these stimuli. Children who received labels during the task made significantly more category choices than those who did not receive labels, possibly due to priming effects. Having investigated influences affecting the developmental trajectory, I continued by exploring the driving mechanism behind the development of category induction. In Study 5, I tested 60 children on a battery of EF tasks as well as my induction task. None of the EF tasks were able to predict inductive variance, therefore EF development is unlikely to be the driving factor behind the transition. Finally in Study 6, I divided 252 children into either a comparison group or an intervention group. The intervention group took part in an interactive educational session at Twycross Zoo about animal adaptations. Both groups took part in four induction tasks, two before and two a week after the zoo visits. There was a significant increase in the number of category choices made in the intervention condition after the zoo visit, a result not observed in the comparison condition. This highlights the role of knowledge in supporting the transition from perceptual to category induction. I suggest that EF development may support induction development, but the driving mechanism behind the transition is an accumulation of knowledge, and an appreciation for the importance of category membership.

Child development; Inductive reasoning; A gradual transition from perceptual to category induction; Interactive education; Executive function.

Acknowledgements

I would like to thank Laura Shapiro for her continued support and guidance, and for always encouraging me to achieve more than expected.

This project was funded by the Economic and Social Research Council, but would not have been possible without the cooperation of the staff and children at the many schools in which I tested. I owe great thanks to the Educational team at Twycross Zoo for their expertise, time, and the Conservation Welfare Fund, but especially to Joanna Hardy, who has continually supported the collaboration. I would also like to thank Michelle Ellefson from the University of Cambridge for her advice and guidance during my executive function work.

Thank you to my family, Sue, Paul, Jo and Jessica, who have always supported me, who have generously trialled my experiments, and who always knew my career would involve playing games. A special thanks must go to my mum for patiently proof reading every chapter of my thesis.

Finally, I would like to thank Graham Smith for the walks, talks, support and endless encouragement, for which I will always be grateful.

List of Contents

List of Figures

List of Tables

Chapter 1: Introduction

In everyday life we constantly come across new and unfamiliar items. The ability to use previous knowledge to categorise and generalise information about these items is called inductive reasoning, and is a fundamental part of human thinking. Induction enables us to generalise properties from a known item onto an unknown item, which in turn allows us to effectively treat distinct items as similar. For example, if I came across an unfamiliar item, I might think, 'it looks and sounds like all cats I have ever seen before, therefore I believe this is a cat. If this is a cat, then I believe it will drink milk and hunt at night'. Of course, although induction allows us to go beyond what we already know through generalisations, our generalisations are not guaranteed to be correct. Human reasoning depends crucially on our ability to attend to relevant information when making generalisations, which becomes most challenging when the relevant information is not immediately obvious. Within the natural world for example, subtle cues are sometimes better predictors of behaviour than overall appearance (Canadian goslings can only be distinguished from mallard ducklings by subtle differences in their bills and feet, even though the two species have different diets and migration habits, and grow up to look highly dissimilar). Adults can use both perceptual cues and category membership to generalise properties or behaviour from a known item onto a novel item, dependant on the context. The extent to which young children are able to use both perceptual and category cues when making induction decisions remains under debate (Gelman, 2003; Gelman & Markman, 1986; Gentner, 1988; Opfer & Bulloch, 2007; Medin & Ortony, 1989; Sloutsky & Fisher, 2004; Sloutsky, Kloos & Fisher, 2007). The ways in which induction strategies develop, and the mechanisms underlying the development of children's induction strategy, are also still unclear.

To understand the development and mechanism behind inductive reasoning is crucial to understanding how we learn to make sense of the overload of information we experience every day. Most of our understanding about the world develops through inferences, rather than through teaching or observation. Exploring inductive reasoning allows us to examine how humans apply 'learned' information to situations and items we have never been taught about directly. The questions that remain are: whether children's induction strategy develops over time, or whether they have the same default induction strategy as adults; whether inductive reasoning is learned through experience, or due to a change in cognitive maturation.

Two broad accounts have been proposed to explain how children develop the ability to focus on category-membership cues when generalising about the properties or behaviour of a new item. Firstly, researchers in support of the category-bias account state that children have a natural bias to use category induction when the relational connections between items are clear (Gelman, 2003; Gelman & Markman, 1986; Opfer & Bulloch, 2007). Secondly, researchers in support of the perceptual-bias account state that children have a natural bias to use perceptual induction, and only at a later stage do they begin to understand the importance of category membership and apply it during induction (Gentner, 1988; Sloutsky et al., 2007). Although these two accounts make clear predictions for young children's reasoning strategies, evidence exists to support both category (e.g., Gelman & Markman, 1986) and perceptual (e.g., Sloutsky et al., 2007) biases in young children, which currently makes it impossible to resolve this conflict. The difficulty is to separate the two potential inductive strategies realistically; in real-life, perceptual similarity is often highly correlated with category membership (Heit & Hayes, 2005). These contradictory findings, with which I am concerned, could be due to a multitude of methodological and stimuli ambiguities and shall be discussed in greater detail in the next section of the introduction.

I shall begin my thesis by discussing previous research into inductive reasoning – framed in terms of the two broad accounts I have just summarised – and by highlighting key remaining questions and issues from the literature. Each subsection of the introduction outlines an outstanding concern which will be further addressed in the experimental chapters. The overall aim of this thesis is to further examine whether children's natural inductive bias is perceptual or category; to investigate the developmental trajectory of induction; to explore whether stimuli and methodological manipulations can influence inductive strategy (and therefore provide some explanation for previous contradictory findings), and finally, provide some explanation about the driving force or mechanism behind the development of inductive reasoning.

1 The two broad accounts of inductive reasoning

The majority of research supporting both the perceptual-bias and the category-bias accounts has used the triad paradigm when investigating induction strategies in young children, as it neatly disambiguates children's inductive strategy preference (Bulloch & Opfer, 2009; Gelman & Markman, 1986; Hayes & Thompson, 2007; Opfer & Bulloch, 2007; Sloutsky et al., 2007). Using three images, the triad paradigm directly pits perceptual preference against category preference. A target item is presented alongside two test items, one of which matches the target in category but not overall appearance (the category choice), whereas the other test item matches the target in overall appearance but not category membership (the perceptual choice). The child is told a hidden property of the target – something that cannot be seen – and asked to generalise this property onto one of the two test items, for example, "If this one [pointing to the target] has thick blood, does this one [pointing to the perceptual choice] or this one [pointing to the category choice] also have thick blood?" (taken from Sloutsky et al., 2007). Some versions of this paradigm present hidden properties for the two test items, and the child has to decide which of the two properties can be generalised to the target item. For example, a participant might see a triad consisting of a flamingo, a bat and a blackbird; underneath the flamingo it says, "This bird's heart has a right aortic arch only", and underneath the bat it says, "This bat's heart has a left aortic arch only". The participant must decide whether the blackbird's heart has a right aortic arch or a left aortic arch (taken from Gelman & Markman, 1986). If a child generalises a hidden property between the target and the perceptual choice, then they are classified as using perceptual induction. If a child generalises a hidden property between the target and the category choice, then they are classified as using category induction. All of the inductive reasoning tasks I have designed and used throughout this thesis use the triad paradigm.

1.1 The Category-bias account

It is widely agreed that adults do not solely make categorisation judgments on perceptual appearance, as often this is not enough to determine category membership (Kripke, 1971, 1972; Mill, 1843; Putnam, 1970). Instead, categorisation is sometimes based on other, often hidden or non-obvious properties such as diet, life expectancy, gestation period and DNA. Using this alternative categorisation method is something that researchers in support of the category-bias account suggest that both children and adults are able to do, as long as the relational connection between the items is clear. Although perceptual cues can serve as an initial indication as to an item's category, these cues provide insufficient information for categorisation to be accurate. The account claims that children's inductive decisions are driven by category membership rather than a featural or perceptual overlap between items; children show an early bias to attend to hidden, non obvious properties, and make generalisations based on kind-information over appearance (Gelman, 2003). Similarly, these accounts suggest that young children are capable of reasoning analogically, although they can be constrained by their knowledge of the relevant relations (Goswami, 1992; 2001).

It has been shown that young children make induction decisions consistent with category membership, even when the relationship is not perceptually obvious (e.g., generalising "bird properties" to a dissimilar looking bird; Gelman & Markman, 1986; see also Gopnik & Sobel, 2000). Category preference has been shown in children as young as 2 years old when items have shared category labels (Gelman & Coley, 1990; see also Deák & Bauer, 1996), and children aged 4-5 years have shown a clear preference for category information, even when items were presented with dissimilar labels (for example, 'rabbit' and 'bunny'; Gelman & Markman, 1986). A pioneering study into inductive preference was conducted in 1986 by Susan Gelman and Ellen Markman. They tested children aged 4-5 years, on a variety of inductive reasoning tasks with real biological kinds. Triad paradigms were used which directly pitted category membership (same category, but perceptually different) against perceptual appearance (high similar appearance, but different category). Children were told properties about the two test items and had to generalise one of the properties to the target item. Category membership was indicated through identical or synonym labels, for example 'puppy (target), fox (perceptual choice), puppy (category choice)'. It was found that young children were most likely to generalise the same category test item's property to the target, rather than the similar appearance test item's property. The researchers concluded that young children have a natural bias to use category induction.

Research has demonstrated that children aged 4 years old are not only aware of the importance of category membership during induction (Gelman & Markman, 1986) and are able to differentiate between internal and external features of an object (Gelman, 2003), but strikingly, can link the internal features to the object's behaviour (Gelman & Gottfried, 1996; Gelman & Wellman, 1991; Opfer & Gelman, 2001). For example, children understood that if one removed 'stuff' from inside a dog, it would no longer be able to bark (Gelman & Wellman, 1991).

Recent research has suggested that young children are not simply category-biased, but instead can vary their strategy depending on the experimental context. Opfer and Bulloch (2007) investigated children's use of perceptual and category cues by making category information relevant for some trials only. Specifically, children were asked questions about juvenile insects that were either the prey of two adults (category information was irrelevant) or the offspring of two adults (category information was relevant). Children aged 5 to 8 were shown triads of these family-groups or prey-groups, and were asked to generalise properties of a target juvenile to one of two test juveniles. Children consistently chose the juvenile of greatest perceptual similarity to the target in the prey trials, but consistently chose the juvenile whose parents had the greatest perceptual similarity to the target's parents in the offspring trials. Opfer and Bulloch (2007; Bulloch & Opfer, 2009) found that when category information was available and relevant (i.e., in offspring trials), children were capable of using this information to make category-based induction decisions and were able to ignore the perceptual characteristics of the juveniles. However, if no category information was available, then children relied on perceptual similarity (see also Hayes & Thompson, 2007). Thus, children could clearly switch between different cues according to the context of the task. However, it is not entirely certain that children switched their focus because they understood the relational context had changed. Instead, it is possible that children were influenced by the differing layout of the two conditions: the juvenile was presented above the adults in the prey conditions, and below the adults in the offspring conditions (this layout difference is shown in the figures presented in the published paper by Opfer & Bulloch, 2007). This slight difference could have meant that the juvenile may have been more salient in the prey conditions, and the adults more salient in the offspring conditions.

Researchers supporting the category-bias account have commented that occasionally children will *appear* to show a perceptual induction preference rather than a category induction preference, but that this is easily explained: although young children have a natural bias to actively seek category membership or relational connections during induction, they may be held back by performance factors such as knowledge about the relevant relational comparisons. Thus, once children understand the specific relationships that are relevant in a particular context, they will naturally use these relationships over surface characteristics to inform their decisions (Goswami, 1992; 2001). It is suggested that if a child does not recognise the relational connection as a reliable predictor, then they will refer back to perceptual cues (Bulloch & Opfer, 2009; Gelman, 2003). Thus, young children may be capable of using relational knowledge, but only use this information when they recognise the relevance of this relational connection to the task.

1.2 The Perceptual-bias account

The perceptual-bias account states that young children intuitively focus on perceptual information when making induction decisions (Sloutsky et al., 2007), and surface characteristics in analogical reasoning tasks (Gentner, 1988). Children maintain this perceptual or surface bias until they have developed the knowledge base necessary to support a shift towards a focus on category and relational information (a relational shift; Gentner, 1988; Gentner & Rattermann, 1991; Rattermann & Gentner, 1998).

Research supporting the category-bias account suggests that children may have an early tendency to focus on category cues beyond overall appearance. Nevertheless, even atypical category members share some perceptual characteristics (Heit & Hayes, 2005). Thus, the apparent category-based decisions observed in the studies described above may have been influenced by perceptual information (Jones & Smith, 1993; McClelland & Rogers, 2003; Rakison, 2000; Rakison & Hahn, 2004; Sloutsky & Fisher, 2004). In order to address this issue, Sloutsky et al., (2007) investigated children’s default strategies when category and perceptual cues conflict, allowing perceptual-based versus category-based decisions to be disambiguated. Children aged 4 to 5 were trained to categorise novel biological stimuli based on a non-obvious rule (relative number of “fingers” to “buttons”: an item is a Ziblet if it has more fingers than buttons, and is a Flurp if it has more buttons than fingers). All other perceptual features were non-predictive of category membership. Although children were highly accurate in categorisation tasks, they made significantly more perceptual than category-based induction choices, thus demonstrating a clear perceptual induction bias. However, the validity of

Sloutsky et al's design has been questioned (Gelman & Waxman, 2007). Although Sloutsky et al., demonstrated that the children could use the category rule very accurately in the initial and final categorisation tasks, it was unclear whether the children recognised the biological relevance of the category rule. Therefore, the children may not have been sufficiently motivated to apply this rule in the induction task. Gelman and Waxman (2007) argue that the apparent perceptual bias Sloutsky et al., observed may have been an artefact of unrealistic stimuli or arbitrary category rules. Thus, although the children were able to use the taught rules, they considered them irrelevant for induction decisions, and instead used simple perceptual cues.

According to researchers in support of the perceptual-bias account, young children's default strategy is for perceptual induction, and only once they develop the sufficient knowledge of causal relations within a domain, can they shift their focus to category and relational cues (Gentner, 1988; Gentner & Toupin, 1986; Inhelder & Piaget, 1958; Sloutsky et al., 2007). For example, when presented with a straw and a plant stem and asked how the two items were similar, children aged 5 focused on the perceptual similarity and said that both items were long and thin, but by age 9 children could shift their focus to the relational similarity and said that both could also carry water: a relational shift (Gentner, 1988; also see Gentner & Rattermann, 1991; Halford, 1993). As discussed in Ratterman and Gentner (1998), this shift reflects changes in knowledge representation. Specifically, young children have better representations of object properties than of the relations between objects. Only once children have developed sufficient knowledge of the causal relations within a domain, can they shift their attention away from common object properties towards common relational structure. Researchers in support of the category-bias account would disagree, and suggest that instead of domain information being sufficient, children need specific knowledge about each of the relations presented (Goswami, 1992). However, there is an alternative explanation. The proposed transition from perceptual to category induction could reflect a qualitative change in the nature of children's thinking, rather than a gradual transition with increasing knowledge. The qualitative change could be a result of a stage-like shift in cognitive maturity akin to a Piagetian shift from preoperational to concrete operational thinking (Piaget, 1964; Inhelder & Piaget, 1958). Piaget claimed that humans move through four stages sequentially: the sensorimotor stage (0-2 years), the preoperational stage

(2-7 years), the concrete operations stage (7-11) and the formal operations stage (11 onwards). He believed that children's cognition must be fully established in one stage before they are able to move into the next, more abstract and complex way of thinking. Therefore, Piaget would claim that children's natural default for induction is perceptual, and only with time and the increasing sophistication of their cognitive capabilities, would they shift into the next sequential stage and learn to focus their attention on category induction.

2 Methodological ambiguities

In the first section of this introduction I explored how conflicting findings are interpreted in terms of two broad accounts of inductive reasoning: the perceptual-bias and the category-bias accounts. Although there appears to be a plethora of supportive research for both accounts, there are various outstanding methodological ambiguities regarding the methodology and stimuli used: the degree of perceptual similarity, the domain, the category structure, and labelling. In the next section of the introduction I will consider these ambiguities as potential explanations for the contradictory findings.

2.1 Degree of perceptual similarity

The influence of item similarity is not a new phenomenon, and it is widely understood that people use general similarity to make judgments and inferences (Heit & Rubinstein, 1994). There have been two main strains of research regarding the influence of similarity on inductive decisions. Firstly, those researchers interested in the effect of similarity between premise and conclusion for a single property or a homogeneous set of properties (Osherson, Smith, Wilkie, López and Shafir, 1990; Osherson, Stern, Wilkie, Stob & Smith, 1991; Rips, 1975). Secondly, those researchers interested in whether different properties of items lead to different strengths of inference (Carey, 1985; Gelman, 1988; Gelman & Markman, 1986; Heit & Rubinstein, 1994; Nisbett, Krantz, Jepson & Kunda, 1983; Springer, 1992).

2.1.1 Premise and conclusion

It is well known that for adults, the likelihood of generalising properties from one item to another depends on the strength of within-category perceptual similarity (Osherson et al., 1990;

Sloman, 1993). Osherson et al., (1990) discovered that when adults were presented with a variety of statements and required to make inductive inferences about the items, the higher the items were in similarity, the stronger the inductive inference. For example, adults generalised more properties between robins, bluejays and sparrows, compared to robins, bluejays and geese, as the latter are more perceptually dissimilar. They also found that adults would extend stronger inductive inferences to a broader grouping, such as 'mammals', when the category level of the items was more similar. For example, adults would make stronger inferences between hippopotami, rhinoceroses, and all mammals, compared to hippopotami, hamsters, and all mammals, because hippopotami and hamsters differ more than hippopotami and rhinoceroses. In a different study adults were again asked to make inferences based on statements, such as 'all horses on some island have a certain disease, what proportion of mice on the island would have the disease?' and found that the strength of the induction was well predicted by the level of perceived similarity between items (Rips, 1975).

2.1.2 Property effects

Another line of research has considered whether people would make different inductive choices based on the property presented: property effects. For example, when participants were asked 'given that one member of a tribe is obese, what percentage of the other tribe members would also be obese?', on the whole participants estimated that fewer than 40% of the tribe would also share the property of being obese. In contrast, when participants were asked 'given that one member of a tribe has [a certain colour of skin], what percentage of other tribe members would also have this skin colour?', on the whole participants estimated that over 90% of the tribe would also share the property of skin colour (Nisbett et al., 1983). Property effects have also been identified when making inferences about biological properties; participants gave stronger inferences about biological properties between two biologically similar animals, for example, two snakes, than they did for two dissimilar animals, for example, a snake and a worm (Gelman & Markman, 1986).

Both of these branches of research found that the higher the level of perceived similarity between items, the stronger the inductive inferences made. Most recent research investigating inductive reasoning in young children has used the triad paradigm, in which perceptual similarity directly conflicts with category membership. In this instance, the most salient perceptually similar choice must be inhibited before a category induction decision can be made. However, even if the category membership of items is known, overcoming highly similar perceptual distracters is likely to be challenging for young children (Richland, Morrison & Holyoak, 2006). My review of previous induction research highlights that the degree of perceptual similarity between the target and the perceptual distracter has never been systematically controlled between studies. This oversight could partly explain the differences in findings from researchers supporting the category-bias account, and researchers supporting the perceptual-bias account. The strongest evidence for a perceptual-bias comes from studies using artificial categories, which allow perceptual and category cues to be isolated (e.g., Sloutsky et al., 2007). The strongest evidence for a category-bias comes from studies using real kind stimuli where it is impossible to separate the influences of category membership information and perceptual similarity (e.g., Gelman & Markman, 1986). For example, Gelman and Markman (1986) used line drawings of real kinds, which meant that although the target and perceptual choice were perceptually more similar than the target and category choice, they were not as closely matched as the computer generated novel stimuli used in Sloutksy et al's (2007) study, which could be manipulated to an exact degree. This typifies the two methodological styles used by researchers in support of the two accounts.

Considering these evident similarity effects, it is striking that the level of perceptual similarity between stimuli used, as well as the hidden properties proposed during inductive reasoning tasks, has not yet been standardised.

2.2 Domain

There is evidence that children as young as 3 years old are sensitive to object domain when making assumptions about the properties of an object (Backscheider, Shatz & Gelman, 1993; Carey, 1985; Inagaki & Hatano, 1996; Jipson & Gelman, 2007). For example, Jipson and Gelman (2007) found that 3-5 year old children clearly distinguished between artefact and natural kinds when shown video clips of unfamiliar items, and asked to generalise 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 artefacts (Gelman, 1988; Greif, Kemler Nelson, Keil & Gutierrez, 2006; Inagaki & Hatano, 1996; Keil, 1992; Kemler Nelson, Egan & Holt, 2004).

Children also make more subtle distinctions between natural kinds versus artefacts when generalising properties or behaviour. Specifically, when confronted with natural kinds, they focus on the items' 'insides' (Gelman, 2003; Gottfried & Gelman, 2005); will be more likely to assign category membership based on internal properties (Gelman & Wellman, 1991), and appear to be sensitive to underlying causal mechanisms explaining behaviour (Springer & Keil, 1991). It has been found that children as young as 14 months focused on category differences when they associated 'drinking from a cup' as something an animal would do, and associated 'a ride' as something a vehicle would do (Mandler & McDonough, 1996). If these different assumptions reflect a fundamental, early-learned distinction between natural kinds and artefacts, this should also be reflected in different induction strategy preferences. Specifically, if 'insides' are assumed to be more important for natural kinds, and children as young as 3 can distinguish between properties of natural kinds and artefacts (Gelman, 2003), then a stronger bias towards non-obvious category membership information should be observed for the domain of natural kinds.

Although studies directly comparing induction of natural and artefact kinds are limited (see Gelman, 2003 for discussion), there is a plethora of research suggesting that inductive reasoning is domain-specific (Gelman, 2003; Goswami, 2001; Keil, 1989; Rattermann & Gentner, 1998; Wellman & Gelman, 1998). These studies suggest that there are critical differences in the way humans conceptualise natural kinds, which elicit an essentialist bias, versus artificial kinds, which do not elicit an essentialist bias (Gelman, 2003; Keil, 1989; Markman, 1989). An artificial kind is man-made and "...unlikely to have a rich cluster of non-obvious properties that are intrinsic to the objects being classified" (Gelman, 2003, p. 49). Whereas, natural kinds are rich in 'essence' below the surface which children appear to instinctively recognise as important. In 1986, Gelman and Markman compared different types of natural kind categories – animals (e.g., squirrel, snake) and natural substances (e.g., honey, stone) – in a triad paradigm with 4 year olds, and found a preference for category induction for both natural kind domains when the category was known. More recently, Newman and Keil (2008) compared young children's notions of essence for natural kinds and artefacts and found a developmental shift from a domain-general to a domain-specific approach: initially children responded in the same way for artificial and natural kinds, but as they gradually understood the distinct differences, their beliefs and reasoning became domain-specific. Thus, given Newman and Keil's findings, we might expect a different developmental trajectory of induction strategy for natural kinds versus artefacts.

Children's focus on non-obvious, essential, characteristics for natural kinds compared to artefacts should cause them to show a stronger bias towards category information in induction tasks. However, although research has investigated these two domains separately, there has not yet been a systematic comparison of children's induction strategies across these two domains, which means that a direct comparison is currently not possible.

2.3 Category structure

Categories can be deterministic, whereby all items of a category share at least one parameter that is exclusive to that category, or probabilistic, whereby only a percentage of items within a category share at least one parameter but no one feature is shared by all (Kittur, Hummel & Holyoak, 2004). There are two key deterministic structures: featural, whereby items within the same category share a unique feature (e.g., knives have a blade), and relational, whereby items of the same category share the same relational connection between features (e.g., the size relation between sting and pincers for lethal scorpions; see also the Sloutsky et al., 2007 artificial categories of Ziblets and Flurps, determined by the ratio of fingers to buttons). The structure of a category is critical when making generalisations about an unknown item. For example, since knives have a common feature (a blade), identifying the existence 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"). Despite category structure providing a crucial clue to the properties of new items, there is a lack of research into the effect of category structure on the development of induction strategy. One key reason why category structure may impact on children's inductive reasoning is that reasoning about more complex category structures is likely to place greater demands on a child's executive function abilities (Bunge & Zelazo, 2006; Halford, 1993; Halford, Wilson & Phillips, 1998; Zelazo & Frye, 1998; Zelazo & Müller, 2002). Executive function (hereafter, EF) is an umbrella term used to describe a multitude of cognitive skills including planning, task-switching, working memory and inhibition. EF development limits how many dimensions, or relations, can be processed in parallel (Andrews & Halford, 2002;

Halford, 1993; Halford et al., 1998), and impacts on our ability to consciously reflect on our plans, for example, to apply rules to a reasoning task (Zelazo, Frye and Rapus, 1996). Researchers supporting the two broad accounts of inductive reasoning typically use different category structures during testing. For example, although in a real life context animal categories can often be determined through relational cues, the real kind stimuli used in research supporting the category-bias account are more likely to fall into the deterministic featural category structure, being disambiguated almost all of the time through individual features. In contrast, the stimuli used in research supporting the perceptual-bias account are more likely to fall into the deterministic relational category structure, being disambiguated through the relationship between features. Therefore, the contradictory findings regarding the natural default strategy for young children's induction may be partly explained by the category structure of the stimuli used. Category induction for the simpler featural category structures (e.g., stimuli used in Gelman & Markman, 1986) may be less dependent on EF than for the more complicated relational category structures (e.g., stimuli used in Sloutksy et al., 2007). Thus, children may be more likely to use category induction strategies when presented with featural versus relational category structures. However, the influence of stimuli category structure has still not been explicitly investigated in inductive reasoning.

2.4 Labelling

It is well known that adults will use labels to identify categories and infer generalisations based on these categories. For example, if presented with one cat that has a certain property, such as, it drinks milk, then adults will infer that other items labelled as cat are also likely to share this property. However, the impact of labels during development and especially during induction is still in debate; great discussion has emerged as to whether labels do or do not guide attention to categories. This discussion has resulted in two main theories: 1) labels are proxies for categories; 2) labels are features for categories. The first suggests that labels are proxies or 'stand-ins' for categories (Gelman, 2003), linking commonality and kind, and therefore allowing for category-based induction by conveying category membership. The second suggests that labels are features for categories (Sloutsky & Lo, 1999), and enhance the similarity between items, which creates label-based inference rather than category or perceptual inference. If a child uses perceptual cues to generalise properties, then he is using perceptual induction; if a child uses category membership to generalise properties, then he is using category induction. However, if a child generalises properties based on the similarity of labels, then he is using neither category nor perceptual induction, and is instead using label-based induction. There appears to be ample evidence for both interpretations.

2.4.1 Labels are proxies for categories

The notion behind this theory is that labels are used for "...communicating kinds, not constructing them" (Gelman, 2003; p. 183). A great deal of research has shown that labels highlight the commonality between items (Waxman, 1999; Waxman & Booth, 2003). In 1986, Gelman and Markman conducted inductive reasoning studies with labels; this paper became the foundation for the 'labels are proxies' theory. Children aged 4-5 were shown triads of images with one target and two test items; one of the test items matched the target on perceptual similarity, the other matched the target on category membership. An example of the test follows: 'This fish stays underwater to breathe. This dolphin pops above the water to breathe. See this fish? Does it breathe underwater, like this fish, or does it pop above the water like this dolphin?' One experiment matched the labels with the perceptual cues, for example, a shark might look like a dolphin, so when labelled 'dolphin', the label matched the perceptual cues. Another experiment did not match the labels with perceptual cues. For example, a shark does not look like a fish, so when labelled 'fish', the label did not match the perceptual cues. They found that no child consistently chose the perceptual choice, preferring instead to base induction on label names even when it conflicted with the perceptual cues. A different set of children were then asked to complete the same task using synonyms, for example 'rabbit' and 'bunny' instead of matching labels, and children still used category induction. The authors concluded that the category preference could not be due to an overlap of labels (label-based inference) because children used category induction as consistently when provided with synonyms rather than identical labels. They claimed that labels were clearly more than just features of categories. Later research replicated this idea by showing that if a pterodactyl was named 'dinosaur' then children were more likely to generalise its properties to other 'dinosaur' images, even though the pterodactyl looked more like a bird (Gelman & Coley, 1990). Interestingly, it has also been shown that the power of labels for communicating kind is not restricted to known labels; novel labels have been found to be just as effective as inference-rich information (Davidson & Gelman, 1990; Graham, Kilbreath & Welder, 2001).

2.4.2 Labels are features for categories

The notion behind this theory is that labels add an extra level of similarity to items: they are perceptual attributes rather than category membership identifiers. It has been suggested that this extra level of similarity causes a new type of induction: label-based (Sloutsky & Fisher, 2004; Sloutsky, Lo & Fisher, 2001). Label-based induction occurs when children generalise properties between items because they have the same label, which increases the overall similarity between the items. This is instead of using labels as a means of identifying category membership. It has been shown that when an identical word refers to two items, young children tend to categorise these items together and generalise properties between them (Balaban & Waxman, 1997; Gelman & Markman, 1986; Markman & Hutchinson, 1984; Sloutsky et al., 2001; Welder & Graham 2001). It was this idea that sparked the formation of the SINC model: Similarity-Induction-Naming-Categorisation (Sloutsky & Fisher, 2005). The model declares that young children's categorisations and inductions are formed from overall perceptual similarity rather than conceptual knowledge, and labelling of items strengthens their similarity as opposed to aiding category-membership. Sloutsky and Fisher (2004; Sloutsky & Lo, 1999) found that children aged 4-5 years rated identically labelled items as more perceptually similar than those without identical labels, an effect not replicated with adults.

Recent research has questioned the validity of the Gelman and Markman (1986) results based on the fact that certain synonym label-pairs were not only semantically similar but were also co-occurrent (Fisher, 2010). It has been suggested that children's responses on these tasks could in fact be based on lexical priming rather than category induction. For example, bunny could be seen as a lexical prime for rabbit – bunny rabbit – as opposed to another option 'squirrel' which would not be primed – bunny squirrel. Thus suggesting that when children were previously given synonyms and found to show a category bias, they may in fact have been showing a priming bias instead. Recent findings provide support for this theory: 4-6 year old children were introduced to non-co-occurring and co-occurring triads consisting of a target, a related item and an unrelated item: a non-co-occurring triad would be toad (target), frog (related), bird (unrelated); a co-occurring triad would be kitty (target), cat (related), pig (unrelated). Children were shown three doors and told that the items were behind the doors, thus encouraging semantic processing and avoiding label-perceptual conflict. Children were then asked to generalise a hidden property from the target to one of the two test items. Results showed that children struggled to reason as well with semantically similar labels compared to identical labels. Children were more likely to make category-based responses when the semantically similar items were co-occurring in child-directed speech, for example, ‘bunny-rabbit’ or ‘kitty-cat’. This shows that the more semantically similar the labels, the more children will use category induction, which suggests that an apparent category induction bias may in fact be a label induction bias (Fisher, Matlen & Godwin, 2011).

It has also been suggested that in early development, auditory stimuli can be more compelling to children than visual stimuli (Robinson & Sloutsky, 2004); if children are given verbal labels for items, this verbal information will have a greater impact than the visual cues. When children are presented with lots of information to process, they attend to auditory features over perceptual features, showing an auditory preference (Sloutsky & Napolitano, 2003). This research has also shown that in early development, even non-speech sounds tend to attract more attention than corresponding visual input, showing the influential nature of auditory cues during early development. However, there is research that challenges this auditory theory, although it is not directly linked to inductive reasoning: it has been found in 9 month olds that although different words can facilitate categorisation, tones do not appear to have the same effect (Balaban & Waxman, 1997, similar findings from Xu, 2002).

Thus, in summary, it is still unclear about the effect labels have on inductive reasoning, and in what way children construe the labels given.

3 The driving mechanism behind induction

I have discussed the two broad accounts of inductive reasoning: research in support of the category-bias account suggests that children have a natural instinct to use category induction if the relational connection between items are clear, whereas research in support of the perceptual-bias account suggests that children will use perceptual induction until they fully appreciate the importance of category membership, at which point they will begin to use category induction. I have also highlighted various methodological ambiguities from previous research that may explain the contradictory findings of these two accounts: the degree of perceptual similarity, the domain, the category structure, and the use of labels. In the next section of the introduction I shall discuss two of the potential mechanisms behind the development of category induction. As discussed throughout the introduction, different researchers have developed different theories regarding the nature of the development of inductive reasoning. Some researchers theorise that children have an essentialist bias to use category induction as long as the relational connection between items is clear. Other researchers theorise that induction begins perceptually, and develops into category induction with an accumulation of knowledge (e.g., the relational domain knowledge theory; Goswami & Brown, 1990; Rattermann & Gentner, 1998; Goswami, 1992), or that it begins perceptually and transitions into category induction with cognitive maturation (e.g., the cognitive capacity theory; Richland et al., 2006; Thibaut, French & Vezeva, 2010; Halford, 1993; Andrews & Halford, 2002; Morrison, Doumas & Richland, 2011). The cognitive capacity theory suggests that in order for children to successfully complete analogical reasoning tasks and inductive reasoning tasks, a developed EF is required: working memory, cognitive flexibility, inhibition and selective attention. This development requires a cognitive maturation. The relational domain knowledge theory suggests that what children really require is an accumulation of knowledge within a domain, good item comparison skills, and relational language. Firstly, I shall discuss the effect of cognitive capacity on inductive reasoning, and secondly, I shall discuss the effect of an accumulation of knowledge on inductive reasoning.

3.1 The role of cognitive control

Executive function (EF) is an umbrella term used to describe a multitude of cognitive control skills that enable understanding of complex cognitions. The relational complexity (RC) theory (Halford et al., 1998) states that the main limiting factor for reasoning based on complex information is memory capacity. A child's ability to process multiple factors in parallel depends on the sophistication of working memory. It has been argued that children aged 2 years old can only reason at a binary level and it is not until age 5 years that children have the capability to reason at a ternary level (Andrews & Halford, 2002; Halford, 1993). Young children struggle with cue abstraction (rules to specify the relationship between items) due to the increased pressure on cognitive functions such as working memory. Instead, they tend to rely on perceptual similarity (DeCaro, Thomas & Beilock, 2008; Juslin, Jones, Olsson & Winman, 2003; von Helversen, Mata & Olsson, 2010). The development of higher levels of reasoning have also been linked to the development of prefrontal brain structures: a child's EF must improve in order to cope with the high levels of complexity (the cognitive complexity and control theory (CCC); Bunge & Zelazo, 2006; Zelazo & Frye, 1998; Zelazo & Müller, 2002). Children aged 3-4 years were able to reason based on simple rules, with one factor, however were unable to integrate factors. Interestingly, even when children were able to understand the complex rule, applying it became difficult, described as "...the existence of an abulic dissociation between knowing rules and using them" (Zelazo et al., 1996, p. 57). It was proposed that a child must have the ability to consider multiple relations, as well as have inhibitory control, before they can overcome a reliance on the more obvious perceptual cues (Richland et al., 2006). It has been suggested that because children have an underdeveloped pre-frontal cortex (PFC), they are unable to recruit selective attention (focus, inhibition, set-shifting) like adults can, and instead use a compression-based system until their PFC is developed (Sloutsky, 2010). The compression-based system "...subserve[s] category learning by reducing perceptually rich input to a more basic format" (Sloutsky, 2010, p. 1249). In other words, in order to learn new categories, category membership features become part of the overall representation of the item, and the defining features are overlooked. This simplifies the representation, and results in a perceptual induction preference until cognitive maturation, i.e., once the PFC is fully developed.

The effects of an increased cognitive load has been studied with regard to analogical reasoning about shapes, and supports the theory that young children's inability to inhibit competing perceptual information over the relational choice is explained by their poor EFs (Thibaut et al., 2010). Thibaut et al., designed four tasks of varying complexity, and found that young children's and teenagers' reasoning did not significantly differ with complexity: young children showed a consistent perceptual preference, teenagers showed a consistent relational preference. However, children aged 8 years showed slower reaction times on the more complex tasks. Thibaut et al., suggest a continuum of executive control when searching for the correct task choice: young children lack executive control so fixate on the obvious solution, whereas slightly older children, whose executive control is improving, begin to consider solutions beyond the obvious which slows down their processing.

Previous research has suggested that when given tasks of increasing complexity, children's performance is linked to the development of their cognitive control and EFs: the more developed their cognitive control (e.g., inhibition), the better their performance on the complex tasks (Halford et al., 1998; Richland et al., 2006; Thibaut et al., 2010; Zelazo & Frye, 1998). In studies of inductive reasoning, the contradictory findings regarding the natural default strategy for young children's induction may be partly explained by the category structure, and therefore the complexity, of the stimuli used. Featural category structures (e.g., stimuli used in Gelman & Markman, 1986) require less cognitive control to complete the task than relational category structures (e.g., stimuli used in Sloutsky et al., 2007). Equally, the contradictory findings could be explained by the degree of perceptual similarity between target and test items: the higher the perceptual similarity of the perceptual distracter choice to the target, the greater the level of inhibition needed to see beyond the obvious. Lower levels of perceptual similarity (e.g., stimuli used in Gelman & Markman, 1986) require less cognitive control than those with higher levels of perceptual similarity (e.g., stimuli used in Sloutsky et al., 2007). Together these studies suggest that children may use different inductive strategies depending on the level of cognitive control required to complete the task.

3.2 The influence of experience and knowledge

The majority of research into enhancing reasoning is carried out through direct training of adults (Bliesnezer, Wills & Baltes, 1981), children (White & Alexander, 1986) or teachers (Alexander et al., 1987) to a specific task. Although this can be effective, it cannot demonstrate whether the knowledge learned highlighted relational membership of items; whether an increased understanding allowed participants to take a new outlook to the tasks, or whether they were simply rote-learning task-specific solutions. More recent research focuses on the effect of educational interaction on reasoning performance, and the ability to transfer newly learned information onto novel situations.

Recent research has investigated whether young children could transfer newly taught biological information from books to real animals (Ganea, Ma & DeLoache, 2010). Children were read stories highlighting the importance of biological camouflage, and were asked before and after the book reading which animal a hungry bird would eat (shown both a camouflaged and non-camouflaged image). They found that children aged 4 were able to transfer biological knowledge learned in picture books onto other picture situations. They could also, although to a lesser degree, transfer this knowledge onto real animals and non-target categories. Although younger children sometimes struggle to see how pictures of items represent reality (DeLoache & Burns, 2004), by age 2, children appear to understand the links between pictures of items, and items themselves (Ganea, Allen, Butler, Carey & DeLoache, 2009; Preissler & Bloom, 2007; Preissler & Carey, 2004). Although it has been shown that when children are provided with biological information through engaging and stimulating media such as well designed television programmes, their learning is increased (Troseth, Saylor & Archer, 2006), it is widely agreed that children learn more through direct experience (DeLoache & Chiong, 2009; Strouse & Troseth, 2008). Design-based learning with high school children has found that interactive, project-based tasks where students must work together to solve a problem, increased the students' knowledge and understanding about certain biological facts and concepts (Ellefson, Brinker, Vernacchio & Schunn, 2008). Importantly, children were found to transfer the newly learned skills onto other problems and tasks in new scenarios. Interactive project-type learning has also been found to improve reasoning that focused on non-obvious biological features (Au et al., 2008). Au et al., taught two groups of children about cold and flu symptoms and avoidance. One group were shown a typical educational programme, and the other group participated in an interactive session about the biological causal mechanisms of cold and flu transmission. They found that those who had participated in the interactive session were able to transfer their understanding of the causal mechanisms onto other topics, thus expanding their reasoning ability. This was not found in those who had been shown a typical educational programme without interaction.

Inductive reasoning is used every day, and is especially important for young children trying to make sense of the world. It seems that interactive experience not only improves ability on one task, but can be transferred to other similar tasks. There is little doubt that direct experience with biological kinds aids knowledge and understanding of biological concepts (Hatano & Inagaki, 1994, 1999, 2002; Tarlowski, 2006), although this exposure is often limited to pets or visits to the zoo (Inagaki, 1990; Rosengren, Gelman, Kalish & McCormick, 1991).

What is still unknown is whether this experience and exposure can boost young children's inductive reasoning in general. The key outstanding question remains: can interactive educational sessions provide sufficient accumulation of knowledge to significantly improve category induction, or will a significant improvement only occur with a qualitative change in children's thinking due to cognitive maturation?

4 Introduction to the experimental work

In the introduction so far I have discussed the two broad accounts of inductive reasoning, the various methodological ambiguities stemming from these accounts, and two potential driving forces behind category induction development: the development of EF, and an accumulation of domain knowledge. In this final section of my introduction I shall outline my experimental chapters, and the motivation for this work.

The experimental work is arranged into four chapters, with the findings from each study motivating the design of the next study. All the studies draw from the outstanding questions and issues arising from the literature, which have been discussed in the earlier part of this introduction. Chapters 2 and 3 explore the possible reasons behind such conflicting findings on inductive reasoning. There are various inconsistencies in the methodology and stimuli used in studies supporting each of the accounts, which may have affected the developmental trajectory of induction. Chapter 2 uses a wide age range to systematically investigate the developmental trajectory of induction. Within this chapter, Study 1 explores whether the degree of similarity of the perceptual distracter and target alters children's strategy preference. Study 2 examines whether children who show an apparent perceptual-bias are in fact basing their choice on category knowledge, just a different type of knowledge than I intended (the maturity of the item depicted). In Chapter 3, Study 3, I investigate other crucial factors affecting the developmental trajectory, such as category structure (featural vs. relational category structure) and domain (natural kind vs. artefact). In Study 4, I introduce the consideration that labelling versus no labelling during induction may change children's induction strategy. Having investigated potential reasons for previous researchers’ contradictory findings, and factors affecting the developmental trajectory of induction, I continue into Chapters 4 and 5 by exploring the driving mechanism behind the development of induction. In Chapter 4, Study 5, I test children on a battery of EF tasks as well as my induction task, to investigate whether EF development could be the driving force behind the inductive strategy preference. Finally in Chapter 5, Study 6, I examine whether experience and an accumulation of domain knowledge could be the driving force behind induction. Here, children are divided into either a comparison condition or an intervention condition and are tested on four induction tasks across two sessions. In between the two sessions, the intervention condition also takes part in an interactive educational session at Twycross Zoo about animal adaptations to habitats. Together, through these studies I investigate potential causes behind the current contradictory findings from researchers in support of the two broad accounts, and move to exploring potential driving forces behind the development of inductive reasoning.

In summary, I have drawn inspiration from the two broad accounts of inductive reasoning. I have highlighted key concerns and potential weaknesses with methodologies, which could be influencing the contradictory findings and affecting the developmental trajectory of induction. In Chapters 2 and 3, I investigate methodological inconsistencies in work conducted by researchers in support of the two accounts. In Chapters 4 and 5, I explore potential mechanisms behind the development of inductive reasoning. In Chapter 6, I will explain how my findings have extended our knowledge about the developmental process and mechanism of inductive reasoning, and whether this crucial cognitive skill can be enhanced.

From the research I have reviewed above, I believe the question of interest is not ‘do children use category-induction or perceptual-induction’ but, ‘under which circumstances do children show which preference, and how does this preference develop’.

Chapter 2: Factors influencing the perceptual to category induction transition

In the introduction I discussed the two broad theories of inductive reasoning: the perceptual-bias account (Sloutsky et al., 2007) and the category-bias account (Gelman & Markman, 1986; Opfer & Bulloch, 2007). Each account can accommodate evidence that children's responses in reasoning tasks change between ages 2 and 9, and overtime children focus more on category and relational information, and less on salient perceptual cues. However, these accounts make different predictions about the natural default and development of inductive reasoning. Researchers in support of the perceptual-bias account suggest that young children have a natural default to focus on obvious perceptual cues when making inductive decisions (Sloutsky et al., 2007) or surface characteristics during analogical reasoning (Gentner, 1988) and only later begin to shift their attention to the importance of category membership through learning and experience. Other researchers in support of the perceptual-bias account would suggest that although perceptual induction is the natural default, the transition to using category induction is due to a qualitative change in children's thinking due to a stage-like shift in cognitive maturity, and not simply an accumulation of knowledge (Inhelder & Piaget, 1958; Piaget, 1964). Whereas, researchers in support of the category-bias account suggest that young children have a natural default to understand the importance of category membership, sometimes called 'essentialism' (see Gelman, 2003 for a review). They claim that children who appear to focus on perceptual cues, do not in fact understand the relational connection between same-category items (Goswami, 2001); once children understand that two items are from the same category, they will automatically know to use category induction. There is continual debate between these two accounts as to the natural default of induction.

In Study 1, I aim to examine developmental changes in children's induction using a task designed to elicit different predictions from the two accounts. Importantly, and unlike previous research, I test children from multiple year groups (nursery: 3-4 years, through to Year four: 8-9 years), which enables me to examine the developmental trajectory of children's induction strategies, and provides a stronger test of category-bias and perceptual-bias accounts. I also considered the effect that degree of perceptual similarity between items has on strategy preference. In Study 2, I investigate whether children are using stimuli maturity as an induction strategy rather than perceptual cues or category membership.

1 Study 1: The development of induction and the influence of similarity

As in Sloutsky et al., (2007), I created novel examples of biological stimuli in order to examine children's early induction biases when the salient perceptual cues are not informative of category membership. Children's understanding of the category rule was confirmed in a stringent test of their categorisation performance before and after the induction task. Perceptual and category induction choices were disambiguated by using the triad paradigm, in which children were asked to generalise properties from a target item to either a perceptual or a category choice (as in Bulloch & Opfer, 2009; Gelman & Markman, 1986; Opfer & Bulloch, 2007; Sloutsky et al., 2007). In order to address concerns about the validity of Sloutsky et al's design (Gelman & Waxman, 2007), I used biologically plausible stimuli (novel insects) and demonstrated a familiar relational connection through an animated transition from juvenile to adult. Previous research demonstrates that young children understand the ageing process (see Inagaki & Hatano, 1996 for a review), therefore growth from infancy to adulthood provides a simple way to demonstrate a relational match, whilst providing a biologically plausible context for changes in perceptual features of the items. Thus, researchers in support of the category-bias account would predict that as long as children clearly understand this relationship, then even the youngest children should generalise properties from an adult insect to the same item depicted as a juvenile, rather than to a different-category adult. In order to demonstrate that the relational connection is a reliable predictor, the transition from juvenile to adult will be shown prior to every induction trial.

I examine whether improvements in cognitive control (specifically inhibition) drive a potential perceptual-category transition, by varying the level of featural similarity between the target and perceptual distracters. If a pattern of decreasing 'distractibility by perceptual similarity' is observed along with a parallel transition from perceptual to category induction, then this would provide an explanation of the transition consistent with the category-bias account. Specifically, young children understand the importance of category membership, yet fail to demonstrate their knowledge due to underdeveloped cognitive control. This outcome would be manifested in an interaction between distracter-similarity and age: the level of distracter similarity should affect induction decisions for younger children more than older children. In contrast, if the transition is independent of featural distraction, then this would support research for the perceptual-bias account, suggesting that children's initial failure to make category-based decisions reflects a lack of understanding about the importance of category membership rather than an inability to ignore perceptual distracters. Thus, children only start to use category membership to inform their induction decisions once they have learned to shift their focus away from perceptual features, and towards relational structure (Gentner, 1988; Gentner & Rattermann, 1991; Sloutsky et al., 2007).

1.1 Method

1.1.1 Pre-study participants

Adults:

Twenty-seven adults (mean age = 24.4 years, range 18-54; 9 males and 18 females) participated in a stimuli similarity pre-test, 23 adults (mean age = 26.0 years, range 18-58; 7 males and 16 females) participated in a domain identification and stimulus naming pre-test, and 37 adults (mean age = 22.2 years, range 18-59; 3 males and 32 females) participated in the pre-test of the induction task.

Children:

Forty-nine primary school children (10 Reception, 27 Year 1 and 12 Year 2; range 4.11-7.10 years; 26 males and 23 females) participated in a stimuli similarity pre-test and 32 primary school children (12 Reception, 10 Year 1 and 10 Year 2; range 4.02-7.00; 16 males and 16 females) participated in a domain identification and stimulus naming pre-test.

1.1.2 Main study participants

Two hundred and sixty-four primary school children participated: 57 Nursery (3-4 years), 65 Reception (4-5 years), 43 Year 1 (5-6 years), 38 Year 2 (6-7 years), 31 Year 3 (7-8 years), and 30 Year 4 (8-9 years); range 3.01-9.03 years; 133 males and 131 females, each of whom was tested individually.

Of these, 146 were tested during my Masters (Reception, Year 1 and Year 2) and 118 were tested during my PhD (Nursery, Year 3 and Year 4).

1.1.3 Stimuli

Computer aided designs (CAD, designed using Microsoft Word and Paint) of 256 novel insect stimuli were used: 58 juveniles, 116 transitional images used in the juvenile to adult animations (see Figure 1) and 82 adults (of which 12 were High Similarity Distracters (HSDs) differing from the relevant adult target by head shape, see Figure 2; 12 were Low Similarity Distracters (LSDs) differing from the relevant target on all dimensions apart from the overall size and shade, see Figure 3). Every stimulus was unique and differed on at least one dimension (head shape: round or angled; eyes: orange or white; body: round or triangular; colour: purple or green; colour of markings: black or white; number of markings: two or four; overall size: small juvenile or large adult; overall shade: pale juvenile or dark adult; see Appendix A for examples). In the transition from juvenile to adult, stimuli changed in size, overall shading and grew legs (see Figure 1).

Figure 1. The juvenile to adult transformation in Study 1. In the category learning and the initial and final categorisation tasks, participants completed randomised trials showing juvenile to adult transformations and were asked whether each juvenile and adult was a Sandbug or a Rockbug.

1.1.4 Pre-tests

Stimuli similarity:

During the induction task in the main experiment, children saw 24 triads consisting of an adult target, a same-category but perceptually dissimilar juvenile, and a different-category but perceptually similar adult (see Figures 2 and 3). To make certain that the target item was perceptually more similar to the perceptual distracter test item than the same-category test item, the similarity between test and target items was validated using adult and child ratings.

Figure 2. An example of an induction triad with a High Similarity Distracter in Study 1: the juvenile is transformed into an adult target (transformation), then the adult target is shown with a category choice and a perceptual choice (induction triad).

Figure 3. An example of a category induction triad with a Low Similarity Distracter in Study 1: the juvenile is transformed into an adult target (transformation), then the adult target is shown with a category choice and a perceptual choice (induction triad).

Adults Forty-eight pairings were shown: every triad created two slides, one with the target and the perceptual choice and the other with the target and the category choice (either HSD or LSD). Three pair matches were created: 1) target and HSD, 2) target and LSD, 3) target and category choice. Adults rated the pairs for perceptual similarity on a 5-point Likert scale. A repeated measures GLM showed a significant difference between the three pair types: F (2, 52) = 321.78; $p < .001$ (target-HSD M = 4.56, SD = .43; target-LSD M = 2.02, SD = .18; target-category choice M = 1.95, SD = .41). Post-hoc t-tests confirmed that the target-HSD pairs were rated significantly more similar than either the target-LSD pairs: t (26) = -23.29; $p < .001$, or the target-category choice pairs: t (26) = -21.19; $p < .001$. Bonferroni correction was set at $p < .025$ as the HSD data was used twice. Although statistically the target-LSD pairs were not rated as more perceptually similar than the target-category choice – t (26) = -.58; p = .564 – the means suggest that overall, they were seen as more perceptually similar.

Children The concept of a Likert scale is too complex for young children therefore a simpler technique was used. On the 24 triads, children had to choose which test item looked most like the target item. One-sample t-tests confirmed that the perceptual choice was chosen significantly more often than chance. In triads with HSDs, t (48) = -205.04; $p < .001$; chosen 98% of the time. In triads with LSDs, t (48) = -74.70; $p < .001$; chosen 71% of the time. Thus, children were significantly more likely to pair together items designed to be more perceptually similar to the target (either LSD or HSD, rather than the category choice), therefore validating both as perceptual distracters. To confirm that this effect was stable across the different age groups, a mixed GLM was conducted to investigate the effects of year (Reception, Year 1 and Year 2: between-subjects) and similarity (HSD and LSD: within-subjects) on the percentage of perceptual choices made. As expected, HSDs were chosen more often than LSDs: *F* (1,46) = 30.52; *p* < .001. However, the percentage of perceptual choices was stable across year groups: *F* (2,46) = 10.40; *p* = .607 (see Table 1), and there was no interaction between year and similarity: *F* (2,46) = .384; *p* = .684. I am therefore confident that children of all ages perceived the distracter as more similar to the target than the category choice, and perceived the HSDs as more similar to the targets than the LSDs.

Table 1. Study 1 mean scores for the percentage of perceptual choices during LSD and HSD trials, for each year group

Year Group	*Mean % LSD choices (SD)*	*Mean % HSD choices (SD)*
Reception	69% (4.79)	91% (2.81)
Year 1	74% (3.81)	99% (.27)
Year 2	67% (3.49)	99% (.29)

Domain identification and stimulus naming:

Adults A domain categorisation pre-test was conducted to check the stimuli were considered to be biologically plausible. Adults were shown four images individually (2 juvenile bugs and 2 adult bugs) and asked to state whether the item was a living or a non-living thing, and what they believed it to be. Adult Sandbugs and Rockbugs were labelled as living kinds 100% of the time, with common responses being ‘beetle’ and ‘bug’; juvenile Sandbugs were labeled as living 87% of the time, and juvenile Rockbugs were labelled as living 91% of the time, with common responses being ‘worm’ and ‘maggot’.

Children The same four images were shown. Adult Sandbugs and Rockbugs were labelled as living kinds 100% of the time, with common responses being ‘spider’ and ‘ladybird’; juvenile Sandbugs were labelled as living 78% of the time, and juvenile Rockbugs were labelled as living 62% of the time, with common responses being ‘fish’ and ‘worm’.

Induction task:

Adults completed the induction task designed for children, to ensure it was easy to understand and complete before it was given to children. Adults scored highly and above chance on the initial and final categorisation tasks: Initial *M* = 99% (*SD* = 3.12), *t* (34) = 93.38; *p* < .001; Final *M* = 99% (*SD* = 1.98), *t* (34) = 148.09; *p* < .001 and made significantly more category choices than expected by chance: *M* = 86% (*SD* = 6.35), *t* (34) = 8.24; *p* < .001.

1.2 Design and Procedure

There was one between-subjects factor: year (with 6 levels: Years N, R, 1, 2, 3, 4) and one within-subjects factor: degree of distracter similarity (with 2 levels: LSD and HSD). The dependent variable was the number of correctly categorised items during the initial and final categorisation tasks, and the number of perceptual or category choices made during the induction tasks.

Each participant completed the task individually, in a single session. The procedure was four-fold (as in Sloutsky et al., 2007): 1) category learning, 2) initial categorisation, 3) induction task, 4) final categorisation. None of the stimuli were repeated across the different task sections, and the order of stimuli presentation in each section was randomised.

Firstly, participants were told that they would see some bugs growing up. They were presented with a Rockbug and a Sandbug animation (from juvenile to adult; see Figure 1), without markings, and were told the category rule for each adult and juvenile:

> *“Here is a Sandbug. Sandbugs live in the sand and have round heads for soft burrows. Let’s watch it grow up. Now it has grown up, it is still a Sandbug, it lives in the sand and has a round head for soft burrows. Here is a Rockbug. Rockbugs live in rocks and have sharp pointy heads for digging. Let’s watch it grow up. Now it has grown up, it is still a Rockbug, it lives in rocks and has a sharp pointy head for digging”.*

To ensure participants did not see the critical feature as arbitrary, a function was provided, for example, “pointy heads for digging”. Participants were told that these bugs came in different colours, shapes and had different markings, and we could only tell the difference by the head shape. This explanation was provided to help children understand that the other differing dimensions were irrelevant. Participants then completed the category learning phase whereby they were shown eight random trials of Sandbugs and Rockbugs and asked to identify the bug at the juvenile stage, and then again at the adult stage once the transformation was complete. Asking for the name at both stages provided a check that children understood the continuity between juvenile and adult. Feedback was provided, and the participants were reminded of the rule after each juvenile and adult answer, for example, “Yes well done, it is a Rockbug because it lives in rocks and has a sharp pointy head for digging”, to re-iterate the functional importance.

In the initial categorisation task, children were shown six Sandbug and six Rockbug animations in random order and were asked whether each juvenile and adult was a Sandbug or Rockbug. No feedback was given and the rule was not reiterated.

The induction task consisted of 24 randomised trials (12 Sandbug targets – of which 6 were HSDs and 6 were LSDs – and 12 Rockbug targets – of which 6 were HSDs and 6 were LSDs), each showing a juvenile to adult animation followed by a triad (see Figures 2 and 3). Each adult target was given a hidden property (relating to its insides e.g., ‘cold blood’, or its behaviour e.g., ‘eats flies’, based on Gelman & Markman, 1986. See Appendix B for full list), and the child was instructed to point to the test item which also had this property. If the child chose the distracter (HSD or LSD), this was coded as a perceptual choice; choosing the juvenile was coded as a category choice. No labels were used to avoid potentially priming the child to select the category choice, simply because it had the same label. Instead, items were referred to as ‘this one’.

The final categorisation task followed the same procedure as the initial categorisation task, but with new stimuli. This task was included to ascertain whether children could remember the category rules for differentiating between Sandbugs and Rockbugs at the end of the study.

1.3 Results

1.3.1 Categorisation Performance

Only children who performed significantly above chance in both initial and final categorisation tasks were included in the final sample. According to a binomial test, scores of 10/12 (83%) and above were significantly above chance (proportion = 0.5, *p* = .04). Forty-three children (34

Nursery, 7 Reception, and 2 Year 1 children) scored at or below chance and were removed. The remaining 221 children scored highly (see Table 2) and above chance on the initial and final categorisation tasks: Initial *M* = 99% (*SD* = 3.20), *t* (220) = 226.20; *p* < .001; Final *M* = 98% (*SD* = 3.80), *t* (220) = 190.67; *p* < .001.

Table 2. Mean scores for the initial and final categorisation tasks for each year group in Study 1

Year Group	*Mean initial score (SD)*	*Mean final score (SD)*
Nursery	96% (4.00)	97% (3.10)
Reception	98% (4.00)	97% (4.53)
Year 1	99% (2.65)	98% (4.55)
Year 2	99% (3.02)	98% (4.02)
Year 3	99% (1.44)	99% (1.00)
Year 4	100% (0.00)	99% (0.73)

1.3.2 Induction Performance

The percentage of category choices made by each participant in the induction task was examined; with participants who made at least 18/24 category choices considered to have a category-bias (binomial test proportion 0.5, *p* = .02). Participants who made less than 6/24 category choices were considered to have a perceptual-bias (binomial test proportion 0.5, *p* = .02). As shown in Figure 4, the percentage of children with a category-bias increased with age. From Year 2 onwards, the pattern is stable, with a clear majority of children showing a category-bias in each year group.

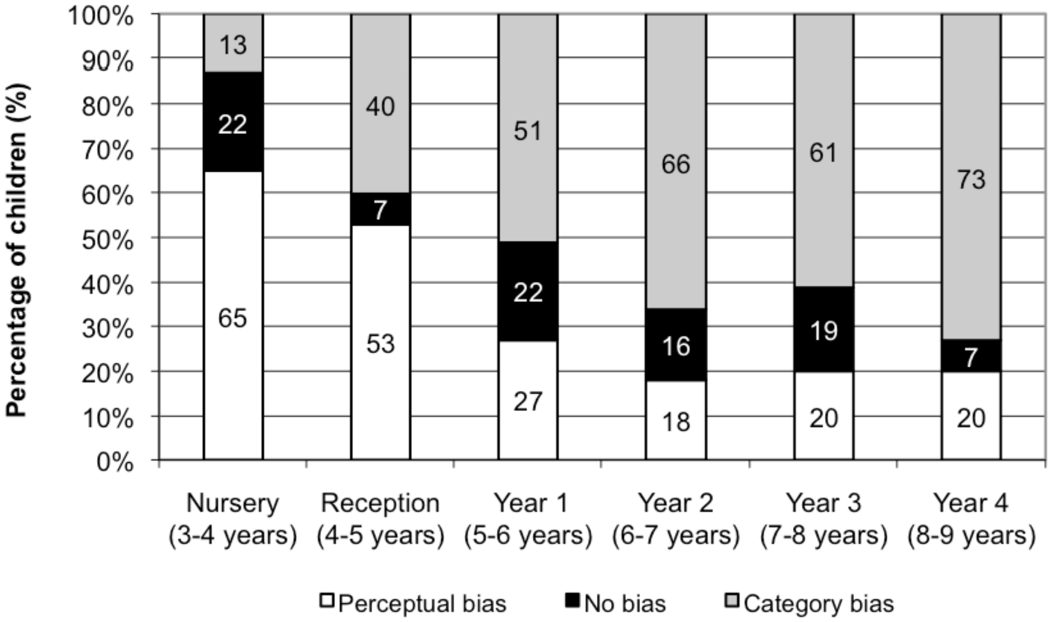


Figure 4. The percentage of children showing each type of bias across six year groups in Study 1.

I used one-sample t-tests to investigate the overall percentage of category choices made by each age group (see Table 3). Nursery children made fewer category choices than expected by chance, indicating a preference for perceptual choices (*t* (22) = -4.02 *p* < .001). Reception children performed at chance, indicating no overall preference for either category or perceptual choices (*t* (57) = -1.14; *p* = .26). Children in Year 1 showed a non-significant trend towards making more category choices (*t* (40) = -1.88; *p* = .07). All groups of older children made significantly more category choices than expected by chance: Year 2, *t* (37) = 3.58; *p* < .001, Year 3, *t* (30) = 3.03; *p* < .05, Year 4, *t* (29) = 3.62; *p* < .001 (see Table 3).

Table 3. The percentage of category induction choices for each year group in Study 1.

Year Group	*Mean initial score (SD)*
Nursery	26% (28.10)
Reception	43% (43.60)
Year 1	61% (38.14)
Year 2	72% (37.57)
Year 3	71% (38.07)
Year 4	75% (37.64)

Twelve hidden properties were each used twice throughout the 24-trial induction task. A Chronbach's Alpha test confirmed that children's responses were consistent across the different properties (α = .98). Most children did not justify their responses, but when verbal comments were made during the induction task, these were noted. Children's responses generally supported the choice that they made by commenting on either the appearance, or the category membership of their chosen test item. For example, a comment from a perceptual-preference Reception child: "That one (*points to distracter*) looks like that one (*points to target*), so even though they don't have the same head, they're the same"; comment from a category-preference Year 1 child: "They look like the same kind (*points to target and distracter*) but that has the same head (*pointing to category choice*) so it must be that one"; comment from a category-preference Year 2 child: "I know [the target and category choice have the same property] as they have the same head".

1.3.3 Effects of Similarity

The mean percentage of category choices were examined as a function of the level of similarity of perceptual distracter (HSD vs. LSD) across the six year groups. A mixed GLM (similarity, year) showed a main effect of similarity (HSD vs. LSD): *F* (1, 215) = 26.75; *p* < .001. As shown in Figure 5, the percentage of category choices for HSD trials was lower than for LSD trials. As expected, the overall percentage of category choices increased for older year groups: *F* (5, 215) = 7.22; *p* = .001. However, there was no interaction between similarity and year: *F* (5, 215) = 148.48; *p* = .404, indicating a stable effect of featural similarity for all year groups.

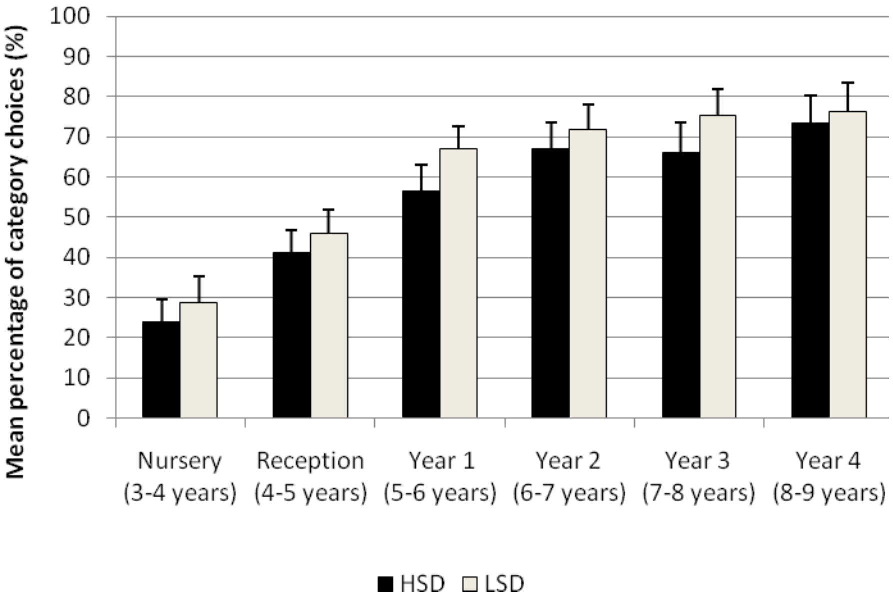


Figure 5. The mean percentage of category choices made when distracter was either a HSD or a LSD for each age group in Study 1.

1.4 Discussion

These findings support the notion that young children's inductive reasoning is perceptually-based (Sloutsky et al., 2007) and over time there is a gradual transition towards category-based induction. These findings also show that older children are equally distracted by higher levels of similarity as younger children, suggesting that the transition towards a category preference is unlikely to be caused by an increased ability to inhibit highly similar distracters. However, it could be argued that the apparent transition is not due to the development of more sophisticated reasoning, and is instead due to changes in the types of category information children consider to be important. Specifically, it is possible that younger children were making choices based on maturation categories (adult vs. juvenile). Taylor and Gelman (1993) created a two-condition study (category-based induction vs. similarity judgments) whereby children in both conditions saw triads of people varying in age and gender (e.g., target: baby girl; test items: baby boy and an adult woman). Children in the similarity condition were asked which test item was "more like" the target; children in the induction condition first learned a new property about the target item and then asked which of the two test items had the same property. They found that children in the similarity condition were more likely to use gender (68%), whereas children in the induction condition only used gender in 20% of cases.

Taylor and Gelman suggest that when children are provided with an inductive situation, they are able to distinguish between what is salient (gender) and what would be a good indication of inference (age). This could explain why younger children tended to generalise properties from the target adult to the distracter choice (which matched the target in maturity), whereas older children tended to generalise properties to the juvenile, which matched the target in terms of taxonomy (bug category). The issue was addressed in Study 2.

2 Study 2: Are children's inductive responses influenced by maturity groupings?

The aim of Study 2 was to examine whether children who showed an apparent perceptual-bias were in fact basing their choice on maturation (pairing the adult target with another adult). In Study 1, an animation from juvenile to adult was used to emphasise the relationship between the two items: the juvenile and adult were from the same category (in fact, they were exactly the same item – the adult was the "grown up" version of the juvenile). I confirmed that children understood this relationship by asking them to label the juvenile and adult items in the initial and final categorisation tasks. I was therefore confident that if children understood the importance of category information, then they would use this information in the induction task. However, this design created a further grouping that could be used to inform category-decisions, specifically, the maturity of the target and test items. Study 2 followed the same format as Study 1, except that two conditions were compared with the category-choice varying in maturity (juvenile or adult). The 'juvenile as category choice' trials (hereafter referred to as juvenile trials) were the same as the HSD condition from Study 1: 12 induction triads with an adult target, an adult HSD and a juvenile category choice. In the 'adult as category choice' trials (hereafter referred to as adult trials) there were 12 induction triads with an adult target, an adult HSD and an adult category choice which differed from the target on shape, colour, markings and marking colour, but had the same head shape. If younger children used maturation to inform their induction decisions, then they would make more maturity-based choices in the juvenile trials (when the target and the distracter were adults and the category-choice was a juvenile) than in the adult trials (when all three items were adults). This is because participants could use maturity information in the former condition, but not in the latter condition. If younger children used perceptual information to inform their induction decisions, then they would choose the distracter item in both conditions, since this item was always the most similar to the target. Since Study 1 demonstrated that most children transition from a perceptual to a category bias sometime between Reception and Year 2, I focused on these three age groups.

2.1 Method

2.1.1 Pre-study participants

Adults:

Twenty-seven adults (mean age = 24.4 years, range 18-54; 9 males and 18 females) participated in a stimuli similarity pre-test and 15 adults (mean age = 21.4 years, range 18-28; 1 male and 14 females) participated in the pre-test of the induction task.

Children:

Forty-nine primary school children participated in a stimuli similarity pre-test (10 Reception, 27 Year 1 and 12 Year 2; range 4.11-7.10 years; 26 males and 23 females).

2.1.2 Main study participants

Fifty-two primary school children participated: 17 Reception (4-5 years), 16 Year 1 (5-6 years) and 19 Year 2 (6-7 years); range 4.04-7.3; 21 males and 31 females, each of whom was tested individually.

2.1.3 Stimuli

Computer aided designs (CAD) of 268 stimuli were predominantly taken from Study 1, the extra were stimuli not previously used during Study 1. All the stimuli were the same except in the induction task. The twelve HSD trials were used as before, but the twelve LSD trials were replaced with 12 adult trials: HSDs replaced LSDs and adult category choices replaced juvenile category choices: 58 juveniles, 94 adults and 116 transitional images used in the juvenile to adult animations.

2.1.4 Pre-tests

Stimuli similarity:

To make certain that the target item was perceptually more similar to the perceptual distracter test item than the same-category test item, the similarity between test and target items was validated using adult and child ratings.

Adults Forty-eight pairings were shown: every triad created two slides, one with the target and the perceptual choice and the other with the target and the category choice (either juvenile or adult). Four pair matches were created: 1) target and adult category choice, 2) target and adult perceptual distracter (adult trials), 3) target and juvenile category choice, 4) target and adult perceptual distracter (juvenile trials). Adults rated the pairs for perceptual similarity on a 5-point Likert scale. Paired t-tests showed that in the adult trials, the adult perceptual distracter was rated as higher in perceptual similarity to the target (*M* = 4.56; *SD* = .43) than the adult category choice (*M* = 2.02; *SD* = .43): *t* (26) = -23.29; *p* < .001. In the juvenile trials, the adult perceptual distracter was also rated as higher in perceptual similarity to the target (*M* = 4.56; *SD* = .42) than the juvenile category choice (*M* = 1.95; *SD* = .41): *t* (26) = -21.46; *p* < .001. Thus confirming that the perceptual distracters in both the adult and juvenile trials were seen as more perceptually similar to the target than the category choices.

Children On the 24 triads, children had to choose which test item looked most like the target item. Twelve of the triads were adult trials and 12 of the triads were juvenile trials. One-way t-tests confirmed that in the adult trials, the adult perceptual distracter was chosen significantly more than the adult category choice: *t* (48) = -472.70; *p* < .001; chosen 99% of the time. In the juvenile trials, the adult perceptual distracter was chosen significantly more than the juvenile category choice: *t* (48) = -205.05; *p* < .001; chosen 98% of the time. Children were more likely to pair together items designed to be more perceptually similar to the target in the adult trials and the juvenile trials, therefore validating both adult distracters as perceptual distracters. There was no significant difference between choosing the adult perceptual distracter when the category choice was an adult or a juvenile: *t* (48) = -.73; *p* = .471. This validates both as perceptual distracters.

Induction task:

Adults completed the induction task designed for children, to ensure it was easy to understand and complete, before it was given to children. Adults scored highly and above chance on the initial and final categorisation tasks: Initial *M* = 100% (*SD* = 0.00); Final *M* = 100% (*SD* = 0.00), and made significantly more category choices than expected by chance: juvenile trials, *t* (14) = 4.00; *p* < .001; adult trials, *t* (14) = 2.81; *p* = .014. A paired samples t-test was conducted; there was no significant difference in inductive strategy between the trials with a juvenile category choice and the trials with an adult category choice: *t* (14) = 8.71; *p* = .398.

2.2 Design and Procedure

There was one between-subjects factor: year (with 3 levels: Years R, 1, 2) and one within-subjects factor: type of category choice (with 2 levels: juvenile and adult). The dependent variable was number of correctly categorised items during the initial and final categorisation tasks, and the number of perceptual or category choices made during the induction tasks. Each participant completed the task individually, in a single session. The procedure replicated Study 1, but with two counterbalanced blocks of trials in the induction task. One of the blocks showed the 12 HSD triads from Study 1, and the other showed 12 new triads including an adult target, an adult HSD and an adult category choice. The four-fold methodology and hidden properties replicated Study 1.

2.3 Results

2.3.1 Categorisation Performance

Only children who performed significantly above chance in both the initial and final categorisation tasks were included in the final sample. According to a binomial test, scores of 10/12 (83%) and above were significantly above chance (proportion = 0.5, *p* = .04). Four children (3 Reception and 1 Year 2) scored at or below chance in either the initial or final categorisation tasks, and were removed from the following analyses. The remaining 48 children scored highly and above chance on the initial and final categorisation tasks: Initial *M* = 99% (*SD* = 2.67), *t* (47) = 126.97; *p* < .001; Final *M* = 98% (*SD* = 3.65), *t* (47) = 92.03; *p* < .001.

2.3.2 Induction Performance

The proportion of category choices made by each participant in the induction task was examined, with participants who made at least 18/24 category choices considered to have a category-bias (binomial test proportion 0.5, *p* = .02). Participants who made less than 6/24 category choices were considered to have a perceptual-bias (binomial test proportion 0.5, *p* = .02).

As in Study 1, a gradual transition was found from perceptual preference to category preference. Figure 6 shows the percentage of children in each year group who made significantly more category choices, perceptual choices, or showed no bias. I used one-sample t-tests to investigate the overall percentage of category choices made by each age group. Reception children made fewer category choices than expected by chance, indicating an overall preference for perceptual choice: *t* (13) = -2.13; *p* = .05. As shown in Figure 6, only 14% of Reception children showed a category bias. This differs from my finding in Study 1, where 40% showed a category bias, and is perhaps because Study 2 was conducted earlier in the school year, so the children were slightly younger than those in the Study 1 sample. Year 1 children performed at chance, indicating no overall preference for either category or perceptual choices: *t* (15) = .21; *p* = .836. The majority of Year 2 children (64%), showed a bias to make the category choice but this bias was not significant for the group as a whole: *t* (17) = 1.55; *p* = .141. Again, this is likely to be because the children were slightly younger in this sample.

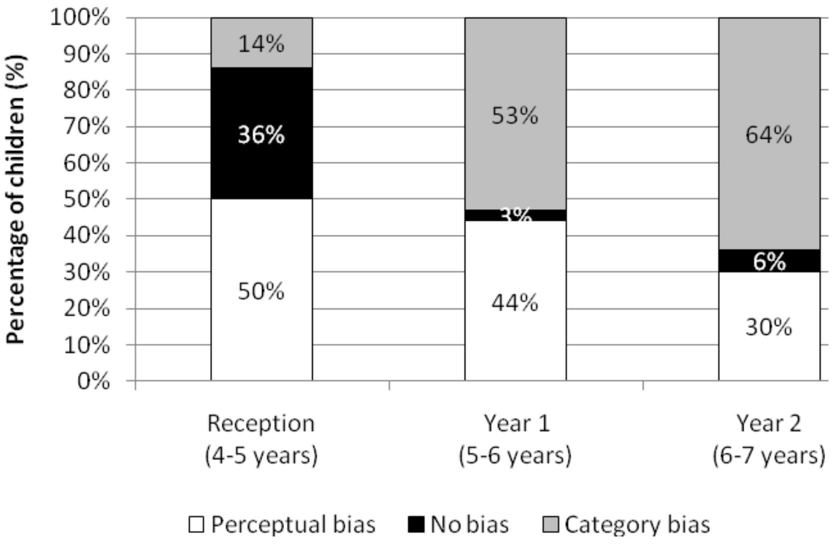


Figure 6. The percentage of children showing each type of bias across the three year groups in Study 2.

2.3.3 Effects of Maturity

A mixed GLM (maturity, year) was conducted and showed no effect of maturity (juvenile trials vs. adult trials): *F* (1, 45) = .071; *p* = .791. As shown in Figure 7, the percentage of category choices for juvenile trials and adult trials was comparable, indicating that children were not making their choices through maturation groupings, and were instead either relying on category or perceptual cues. There was a non-significant trend towards an effect of year (Reception, Year 1, Year 2) with the number of category choices increasing for older year groups: *F* (2, 45) = 2.70; *p* = .078. There was no interaction between maturity and year: *F* (2, 45) = 1.18; *p* = .316, indicating that all year groups were equally unaffected by the maturity of the category choice.

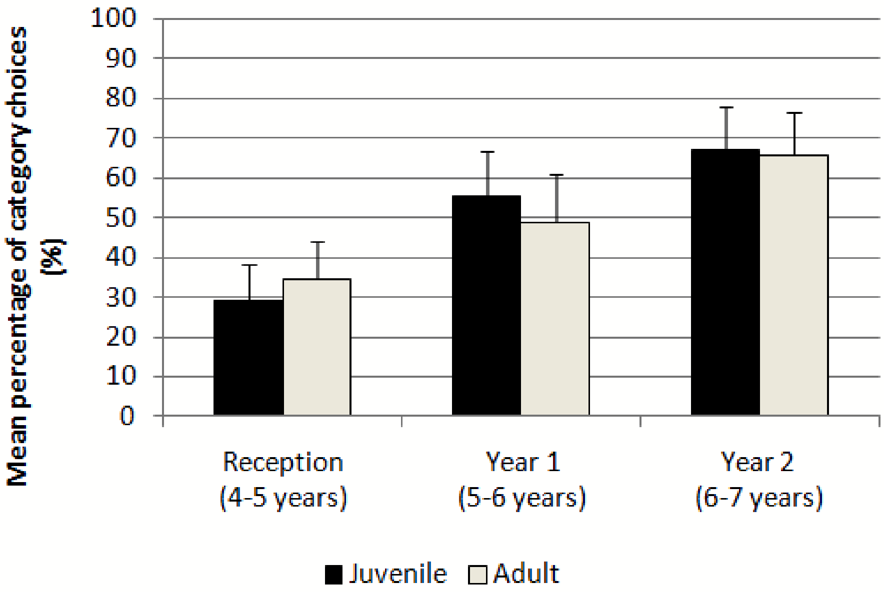


Figure 7. The mean percentage of category choices made when the category choice was either a juvenile or an adult in Study 2.

2.4 Discussion

As in Study 1, I found a gradual transition from the majority of children showing a bias towards perceptual information through to the majority of children showing a bias towards category information. The same pattern was shown regardless of the maturity of the category choice. These findings suggest that there are no clear developmental changes in the types of category information children consider to be important. Therefore, it is unlikely that the younger children in Study 1 were making choices based on maturation categories. Instead, it is much more likely that the transition reflected a genuine change in focus from perceptual to category induction.

3 General discussion of Studies 1 and 2

I investigated two broad accounts of the development of children's inductive strategies. Specifically, I examined whether young children's default preference is for perceptual or category induction, and by testing multiple year groups I investigated whether a developmental trajectory was observed. As in Sloutsky et al., (2007), I used novel examples in order to disambiguate category-based and perceptual-based induction choices. However, I also addressed the concerns of Gelman and Waxman (2007) by using examples of plausible biological kinds, and demonstrating the relational connection between same-category items using a highly familiar relationship (growth from juvenile to adult). Importantly, I also addressed whether the perceptual to category transition was driven by a decrease in featural 'distractability'. Specifically, whether the transition was due to children becoming better at inhibiting high similarity distracters through improved cognitive control (Richland et al., 2006).

This work provides support for perceptual-bias and relational shift accounts (Gentner, 1988; Gentner & Toupin, 1986; Sloutsky et al., 2007), as I observed a gradual developmental transition in inductive reasoning strategy away from a perceptual-bias, towards a category-bias in 3 to 9 year old children. As expected, children were more distracted by higher perceptual distracters (Osherson et al., 1990; Sloman, 1993). However, an interesting and novel observation is that children at all ages were equally influenced by the level of similarity of the distracter. Children of all ages made fewer category choices when the distracter was a HSD compared to when the distracter was a LSD, regardless of their overall induction strategy preference. Thus, this gradual transition cannot be explained by an increased ability to inhibit or ignore salient, but irrelevant, perceptual cues. Instead, the perceptual to category transition occurs independently of children's ability to ignore highly similar perceptual distracters.

3.1 Young children's default induction strategy

These findings make an essentialist bias towards non-obvious category information in early induction unlikely. Gelman (2003) argues that children have an early bias to attend to hidden, non-obvious properties and thus, make generalisations based on kind-information over appearance. Similarly, Goswami (1992; 2001) argues that young children's natural default is for relational interpretations. Once children understand the specific relations relevant to the task, then their decision will be based on these. Bulloch and Opfer (2009) also argue that young children are capable of using relational information when they recognise this as a reliable predictor. In Study 1, the target had a very strong relationship to the category-choice. It was not only the same type of bug; it was the very same bug, grown up. The connection between the juvenile (category choice) and adult target was demonstrated prior to every trial when the child witnessed the juvenile bug being transformed into the adult target bug. In addition, children's understanding of the two different types of bugs (Sandbugs and Rockbugs), and the continuity between juvenile and adult bugs was confirmed in the categorisation task both before and after the induction task. Instead, my findings provide support for research suggesting that the natural default for induction in young children is perceptual (Inhelder & Piaget, 1958; Piaget 1964; Sloutsky et al., 2007). It is striking to note that the very young children who passed the strict categorisation criteria showed a significant bias not to apply this category rule during the induction task, and instead reverted to induction based on perceptual cues.

Nevertheless, there are three main alternative explanations for my findings in Study 1. Firstly, the perceptual choice was arguably more distinctive than the category choice (larger, with more salient features), thus young children may have been drawn to this item. Secondly, young children may have been using category information, just not the information I intended them to use. Specifically, it is possible that they chose the perceptual choice because this matched the target in maturity and was therefore more likely to share properties with the target than the juvenile. Finally, children may have been basing their responses on basic-level category information, and thus considered either type of bug to plausibly share properties with the target (Waxman, Lynch, Casey & Baer, 1997). The first two concerns are neatly addressed in Study 2, which confirmed that the same percentage of category choices were made in the adult trials as well as in the juvenile trials. Thus, children's decisions were stable, even when the category choice and distracter were equally salient, and even when the category choice was the same age as the target. Finally, it is unlikely that children's performance can be explained by a focus on basic-level information since children were trained to focus on the differences between the categories, and were proficient in defining items by their specific name by the time they reached the induction task. In addition, if children believed that either choice were a plausible basis for induction, then they should have responded at random, choosing either the juvenile or adult distracter because both were bugs and therefore either could share the properties at the basic level. However, the majority of children showed a bias either for the perceptual or the category choice. Only a minority showed no clear bias at each age group. In contrast, there is a clear transition from the majority of children showing a bias for perceptual choices through to a majority bias for the category choice. Thus, by far the most plausible interpretation of my findings is that young children showed a bias towards the perceptual distracter because they did not appreciate the importance of category information in making induction decisions.

3.2 The influence of perceptual similarity on induction

Previous studies have used varying levels of perceptual similarity between target and test items which could be a key factor behind their contradictory findings. For example, Sloutsky et al., (2007) used an explicit manipulation of appearance similarity such that the perceptual (distracter) choice had the same overall appearance as the target. In contrast, Gelman and Markman (1986) selected real biological kind examples (e.g., squirrel one, squirrel two, rabbit). Although these were selected so that the overall appearance similarity between target and distracter was greater than between target and category choice, the salience of the perceptual distracter will have been much less than in Sloutsky et al’s study, allowing more children to successfully make the category choice. Thus, a comparison of these studies reveals a pattern consistent with Osherson et al., (1990) and Sloman’s (1993) work with adults. Specifically, fewer inductive category choices were made when perceptual similarity was higher (as in Sloutsky et al., 2007), and more inductive category choices were made when perceptual similarity was lower (as in Gelman & Markman, 1986). This crude comparison demonstrates how the choice of stimuli can affect the apparent inductive preference choice, and ideally, researchers need to compare stimuli at different levels of similarity within the same study. In Study 1, I specifically investigated whether different levels of perceptual similarity between the target and distracter item could encourage different inductive preferences. My findings demonstrated a clear, consistent influence of perceptual similarity on induction decisions. Crucially, there was no interaction between age and level of similarity. Having a highly similar distracter present during the induction task caused children of all ages to make fewer category choices. Thus, an increased ability to ignore highly similar distracters is unlikely to drive the transition from a perceptual to a category bias. I do not mean to rule out the influence of inhibition in supporting children's induction decisions. Clearly, the children who made category choices must have successfully inhibited the more salient perceptual choice. However, I cannot interpret the failure of the younger children to make category choices simply in terms of poor inhibition, because the level of similarity of the perceptual distracters equally affected all ages of children. Thus, although inhibition may play a role in supporting children's induction choices, the ability to ignore perceptual distracters is not enough to drive the transition from perceptual to category induction. Below, I discuss more likely interpretations of this transition.

3.3 Developmental changes in induction strategy

Although I am not the first to suggest a qualitative shift in children's thinking during their first few school years (Gentner, 1988; Piaget, 1964; Sloutsky et al., 2007), I have presented the first systematic evidence of a perceptual to category transition in induction strategy. The key question remaining concerns the driving force behind this transition; my data adds to the debate towards deciding between potential mechanisms. Firstly, I have demonstrated that the development of category induction is unlikely to be driven by an early bias to attend to non-obvious categorical information, since there is no evidence of an early focus on category information in induction. In contrast, there is clear evidence of an early bias to use perceptual information, even though young children are highly proficient in categorising the items. Secondly, the development of category induction, although possibly supported by inhibition, is not triggered by an increased ability to inhibit high similarity distracters, since all age groups were equally influenced by the level of similarity between distracter and target. Thirdly, children are not making their inductive decisions based on maturity groupings or attention grabbing features such as size, and are instead focusing on either the appearance or the category of the items presented. Instead, the most likely explanation of this transition is that young children's reasoning is biased toward perceptual features, and only later do they start to incorporate category information into their induction decisions. Thus, the younger children in my sample did not make category choices, because they did not appreciate the importance of category membership when making inductive decisions.

These findings are potentially consistent with either a stage-like cognitive maturation account (Piaget, 1964) or a gradual shift in the development of analogical reasoning (Gentner, 1988) and category induction (Sloutsky et al., 2007) due to an accumulation of knowledge. I argue that the gradual nature of the transition makes the second account more likely. It is inevitable that the older children in my sample had greater knowledge of biological kinds and were therefore more attuned to the relationship between category membership and behaviour. Thus, this heightened understanding of the importance of category relationships in the biological domain caused the older children to become increasingly more attentive to relational structure, biasing them to use category information to make induction decisions (Gentner, 1988; Gentner & Rattermann, 1991; Halford, 1993; Rattermann & Gentner, 1998).

In conclusion, my findings demonstrate that a qualitative change occurs in children's reasoning between the ages of 3 and 9, independent of inhibition capabilities, in which they begin to appreciate the importance of category information in explaining the properties and behaviour of biological kinds. This appreciation enables children to look beyond the obvious when making induction decisions about natural kinds, and is likely to scaffold increasingly sophisticated reasoning more generally.

Chapter 3: The effects of category structure and domain on the developmental trajectory of induction

In Chapter 2, I examined the developmental trajectory of inductive strategy with children aged 3-9 years, and observed a gradual transition from a perceptual-bias to a category-bias. It was also found that stimuli maturity groupings did not affect inductive choice; children used the same preferred strategy when the category choice was an adult and when it was a juvenile. Therefore, the perceptual preference in young children could not be explained by induction through a new category 'maturity'. The gradual transition provides support for the perceptual-bias account, making an early category bias very unlikely. I also investigated whether the level of perceptual similarity of stimuli affects strategy preference, and found that children of all ages were equally distracted by the HSDs compared to the LSDs. Children made more perceptual choices during HSD trials regardless of whether their overall strategy preference was perceptual or category. Interestingly, the gradual transition of strategy preference was independent of distractibility, which suggests that improvements in cognitive control (e.g., inhibition) do not drive a potential perceptual to category transition: it is not children's inability to inhibit perceptual distracters due to underdeveloped cognitive control that causes a transition.

In this chapter I further investigate factors that may affect the development of category induction, and could be linked to the transition from perceptual to category preference. Study 3 investigates the issues of category structure and domain. When looking through previous methodologies, it became apparent that the category structure of the stimuli often differed – either a simple structure (featural; Gelman and colleagues) or a complex structure (relational; Sloutsky and colleagues) – which made direct comparison between studies impossible. I also found that although previous researchers have speculated as to whether inductive reasoning is domain-specific or domain-general, they have not directly compared natural kind and artefact domains. Researchers from neither account have systematically investigated the effects of category structure or domain on inductive strategy preference. My aim was to investigate the impact that category structure and domain have on the developmental trajectory of inductive reasoning. Study 4 checks the effects of labelling on children’s strategy preference during the induction task. One criticism against certain researchers supporting the category-bias account is the use of stimuli labels during their induction trials. They claim that labels are proxies for induction, highlighting category membership which leads to category induction. However, it has been argued that labels contribute to the level of item similarity, resulting in children using label induction rather than category induction (Fisher & Sloutsky, 2005). Although studies have been conducted to prove that labels do not affect induction (Gelman & Markman, 1986), the debate continues.

1 Study 3: The effect of domain and category structure

As mentioned in the introduction, there is evidence that children as young as 3 years old are sensitive to domain when making assumptions about the properties of an object (Backscheider et al., 1993; Carey, 1985; Inagaki & Hatano, 1996; Jipson & Gelman, 2007), for example, children can make subtle distinctions between the categories of natural kind and artefact when generalising behaviours; knowing that ‘insides’ are important for natural kinds only (Gelman, 2003). A plethora of research suggests that it is this essentialism or instinctive understanding about natural kinds, which leads induction to be domain-specific (Gelman, 2003; Goswami, 2001; Keil, 1989; Rattermann & Gentner, 1998; Wellman & Gelman, 1998). However, there has not yet been a direct systematic comparison of children’s induction strategies across the domains of natural kind and artefact. The current study directly compared well-matched natural kind (bugs) and artefact (vehicles) domains in tasks of inductive reasoning.

Category structure is also an important feature of categorisation. As mentioned in the introduction, categories can be deterministic or probabilistic (Kittur et al., 2004). This study focused on two key deterministic structures: featural, whereby items of the same category share a unique feature, and relational, whereby items of the same category share the same relational connection between features. It has been suggested that the intensity of the request put on a child’s cognitive control – executive function (EF) – could affect children’s ability to carry out tasks (Bunge & Zelazo, 2006; Halford, 1993; Halford et al., 1998; Zelazo & Frye, 1998; Zelazo & Müller, 2002). Featural structures are less complex and are therefore less demanding on EF compared to more complex relational structures. Children may find it harder to successfully complete the more complex tasks until their EF is well developed.

As I have outlined above, both category structure and domain are likely to critically affect children's category induction decisions and could help to explain previous contradictory findings. I manipulated category structure (featural vs. relational) and domain (natural kind vs. artefact) orthogonally using artificial categories. However, bearing in mind criticisms of previous studies using artificial categories, I took care to ensure that the items were plausible examples of natural kind and artefact categories. As in previous studies of inductive reasoning (Gelman & Markman, 1986; Opfer & Bulloch, 2007; Sloutsky et al., 2007), I used the triad paradigm in which children are asked to generalise the properties of a target item to one of the two test items. In order to create plausible examples of same-category items that were very different in appearance, I created juvenile natural kinds and incomplete artefacts, and adult natural kinds and complete artefacts. The relationship between these category members was demonstrated through an animated transformation (see Figures 8 and 9). Finally, in order to compare the developmental trajectory of inductive reasoning for each of these conditions, I tested children spanning the key age range used in previous studies (ages 4 to 9 years).

Aston University

Illustration removed for copyright restrictions

Figure 8. Example of a natural kind transformation used in Study 3.

Figure 9. Examples of an artefact transformation used in Study 3.

I had two predictions: firstly, young children would make more category choices when considering the natural kind examples because of their essentialist beliefs about the importance of non-obvious, internal properties (see Gelman, 2003 for an overview); secondly, young children would make more category choices when reasoning about categories with featural structures, since these are less complex and therefore less demanding on EF than relational structures (Halford, 1993; Richland et al., 2006; Thibaut et al., 2010; Zelazo & Müller, 2002).

1.1 Method

1.1.1 Pre-study participants

Adults:

Thirty adults (mean age = 19.7 years, range 18-34; 5 males and 25 females) participated in a stimuli similarity pre-test, 31 adults (mean age = 19.7 years, range 18-34; 5 males and 26 females) participated in a domain identification and stimulus naming pre-test, and 25 adults (mean age = 20.4, range 18-35; 5 males and 20 females) participated in the pre-test of the induction tasks.

Children:

Thirty primary school children (10 Reception, 10 Year 1 and 10 Year 2; range = 4.02-7.01; 15 males and 15 females) participated in a stimuli similarity pre-test and 32 primary school children (12

Reception, 10 Year 1 and 10 Year 2; range 4.02-7.00; 16 males and 16 females) participated in a domain identification and stimulus naming pre-test.

1.1.2 Main study participants

Four hundred and three primary school children participated: 105 Reception (4-5 years), 109 Year 1 (5-6 years); 102 Year 2 (6-7 years); 41 Year 3 (7-8 years), and 46 Year 4 (8-9 years); range 4.10-9.1 years; 214 males and 189 females. Every child participated in either the featural or relational rule condition, and completed both a natural kind and an artefact task within their given condition; all tasks were counterbalanced.

1.1.3 Stimuli

Computer aided designs (CAD, designed using GMax design software) of 392 novel items were used, of which 196 were for the natural kind condition (46 juvenile bugs, 58 adult bugs and 92 transitional bug images used in the juvenile to adult animations, see Figure 8) and 196 were for the artefact kind condition (46 part sets for trudges, 58 complete trudges and 92 transitional trudge images used in the parts to complete animations, see Figure 9). The stimuli were designed to be used across category structure tasks and every stimulus was unique and differed on at least one dimension (**Natural kind:** head shape: round or angled; colour: green, yellow or purple; body: round or triangular; colour of markings: black or white; mandible-sting relation: large mandible and small sting or small mandible and large sting. **Artefact**: wheels: tracked or round; colour: green, yellow or purple; cabin: round or triangular; rim colour: black or white; front loader and back loader relation: large front, small back or small front, large back, see Appendix A for examples). Every starting image of the transformation was smaller and lighter in colour than the final image. There were three sizes of each feature in the relational condition (the trudges’ front and back loaders; the bugs’ mandibles and stings) – small, medium and large. Children had to assess which feature was the largest and smallest, and could not recognise one size as always being the largest or smallest.

1.1.4 Pre-tests

Stimuli similarity:

During the induction task in the main experiment, children saw 12 triads consisting of a target, a same-category but perceptually dissimilar category choice, and a different-category but perceptually similar distracter (see Figures 10 and 11). To make certain that the target item was perceptually more similar to the perceptual distracter test item than the same-category test item, the similarity between test and target items was validated using adult and child ratings.

Figure 10. Example of a featural artefact triad from Study 3. The trudge parts are transformed into a complete trudge (transformation), then the complete target is shown with a category choice and a perceptual choice (induction triad).

Figure 11. Example of a relational natural kind triad from Study 3. The juvenile is transformed into an adult target (transformation), then the adult target is shown with a category choice and a perceptual choice (induction triad).

Adults Pairings were shown from both domains and category structures resulting in a total of 96 pairings: every triad created two slides, one with the target and the perceptual choice and the other with the target and the category choice. Four pair matches were created: 1) natural kind target and perceptual distracter, 2) natural kind target and category choice, 3) artefact kind target and perceptual distracter, 4) artefact kind target and category choice. Adults rated the pairs for perceptual similarity on a 5-point Likert scale. A repeated measures GLM was conducted with three independent variables: 1) pair-type (target vs. perceptual distracter (TP); target vs. category choice (TC); 2) domain (natural kind vs. artefact), and 3) category structure (featural vs. relational). A main effect was found for pair-type *F* (1, 29) = 39.51; *p* < .001, with TP pairs receiving higher similarity scores than TC pairs. No effect was found for domain *F* (1, 29) = .14; *p* = .707. A main effect was identified for category structure *F* (1, 29) = 25.04; *p* < .001, an effect which seems to have been driven by the much higher score for the TP artefact relational condition (4.70) than in the featural condition (3.54). All other comparisons across the natural kind and artefact conditions seem to be more comparable in score (see Figure 12), therefore the main effect observed overall, seems irrelevant. This confirmed that adults were pairing items designed to be higher in perceptual similarity across domain and category structure.

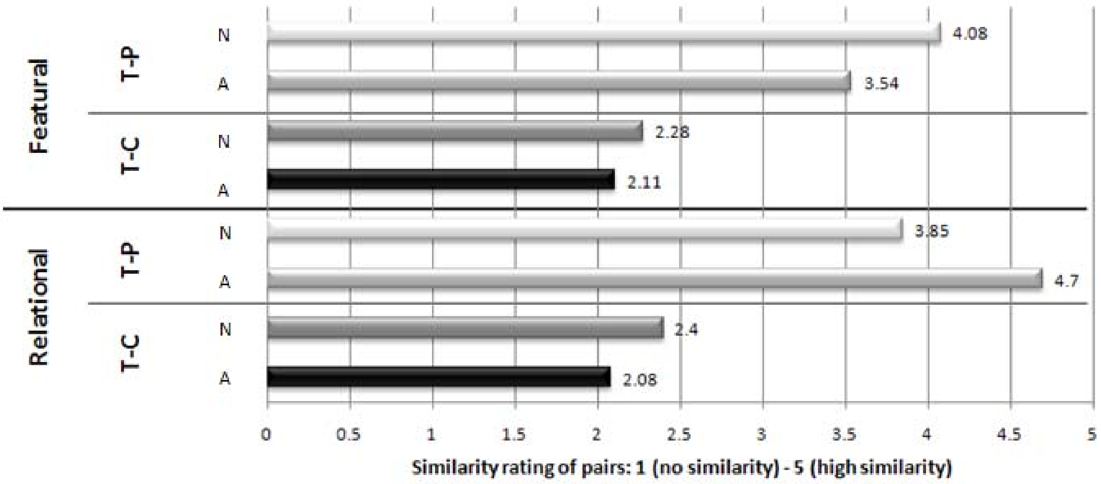


Figure 12. Adult ratings of pair similarities used in Study 3. Featural and relational conditions; target versus perceptual distracter (TP) and target versus category choice (TC) pairings; natural kind (N) and artefact (A) conditions.

Children The concept of a Likert scale is too complex for young children therefore a simpler technique was used. On the 48 induction triads shown during the main experiments (12 featural natural; 12 featural artefact; 12 relational natural, and 12 relational artefact), children had to choose which test item looked most like the target item. One-sample t-tests showed that in all triads the distracter was seen as more perceptually similar to the target than the perceptual choice (see Table

4). Children were significantly more likely to pair together items designed to be more perceptually similar to the target.

Table 4. Statistics confirming that children were pairing items designed to be higher in perceptual similarity across domain and category structure – Study 3. The t-values show that the target and the perceptual choice were paired frequently, which was significant; the target and the category choice were paired infrequently, which was also significant.

Category structure	*Domain*	*Test item*	*Times chosen (%)*	*Statistics*
Featural	Natural	Perceptual	99%	*t* (29) = -481.06; *p* < .001
Featural	Natural	Category	1%	*t* (29) = -629.08; *p* < .001
Featural	Artefact	Perceptual	91%	*t* (29) = -83.51; *p* < .001
Featural	Artefact	Category	9%	*t* (29) = -104.45; *p* < .001
Relational	Natural	Perceptual	100%	*t* (29) = -141.00; *p* < .001
Relational	Natural	Category	0%	-
Relational	Artefact	Perceptual	96%	*t* (29) = -138.22; *p* < .001
Relational	Artefact	Category	4%	*t* (29) = -177.71; *p* < .001

Domain identification and stimulus naming:

Adults A domain categorisation pre-test was conducted to check the stimuli were considered to be plausible natural kinds and artefacts. Adults were shown eight images individually (2 juvenile bugs, 2 adult bugs, 2 trudge parts and 2 complete trudges) and asked to state whether the item was a living or non-living thing, and what they believed it to be. Adult bugs were labelled as living kinds 98% of the time, with common responses being ‘scorpion’ and ‘insect’, and juvenile bugs were labelled as living kinds 92% of the time, with common responses being ‘insect’ and ‘ant’. Complete trudges were labelled as non-living 92% of the time, with common responses being ‘tractor’ and ‘digger’, and trudge parts were labelled as non-living 95% of the time, with common responses being ‘shapes’ and ‘machine’.

Children The same eight images were shown. Adult bugs were labelled as living kinds 95% of the time, with common responses being ‘bug’ and ‘ant’, and juvenile bugs were labelled as living kinds 94% of the time, with common responses being ‘caterpillar’ and ‘ant’. Complete trudges were labelled as non-living 83% of the time, with common responses being ‘scales’ and ‘digger’, and trudge parts were labelled as non-living 92% of the time, with common responses being ‘shapes’ and ‘plates’. There was no significant difference in the percentage of correct classifications between natural kinds and artefacts: t (31) = 1.42; p = .165.

Induction task:

Adults completed the induction tasks designed for children to ensure they were easy to understand and complete before they were given to children. Adults scored highly and above chance on the initial and final categorisation tasks: **Featural natural** Initial M = 100%; Final M = 99% (SD = 2.07), t (14) = 92.75; $p < .001$. **Featural artefact** Initial M = 98% (SD = 4.72), t (14) = 39.70; $p < .001$; Final M = 98% (SD = 4.96), t (14) = 37.33; $p < .001$. **Relational natural** Initial M = 96% (SD = 7.19), t (9) = 92.75; $p < .001$; Final M = 91% (SD = 7.47), t (9) = 92.75; $p < .001$. **Relational artefact** Initial M = 99% (SD = 2.53), t (9) = 39.70; $p < .001$; Final M = 94% (SD = 6.94), t (9) = 37.33; $p < .001$, and made significantly more category choices than expected by chance.

1.2 Design and Procedure

There were two between-subjects factors: year (with 5 levels: Years R, 1, 2, 3, 4) and category structure (with 2 levels: relational and featural). There was one within-subjects factor: domain (with 2 levels: natural and artefact). The dependent variable was the number of correctly categorised items during the initial and final categorisation tasks, and the number of perceptual or category choices made during the induction tasks.

Every participant completed each task individually, in a single session. The procedure followed the same four-fold methodology as Studies 1 and 2, except this time children were randomly assigned to either the relational or featural condition and completed both natural and artefact conditions within this category structure. This design ensured that a child was only exposed to each category of stimuli once, and avoided the child having to learn different types of rules for classifying the stimuli in each session.

Featural Study:

Children were taught a rule to categorise two item kinds – in this case either Desert bugs from Rocky bugs (natural kind task), or Country trudges from Town trudges (artefact task) – based on a feature: head shape for the natural kinds, and wheel type for the artefact kinds. Functions of the critical feature were provided to ensure participants did not see the features as arbitrary:

> *"Desert bugs live in the sand and have round heads for soft burrows"; "Rocky bugs live in rocks and have pointy heads for digging"; "Country trudges are used in fields and have tracked wheels for muddy surfaces"; "Town trudges are used in cities and have round wheels to work on paved surfaces".*

Relational Study:

The relational task involved more complicated rules using feature-size comparison:

> "*Desert bugs live in the sand and have larger mandibles to catch prey walking 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 in fields, and a smaller backhoe to move small stones"; "Town trudges are used in towns and have a smaller front loader to carry small materials, and a larger backhoe to break concrete"*

The procedure followed the same four-fold methodology as previous studies (1) category learning, 2) initial categorisation, 3) induction tasks, 4) final categorisation; see Appendix C for all hidden properties), with 12 trials per task.

1.3 Results

1.3.1 Categorisation Performance

Only children who performed significantly above chance in both initial and final categorisation tasks were included in the analyses (scores of 10/12, 83% and above; binomial test proportion = 0.5, *p* = .04). Fifty-five children were unable to pass the categorisation tasks for either domain (43 Reception: 10 featural and 33 relational; 9 Year 1: all relational, and 3 Year 2: all relational); 37 children were unable to pass the categorisation tasks for one of the two domains (17 Reception; 15 Year 1; 4 Year 2, and 1 Year 4, of which 15 children failed the artefact task and 21 failed the natural task). Since I cannot be sure that these children fully understood the categorisation rule, their data was removed before conducting my analyses of induction performance. The remaining 311 children (45 Reception: 42 featural and 3 relational; 85 Year 1: 51 featural and 34 relational; 95 Year 2: 52 featural and 43 relational; 41 Year 3: 21 featural and 20 relational, and 45 Year 4: 25 featural and 20 relational) scored highly and above chance on the initial and final categorisation 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. **Artefact:** 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. Since only 3 Reception children remained for the relational condition, only Years 1-4 were used in the induction analyses.

1.3.2 Induction Performance

The proportion of category choices made by each participant in the induction task was examined, with participants who made at least 10/12 category choices considered to have a category-bias (binomial test proportion 0.5, *p* = .04). Participants who made less than 2/12 category choices were considered to have a perceptual-bias (binomial test proportion 0.5, *p* = .04). Due to the removal of children who were unable to pass both of the categorisation tasks, the number of reception children remaining in the relational condition was three, therefore they were not included in any of the analyses.

A mixed GLM was conducted with children from Year 1-Year 4 (266 in total) to examine whether the proportion of category choices made in the induction task differed by year group, domain and category structure. A significant main effect of year group was found: *F* (3, 258) = 10.723; *p* < .001. As expected, the proportion of category choices increased with year (Year 1: *M* = 39.30, *SD* = 42.45; Year 2: *M* = 50.48, *SD* = 44.29; Year 3: *M* = 70.24, *SD* = 39.14; Year 4: *M* = 71.50, *SD* = 39.99). No significant main effect of domain was found (natural vs. artefact): *F* (1, 258) = 2.551; *p* = .111. A significant main effect of category structure was observed (featural vs. relational): *F* (1, 258) = 20.76; *p* < .001, with children making fewer category choices in the relational condition (featural: *M* = 64.56, *SD* = 41.40; relational: *M* = 41.50, *SD* = 45.35). These findings suggest that domain made no difference on strategy preference (perceptual or category). However, the category structure of the stimuli did have an effect. All interactions were non-significant.

In order to examine changes in children's strategy with year, I calculated the percentage of children showing each type of bias in each year group. Since the GLM indicated that the pattern varied by category structure but not domain, I calculated these percentages separately for the featural and relational conditions, collapsed across the two domains. Figures 13 and 14 show a gradual transition from perceptual to category bias for both conditions, but with greater numbers of children making category choices in the featural condition. Although the Reception group was not included in the GLM analysis because of low numbers of participants (N = 3) in the relational condition, it is interesting to note that the majority of the 42 children left in the featural condition showed a bias to select the perceptual choice. Similarly, the majority of Year 1 children in the relational condition showed a significant perceptual bias. From Year 2 in the featural condition, and from Year 3 in the relational condition, the majority of children showed a significant bias to select the category choice.

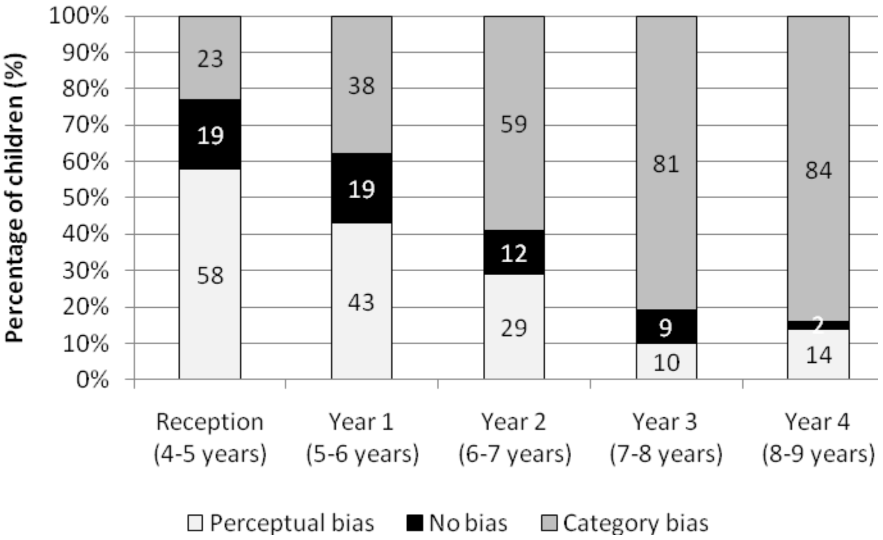


Figure 13. The percentage of children making different induction choices across the year groups in the featural rule condition – Study 3.

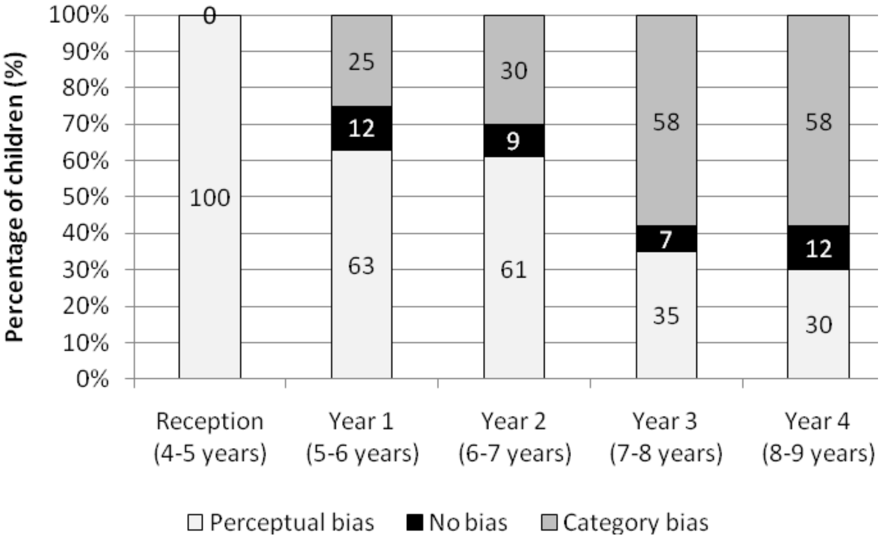


Figure 14. The percentage of children making difference induction choices across the year groups in the relational rule condition – Study 3.

1.4 Discussion

The data yielded three main findings. Firstly, a clear preference for perceptual induction in the younger year groups was observed, which transitions to a clear preference for category induction in the older year groups. This finding is consistent with research supporting the perceptual-bias account (Sloutsky et al., 2007) and the relational shift (Gentner, 1988; Gentner & Rattermann, 1991; Rattermann & Gentner, 1998). Specifically, young children focus on obvious perceptual characteristics to inform their reasoning, and only learn to focus on less obvious category information as they develop a greater knowledge base, and gain a heightened understanding of the importance of category membership. Secondly, there were no differences in inductive strategy preference between the domains of natural kind and artefact, suggesting that inductive reasoning develops in a domain-general fashion (counter to predictions derived from Goswami, 2001; Keil, 1989; Rattermann & Gentner, 1998; Wellman & Gelman, 1998). Finally, the rate of transition from a perceptual to a category induction bias was dependent on category structure. Although the developmental trajectory was similar for both category structures, fewer category choices were made overall for the relational condition, and it took a further year for the majority of children to show a category bias, compared with the featural condition. Since the relational condition is likely to tap into EF abilities more heavily (due to the complexity of the category structure), this finding suggests that EF may be a limiting factor in the development of category induction (Bunge & Zelazo, 2006; Halford, 1993; Halford et al., 1998; Zelazo & Frye, 1998; Zelazo & Müller, 2002). This could be a factor linked to the driving force behind inductive reasoning. However, before starting investigation into the mechanisms behind induction development, I decided to explore one final methodological ambiguity from previous research. Having so far avoided using labels during the induction tasks, in case of priming, I decided to investigate whether labels could influence the developmental trajectory of induction.

2 Study 4: The effect of stimuli labels

As I mentioned in the introduction, there is an on-going debate as to whether children need linguistic information to use category induction. Some researchers believe that providing item labels merely identifies the category of items, which allows people to apply category induction. In fact, a preference for category induction has been shown in children as young as 2 years old when items have shared category labels (Gelman & Coley, 1990; also Deák & Bauer, 1996). Whereas, other researchers have suggested that providing category labels merely contributes to the similarity judgement: children see labels as a good predictor of similarity rather than using it to understand the relational connection between items (see Sloutsky et al., 2001 for a review). However, children aged 4-5 years have also been shown to have a clear preference for category information, even when items were presented with synonym labels (for example, ‘rabbit’ and ‘bunny’; Gelman & Markman, 1986). Nonetheless, synonyms still provide more similarity than no labels, and it has been argued that children's "…induction could be label based without being category based" (Fisher & Sloutksy, 2005 p. 584). It has also been proposed that an auditory label is seen as the best predictor of similarity (more so than visual cues) due to it grabbing children's attention (Napolitano & Sloutsky, 2004; Sloutsky & Napolitano, 2003).

Study 4 follows the same procedure as Study 3, using the featural natural kind task only. Children were divided into a 'with labels' condition and a 'without labels' condition. Based on this design, researchers from the two broad accounts would again predict different results. Researchers in support of the perceptual-bias account would suggest that young children would choose the category choice more often in the 'with labels' condition because they would see the auditory label as the best predictor of similarity (Sloutsky et al., 2001). When the label was not available in the 'without labels' condition, young children would revert to using perceptual cues for induction. In contrast, researchers in support of the category-bias account would suggest that as long as children understood the relational connection between items (i.e., those that could pass the categorisation tasks), they would have a natural bias to use category induction regardless of whether the items were labelled during the induction task (Gelman & Markman, 1986).

2.1 Method

2.1.1 Pre-study participants

All pre-tests were completed during the Study 3 set-up.

2.1.2 Main study participants

Ninety-eight primary school children participated: 31 Reception (4-5 years), 33 Year 1 (5-6 years) and 34 Year 2 (6-7 years); 50 males and 48 females, each of whom was tested individually.

2.1.3 Stimuli

This Study used the featural natural kind task (and stimuli) from Study 3.

2.2 Design and Procedure

There were two between subjects factors: year (with 3 levels: Years R, 1, and 2) and labelling (with 2 levels: 'with labels' and 'without labels'). The dependent variable was the number of correctly categorised items during the initial and final categorisation tasks, and the number of perceptual or category choices made during the induction tasks. Each participant completed the task individually, in a single session. The procedure and stimuli were identical to the featural natural kind task in Study 3, except this time children either did or did not receive labels during the induction task:

Labels: *"If this Desert bug* [point to adult target] *has [stated property] then does this Desert bug* [point to test image 1] *or this Rocky bug* [point to test image 2] *also have [stated property]?"*.

No labels: *"If this one* [point to adult target] *has [stated property] then does this one* [point to test image 1] *or this one* [point to test image 2] *also have [stated property]?"*.

2.3 Results

2.3.1 Categorisation Performance

Only children who performed significantly above chance in both initial and final categorisation tasks were included in the final sample. According to a binomial test, scores of 10/12 (83%) and above were significantly above chance (proportion = 0.5, p = .04). Six children (2 Reception, 'with labels' condition; 3 Reception and 1 Year 1, 'without labels' condition) scored at or below chance and were removed from the analyses. The remaining 92 children scored highly and above chance on the initial and final categorisation tasks: Initial M = 98% (SD = 4.07), t (91) =113.39; p < .001; Final M = 98% (SD = 3.99), t (91) = 115.12; p < .001.

2.3.2 Induction Performance

The proportion of category choices made by each participant in the induction task was examined, with participants who made at least 10/12 category choices considered to have a category-bias (binomial test proportion 0.5, p = .04). Participants who made less than 2/12 category choices were considered to have a perceptual-bias (binomial test proportion 0.5, p = .04). As shown in Figure 15, the percentage of children with a category-bias increased with year in both conditions, although the percentage of category choices is clearly higher overall in the 'with labels' condition.

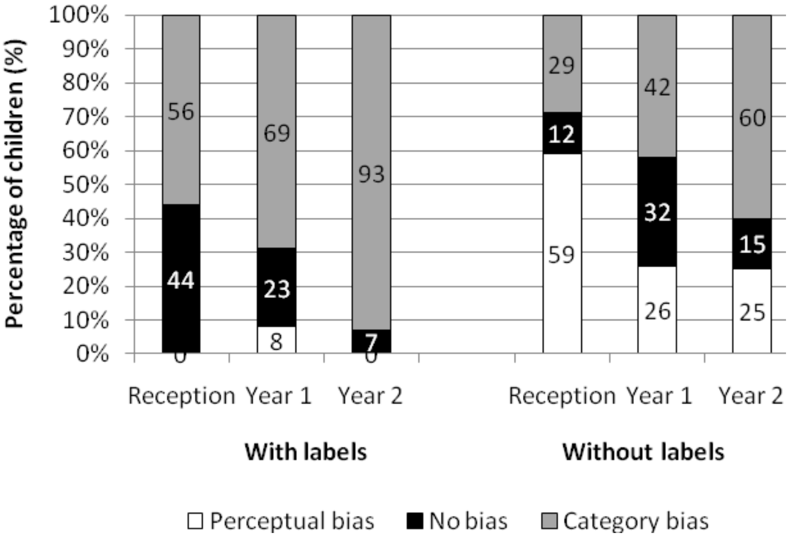


Figure 15. The percentage of children showing each type of bias across the three year groups in the 'with labels' and 'without labels' conditions – Study 4.

I used one-sample t-tests to investigate the overall percentage of category choices made by each year group and condition. '**With labels':** all year groups made significantly more category choices than expected by chance: Reception, *t* (8) = 4.01; *p* = .004, Year 1, *t* (12) = 2.67; *p* = .02, Year 2, *t* (13) = 8.93; *p* < .001. '**Without labels':** none of the groups showed a significant preference for either perceptual or category induction: Reception, *t* (16) = -1.68; *p* = .112, Year 1, *t* (18) = .516; *p* = .612, Year 2, *t* (19) = 1.31; *p* = .206, however, the mean percentages show an increased preference towards category induction (see Figure 15 for percentage of children and Tables 5 and 6 for mean percentages).

Table 5. 'With labels' condition: the percentage of category induction choices for each year group – Study 4.

Year Group	*Mean initial score (SD)*
Reception	80% (22.13)
Year 1	76% (35.63)
Year 2	93% (17.96)

Table 6. ‘Without labels’ condition: the percentage of category induction choices for each year group – Study 4.

Year Group	*Mean initial score (SD)*
Reception	33% (40.90)
Year 1	55% (40.48)
Year 2	63% (42.86)

Effects of Labelling

A two-way between-subjects ANOVA was used; the two independent variables were year (Reception, Year 1 or Year 2) and labelling (‘with labels’ or ‘without labels’). The dependent variable was the percentage of category choices made. The analyses revealed a main effect of labels, with more children making category choices in the ‘with labels’ condition, compared to those in the ‘without labels’ condition: F (1,86) = 17.19; $p < .001$ (see Figure 15). There was a trend towards a year effect, with older children making more category choices, however this effect was non-significant: F (2,86) = 2.41; $p = .10$. There were no significant interactions.

2.4 Discussion

Researchers from both accounts suggest that labels affect the inductive strategy used by young children. Those in support of the category-bias account claim that labels are proxies for categories, highlighting relational connections between items, which encourage category-induction. Those in support of the perceptual-bias account claim that labels are features for categories, adding an extra level of overall similarity, which creates label-based induction. My findings showed a great difference between induction strategy preference with and without labels. Children in both conditions were highly proficient at differentiating between the categories shown, yet only in the ‘with labels’ condition did all year groups show a significant preference for category induction. If the theory made by researchers supporting the category-bias account were correct, then all children in both conditions should have shown a category-bias, because all children were aware of the category name for every item. However this was not the finding, which leads me to lean towards the prediction made by researchers in support of the perceptual-bias account. It seems more likely that children in the ‘with labels’ condition were provided with an extra level of similarity between items, on which they based their decision. Although this study cannot provide me with the exact reason for children’s category preference in the ‘with labels’ condition, I have decided to avoid using labels in the rest of my studies to avoid potential priming.

3 General discussion of Studies 3 and 4

Previous research into inductive reasoning has produced contradictory results (Gelman & Markman, 1986; Opfer & Bulloch, 2007; Sloutsky et al., 2007). Where Gelman and colleagues have found that young children’s inductive preference was category-biased, Sloutsky and colleagues have found that young children’s inductive preference was perceptually-biased. Since these studies did not systematically control the nature of the categories children were confronted with, it is not possible to directly compare these findings. I present the first systematic investigation of the development of category induction in which category structure and domain were explicitly manipulated. I also present findings about the impact of labels during my inductive reasoning tasks.

3.1 Domain

I found a domain-general transition from a perceptual to a category bias in inductive reasoning. There are several possible explanations for this finding. Firstly, it is possible that children could not differentiate between domains of the stimuli 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 very unlikely. In fact, the pre-test was a very harsh test of the plausibility of my stimuli since children were only shown the images briefly, having 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 behaviour. Secondly, it is possible that causal assumptions are not elicited by novel stimuli, as children do not believe the items to be real. However, the stimuli were no less realistic than pictures children are frequently exposed to in books and on television, and children readily engage with such stories appropriately (Ganea et al., 2010). In fact, during the experiment, children often spontaneously commented on the stimuli as if they were pictures depicting real things, for example, *"I've seen a Rocky bug in my garden"* or *"I know that's a City trudge because they go down my road sometimes".* Thus suggesting that the images were seen as plausible and real, making this second explanation unlikely. Finally, this domain-general reasoning could suggest that heightened understanding of the importance of hidden properties and 'essences' in natural kind category membership is learned. Contrary to previous research (Gelman, 2003; Keil, 1989; Markman, 1989), this study suggests that essentialism is not a natural default: children were no more advanced at assigning category induction for the natural kinds than they were for the artefact kinds. This suggests that young children did not instinctively recognise an 'essence' in the natural kinds as being more important for induction decisions. In addition, I found no evidence that domain-specific reasoning strategies emerged with age, counter to predictions made based on the work of Newman and Keil (2008). Possibly the best explanation for these unexpected findings is that I had carefully matched the natural kind and artefact stimuli so that the key features determining category membership were equally non-obvious in each case, and equally important for the function or behaviour of the item. Thus, apparently domain-specific reasoning may simply reflect children's sensitivity to the specific stimuli chosen. For many real-life artefact categories, obvious surface characteristics are sufficient to determine their function, for example, the blade of a knife; the seat of a chair. However, this is not always the case, and especially for more complex artefacts (e.g., machines and vehicles). In this study, the key features determining category membership were equally non-obvious for natural kinds and artefacts, and equally strongly associated with the function or behaviour of an item. Children readily adapted to the characteristics of the specific item they were confronted with, regardless of domain. The data suggests that there is, in fact, no inherent essentialist bias when reasoning about natural kinds, and instead children develop an understanding of the importance of category membership through learning about the relevant characteristics of categories. These findings are therefore consistent with Sloutsky et al's (2007) hypothesis that categorisation precedes understanding about the importance of category significance, and knowledge is essential for the progression from perceptual induction to category induction.

3.2 Category Structure

A gradual transition was observed from a perceptual to a category preference for both category structures. However, the progression was delayed a year for the relational category structure. There are three possible explanations for this lag in the relational categories. Firstly, it could be argued that children struggled more with the relational category structures because as young children, they are more used to categorising based on featural structures. However, several researchers have stated that in the real world, categories are more often defined in terms of relational structures rather than featural structures (Kittur et al., 2004). This would suggest that children would be used to recognising and identifying items based on relational structure, and therefore makes this explanation unlikely. Secondly, the categorisation rule in the relational condition may have been too complicated for children to follow. Although more children failed to meet my categorisation criteria in the relational condition, suggesting that learning the categorisation rule was harder, these children who were unable to categorise the items were not included in my analyses of induction performance. In fact, the children included in my analyses scored 10/12 or above in both initial and final categorisation tasks. There is no doubt that they were proficient in following the categorisation rule, and remembered the rule throughout the duration of the experiment without reminder or prompt. Finally, it is possible that the extra cognitive load required for processing the relational category structure resulted in greater demands on EF. According to Zelazo and colleagues, this type of performance is due to a "…lack of reflection on rules, not to a lack of consciousness of rules" (Zelazo et al., 1996; p. 41). Children fail to apply their category knowledge in the induction task because their EF has not developed enough to allow them to reflect on the importance of the category rules. Previous research suggests that children aged 2 years old can only reason at a binary level and it is not until age 5 years that children have the capability to reason at a ternary level (Andrews & Halford, 2002; Halford, 1993). Nevertheless, the majority of the youngest children in the featural condition, where the EF demands are much less (arguably in line with a binary cognitive load), still showed a preference for perceptual induction. Importantly, I have replicated Sloutsky et al's (2007) finding of an early bias for perceptual induction, even for the featural condition in which the cognitive load is significantly reduced. Overall, these findings suggest that although an increase in

EF ability is likely to support the development of category induction, it is unlikely to be the driving force behind the development of category induction. The category structure of items may limit children's ability to focus on category-relevant information during induction, but knowledge and experience are essential to enable the child to understand the premise that category membership is important. These findings are therefore consistent with the hypothesis that categorisation precedes understanding about the importance of category membership, and an accumulation of knowledge and experience are necessary for sophisticated analogical reasoning (Gentner, 1988) and category induction (Sloutsky et al., 2007). Importantly, this data has highlighted the effect that differing levels of category structure and complexity have on young children's inductive strategy. It goes some way to explaining why researchers using more featural-like stimuli (Gelman and colleagues) found evidence of category induction at a younger age than those using more complex relational stimuli (Sloutsky and colleagues). Category structure affects the developmental trajectory of induction.

3.3 Labels

An increase in the proportion of category-based induction decisions was observed for all year groups completing the 'with labels' condition. In fact, all year groups in this condition made significantly more category choices than expected by chance, compared to the 'without labels' condition, where only Year 2 showed a strong trend towards category induction. Thus clearly showing that adding labels during the induction trial biases category choice. However, as mentioned in the introduction, there are two possible reasons behind this increase. Firstly, it could be suggested that adding labels during the induction trials links the commonality between items, and allows category-based induction by conveying category membership (Gelman, 2003). Secondly, it could be suggested that adding labels during the induction trials creates an extra feature for the categories, which enhances the overall similarity between items, and increases the number of category choices made. However, this increase is due to children making inferences based on a label-based preference rather than a category-based preference (Sloutsky & Lo, 1999).

Study 4 cannot identify the exact reason for children's increased category preference in the 'with labels' condition, however the findings lend more support to those supporting the perceptual-bias account. All children completing the task, in both conditions, were highly proficient at differentiating between the categories presented, but only in the ‘with labels’ condition did all year groups show a significant preference for category choice. The children already knew each item’s category membership, therefore it is unlikely that the labels were providing any new information. It seems more likely that the labels added an extra level of similarity between items, on which children based their decisions. Importantly though, it is clear that the trajectory of induction can be affected by the inclusion of labels, which leads me to believe that more informative data will be observed without them.

3.4 Conclusion

I have presented the first systematic evidence of a gradual transition from a perceptual to a category bias for relational and featural categories from both artificial and natural kind domains. I found that the development of category induction strategy is domain general, but delayed by approximately a year for relational category structures. The delay is possibly due to the greater demands on processing and EF. These findings make an early category induction bias unlikely. Instead, I argue that children gradually develop an appreciation of the importance of category membership, through an accumulation of knowledge, which drives the development of category-based inductive reasoning. Finally, due to a significant alteration in strategy preference with the inclusion of labels during the induction trials, I conclude that in order to determine when children spontaneously use category induction, without prompting, labels are best avoided.

Chapter 4: Could the development of cognitive control affect the development of induction?

In Chapters 2 and 3, I investigated the influence of methodological manipulations on the developmental trajectory of induction, and explored the contradictory findings about the natural default of inductive reasoning. In Chapter 2, I explored the developmental trajectory of inductive reasoning across ages 3-9 years, and discovered a gradual transition from perceptual to category induction, which supports researchers in favour of the perceptual-bias account. I also found that children were not making decisions based on maturity groupings, and the transition could not be due to an inability to inhibit highly similar distracters, as all year groups were equally distracted. In Chapter 3, I considered whether previous contradictory findings could be due to the category structure of stimuli used, and found that when children were presented with relational category structures, their use of category induction was delayed by a year compared to those presented with featural category structures. Interestingly, I discovered that label use during induction trials does affect inductive preference, with significantly more category choices being made when labels were present. Importantly, children's strategy did not differ between the natural kind and artefact domains, suggesting that inductive reasoning is domain-general. Through these studies, I have highlighted crucial differences between stimuli and methodology in studies supporting each of the two accounts, which could have contributed to previous conflicting results.

In this chapter, I turn my attention to the driving force behind the development of inductive reasoning. Although my findings from Study 1 suggested that the gradual transition is unlikely to be due to young children's inability to inhibit, I wondered whether there were other aspects of an undeveloped cognitive control that may be influencing induction. The hypothesis was motivated by my findings from Study 3, which showed that when faced with more cognitively demanding tasks – relational versus featural category structures – children's category induction was delayed by a year. Study 5 explores the potential relationship between the development of cognitive control through EF tasks, and the development of category induction through an inductive reasoning task.

1 Study 5: The influence of executive functions on the development of induction

Executive function (EF) is a term used to describe various higher order cognitive processes, including inhibition, planning, working memory and shifting (Denckla, 1994; Dennis, 1991; Lezak, 1993; Welsh, Pennington & Grossier, 1991). It is well established that the development of EF shows the most dramatic increase during the pre- and early school years (see Espy, Bull, Martin & Stroup, 2006), and certain researchers have linked young children's difficulties in general reasoning tasks to certain underdeveloped EFs such as inhibition (Richland et al., 2006) and working memory (Halford and colleagues; Zelazo and colleagues). However, there is no literature directly assessing the relationship between EF and inductive reasoning. Two categories of EF have been proposed. Firstly, 'hot' affective EFs associated with the ventral and medial regions of the prefrontal cortex (VM-PFC), which are elicited when a person is faced with "…regulation of affect and motivation" (Hongwanishkul, Happaney, Lee & Zelazo, 2005; p. 618), for example trying to choose advantageously during a gambling task. Secondly, the purely cognitive 'cool' EFs associated with the dorsolateral regions of the prefrontal cortex (DL-PFC), for example decontextualised problems such as sorting colours or remembering abstract number sequences (Zelazo & Müller, 2002). Some research has suggested that cool and hot EFs are in fact distinct, and show different patterns of development (Hongwanishkul, et al., 2005). However, research into the development of hot EF is limited, and the exact developmental trajectory and relationship with cool EFs is still unclear.

Since the development of certain EFs may affect children's induction strategy, I decided to test children on the featural natural kind inductive reasoning task used in Study 3, alongside a battery of EF tasks: three cool simple, one cool complex, and one hot complex. The simple EFs (working memory, shifting, inhibition) were chosen because in the literature they are considered to be three of the most important EFs for cognition, and can be assessed through reliable standardised tests (Baddeley, 1996; Logan, 1985; Miyake, Friedman, Emerson, Witzki & Howerter, 2000; Rabbitt, 1997). I included two complex tasks – one hot and one cool – that measure a mixture of the three simple EFs (Miyake et al., 2000), as well as other cognitive skills such as planning and decision-making. Every child took part in all tests over three sessions. I decided to include a battery of EF tasks because it is still in debate as to what extent EFs can be consider unitary, in the sense that they tap into the same underlying mechanism or ability. There are mixed views as to whether EFs are unitary and show significant intercorrelations, or whether they are nonunitary and tap into different abilities (see Miyake et al., 2000, for a review).

I had three predictions. Firstly, that age will be a significant predictor of not only inductive reasoning (showing a gradual transition from perceptual-based to category-based induction), but also a significant predictor of performance on the EF tasks, with children improving with age. Secondly, there will be an association between the use of category induction and the performance on EF tasks, with those children who perform better on the EF tasks showing a stronger category preference during the induction task. Finally, the cool EF tasks will be unitary and strongly correlated with each other, as they tap into the same underlying mechanism – the DL-PFC – but they will not be as strongly correlated with the hot EF task which taps into the VM-PFC.

1.1 Method

The tasks in this experiment have been standardised, therefore no pre-tests were conducted.

1.1.2 Main study participants

60 primary school children participated in total: 20 Reception (4-5 years), 20 Year 1 (5-6 years), and 20 Year 2 (6-7 years); range 4.5 – 7.4 years; 32 males and 28 females, each of whom was tested individually, and completed all EF tasks and an inductive reasoning task, over 3 sessions.

1.2 Design and Procedure

There was one between-subjects factor: year (with 3 levels: Reception, Year 1 and Year 2). There was one within-subjects factor: task (with 5 levels: inductive reasoning, Digit Span (working memory), Shape School (inhibition and shifting), Tower of London (cool complex) and Children's Gambling (hot complex). The dependent variable was the score awarded for each task.

1.2.1 Main study participants

Children took part in 5 tasks; 1 inductive reasoning task and 4 EF tasks (the Shape School is one task but taps into two of my measures – inhibition and shifting – in different sections of the task).

Each child was tested three times over a three-week period to shorten each testing session to 20-25 minutes maximum. Session one comprised of the inductive reasoning and Tower of London tasks, session two comprised of forwards Digit Span, the Shape School and backwards Digit Span, session three comprised of the Children's Gambling task. The order was assigned based on length of set-up and task, and all children completed the sessions and tasks in the same order. Figure 16 graphically shows the relationships to be explored.

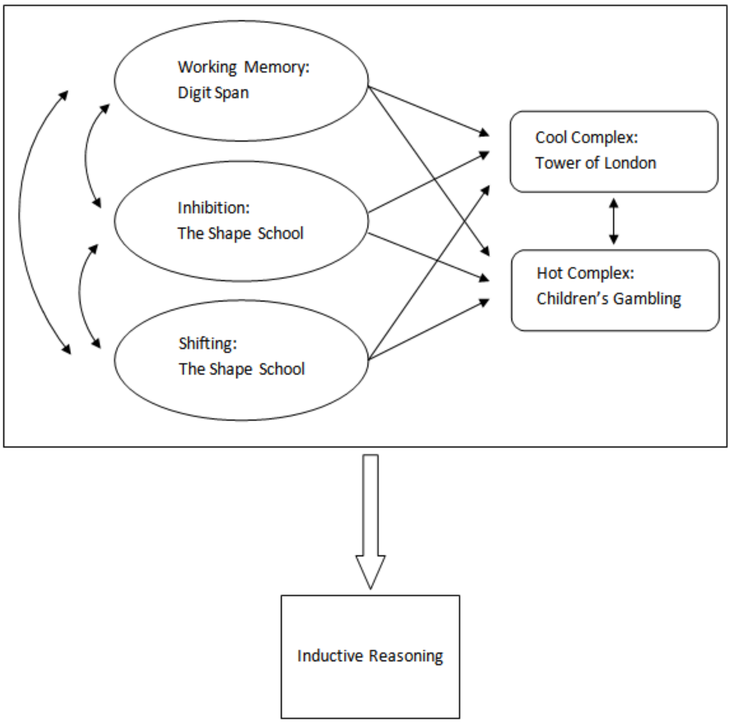


Figure 16. The relationships observed in Study 5: the influence of simple and complex EFs on inductive reasoning, and whether any of the EF tasks correlate with the other EF tasks. The chart was inspired by Miyake et al., (2000).

Inductive reasoning

Sandbug-Rockbug. Each participant completed the task individually, in a single session. The procedure followed the same four-fold methodology as previous studies (1) category learning, 2) initial categorisation, 3) category induction, 4) final categorisation; see Appendix B for hidden properties). This time children only viewed the natural featural condition used in Study 3, but the items were called Rockbugs and Sandbugs.

Simple EF tasks

Digit Span. Digit Span is a cool EF task tapping into working memory, and comprises a forwards and a backwards task (The British Ability Scales 2; Elliott, 1996). One list of 36 number items (forwards Digit Span) and one list of 30 number items (backwards Digit Span) were used (see Appendix D). Z-scores were calculated for each task: $\frac{Score - Mean\ of\ all\ scores}{SD\ of\ all\ scores}$ and were averaged, creating one Digit Span score for each child, based equally on their forwards and backwards Digit Span performances.

The Shape School. The Shape School is a cool EF task tapping into inhibition and shifting (extended U.K. version; Espy, 1997; Ellefson, Blagrove & Espy, in preparation). A folder of the Shape School task story was used. The task assesses working memory and shifting through four story sections and "…colorful, affectively engaging stimuli presented in an age-appropriate and appealing format, a storybook", see Figure 17 (Espy et al., 2006; p. 375). The Shape School comprises of four sections: control (the baseline), inhibition (children must inhibit certain images), shifting (children must shift between naming images based on colour or on shape) and both (children must perform the shifting task, but are also required to inhibit certain images). These will be explained individually, but during the Shape School task, follow a continual story-like format. To avoid confusion, images of the Shape School children will be referred to as 'SSc'; participating children will be referred to as 'children'.

Figure 17. Images from the Shape School used in Study 5 (Espy, 1997; images presented with permission from Dr. Kimberly Andrews Espy). From left to right: a neutral SSc from the control condition; a sad SSc from the inhibition condition; a SSc with a hat from the shift condition, and a happy SSc with a hat from the both condition.

Control condition

Children were introduced to two classes of SSc, Mr Square's and Ms Circle's (red and blue square and circle shapes with faces, arms and legs; see Figure 17), and were told that their names were their colours, for example a red SSc was called Red. Children were told that the SSc are lining up for break; after a practice of six pictures (corrected by the experimenter if needed), the children were shown 48 images of SSc queued up for break and were asked to name them as quickly as possible without making any mistakes. Children were timed.

Inhibition condition

As the story continued, children were told that it was lunchtime and all the SSc were lining up for lunch. However this time, if a SSc had a sad face it meant that they did not finish their work and were therefore not ready for lunch. The children were told that when naming the SSc, they must not call to lunch the children with sad faces (thus inhibiting certain images). There was a practice of six SSc (corrected by the experimenter if needed), before being shown the 48 images made up of SSc with happy and sad faces, and were asked to name only the happy faced SSc as quickly as possible without making any mistakes. Children were timed.

Shifting condition

As the story continued, children were told that the SSc were lining up for story time with another class (Ms Hat's); this group of children were wearing hats. Children were told that the names of the SSc with hats are their shapes (not their colours), for example, a square SSc with a hat is called Square. Children were given six practice images of SSc with and without hats (corrected by the experimenter if needed), then shown 48 images made up of SSc with and without hats, and were asked to name them as quickly as possible without making any mistakes. Children were timed.

Both condition

As the story continued, children were told that the SSc were lining up for art, however those that had not finished their stories could not go to art and had sad faces; the children must not call the names of the SSc with sad faces. They were given six practice images (corrected by the experimenter if needed), then shown 48 images made up of SSc with and without hats, with happy and sad faces, and were asked to name only the happy faced SSc (these could be SSc with or without hats) as quickly as possible without making any mistakes. Children were timed. An efficiency score was calculated for every condition using the following equation: $\frac{Number\ correct - Number\ of\ errors}{Response\ time\ (in\ seconds)}$.

Complex EF tasks

Tower of London. The Tower of London (ToL; Shallice, 1982) is a cool complex measure of EF tapping into strategic planning and problem-solving (Sikora, Haley, Edwards & Butler, 2002). Children were presented with two ToL sets; each set had a wooden base with 3 pegs of ascending size and 3 balls of the same size, 1 green, 1 blue and 1 red (see Figure 18). Children were given one ToL set and the experimenter had the other. The experimenter set up a 'target' position and set up the children's ToL set in the 'start' position; the children were told that they must make their ToL set look exactly like the experimenter's ToL set, in as few moves as possible. They were told that there are some rules: 1) they may only move one ball at a time, 2) all balls must be on a peg before another ball is moved, 3) balls cannot be balanced on another ball if there is no peg left (e.g., peg one only has room for one ball, therefore children must not balance a second ball on top).

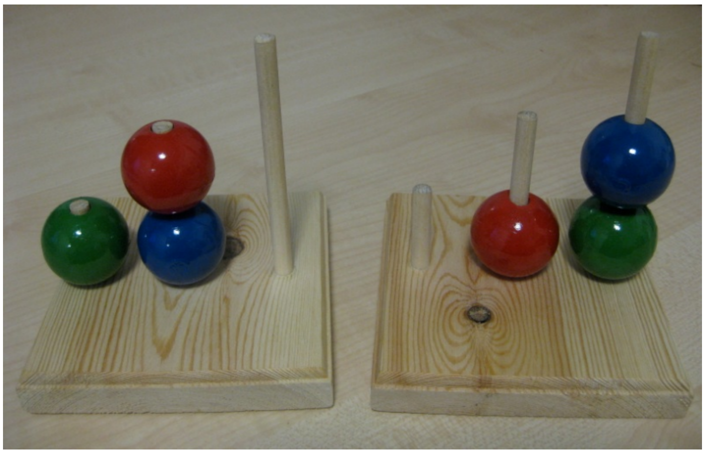


Figure 18. The Tower of London sets used in Study 5. These sets were created specifically for this experiment, but were based on the original Tower of London task by Shallice (1982). This is an example of a 4-move trial. The experimenter's set is on the left; the child must make his set (on the right) look exactly like the experimenter's set in only four moves.

This task consisted of 15 trials, 3 requiring 1 move, 3 requiring 2 moves, 3 requiring 3 moves, 3 requiring 4 moves and 3 requiring 5 moves. Trials were presented in order of difficulty (easiest first). The experimenter recorded the trial, the required number of moves, the number of moves made, the time taken (in seconds), and whether or not the child completed the trial. Children were awarded one mark for every trial they successfully completed in the number of moves required; if a child completed a trial but went over the moves required they did not get a mark.

The Children's Gambling task. The Children's Gambling task is a complex hot EF task tapping into affective decision-making (Kerr & Zelazo, 2004), based on the Iowa Gambling task designed for adults (Bechara, Damasio, Damasio & Anderson, 1994). This game placed children in a situation which caused emotional consequence: they were told that if they did well, they could take home lots of stickers; if they did not do well, they could end up taking home no stickers.

Two sets of laminated cards were used, each containing 2 decks: one advantageous and one disadvantageous deck (see Figure 19); each deck consisted of 50 cards. On the back of one set of cards, the advantageous deck was white with black spots and the disadvantageous deck was white and black stripes; on the back of the other set of cards, the disadvantageous deck was white with black spots on the back and the advantageous deck was white and black stripes. Each card was divided with the top half white and the bottom half black. Black happy faces were on the white half, and white sad faces were on the black half. Children won a sticker (the reward) for every happy face, and lost a sticker for every sad face. The black section was covered with a Post-It note so that the experimenter had to lift it to reveal any sad faces (and therefore sticker losses). Children were forced to consider their sticker wins before any sticker losses. The original study by Kerr and Zelazo (2004) used M&M chocolate candies as rewards, but I decided that for hygiene reasons, I would use stickers as rewards instead.

Figure 19. The Children's Gambling task used in Study 5. This game was created specifically for this experiment, to replicate materials used in Kerr and Zelazo's (2004) Children's Gambling game.

Two decks of cards were placed face down in front of the child (the order was counterbalanced). Also in front of the child were two containers, one for the stickers children won, and one in front of the experimenter where stickers were kept, taken from or returned. When children won stickers, stickers from the experimenter's container would be placed in the child's container; when children lost stickers, stickers from their container would be removed and placed back in the experimenter's container. Children had to choose a card from either deck of cards and place it on the table face up. A sticker was awarded for every happy face; a sticker was lost for every sad face under the Post-It note. If a child consistently chose cards from the advantageous deck, they would end up winning more rewards overall; if a child consistently chose from the disadvantageous deck, they would end up with few or no rewards overall.

1.3 Results

Although all children were asked to take part in every task, there were occasions where children refused, were unable to complete a task, or were absent during the final stages of data collection. This led to missing data: 4 children were missing data on the inductive reasoning task (3 Reception, 1 Year 1); 2 on the Shape School Shift task (2 Reception) and 1 on the Children's Gambling task (1 Year 2).

1.3.1 Inductive reasoning Categorisation Performance

Only children who performed significantly above chance in both initial and final categorisation tasks of the inductive reasoning task were included in the final sample, as this was the dependent variable during regression. According to a binomial test, scores of 10/12 (83%) and above were significantly above chance (proportion = 0.5, *p* = .04). Four children (3 Reception and 1 Year 1) scored at or below chance and were removed from analyses (therefore these four children were missing data on this task, as stated above). The remaining 56 children scored highly and above chance on the initial and final categorisation tasks: Initial *M* = 98% (*SD* = 4.16), *t* (55) = 86.47; *p* < .001; Final *M* = 98% (*SD* = 4.36), *t* (55) = 81.89; *p* < .001.

My results are split into two sections, 1) exploratory screening of the data to ensure normality, 2) statistical analyses of correlation and regression.

1.3.2 Exploratory Analyses

Before any exploratory analysis to check for normality and distribution, the data were processed to create one variable (covariate) per measure. **Inductive reasoning**: the percentage of category choices made during the induction trial for each participant. **Tower of London**: the raw score for each participant. **Digit Span**: the raw scores for each participant's forwards and backwards tests were transformed into z-scores using the following equation: $\frac{Score - Mean\ of\ all\ scores}{SD\ of\ all\ scores}$. These two z-scores were averaged to create one z-score which was equally influenced by both forwards and backwards performance. **Children's Gambling**: a proportional score was calculated using the following equation: $\frac{Number\ of\ advantageous\ choices - Number\ of\ disadvatageous\ choices}{50}$, with a range between -1 and 1, where -1 means all 50 cards chosen were disadvantageous; 1 means all 50 cards chosen were advantageous, and 0 means that 25 cards chosen were advantageous and 25 cards chosen were disadvantageous. **Shift**: raw efficiency scores were calculated using the following equation: $\frac{Number\ correct - Number\ of\ errors}{Response\ time\ (in\ seconds)}$. These data were high in skewness and kurtosis so the square root transformation was applied. In order to use this transformation all scores had to be positive; 'plus one' was added to every score. **Inhibition**: the raw efficiency scores were calculated using the equation above, and were also found to be high for skewness and kurtosis, so as above, the scores were made positive, and the square root transformation was applied.

The data was then screened for normality to ensure the skewness and kurtosis ratings were as close to <1.5 (Osborne, 2002) as possible. Table 7 shows that the kurtosis levels for inductive reasoning and Tower of London were slightly above the recommended <1.5, but these could not be improved through transformations. Inhibition was slightly above, and Shift was much higher, in spite of the square root transformation used.

Table 7. Descriptive data for all tasks used in Study 5

Measure	Mean	S.D.	Number	Transformation used	Skewness (S.E.), post transformation	Kurtosis (S.E.), post transformation
Inductive Reasoning (out of 100)	51.04	42.52	56	None	-.009 (.319)	1.838 (.628)
Digit Span Forwards (out of 36)	15.46	4.01	56	Into one combined z-score	-.048 (.319)	.066 (.628)
Backwards (out of 30)	6.07	2.95				
Tower of London (out of 15)	10.25	2.03	56	None	-.899 (.319)	1.664 (.628)
Children's Gambling (high = 50, low = -50)	8.11	16.56	56	Into proportional scores	-.228 (.319)	.016 (.628)
Shape School Inhibition (efficiency score)	0.93	1.69	56	Sqrt	1.167 (.319)	1.567 (.628)
Shape School Shift (efficiency score)	0.16	0.21	55	Sqrt	.741 (.322)	3.166 (.634)

Statistical analyses: Correlation

A Pearson correlation matrix was performed to explore the strength of the relationship between the inductive reasoning task (the dependant variable), the EF tasks and age (the covariates), see Table 8.

The correlation figures in black show no correlation; the figures in blue show small correlations (r = .10 to .29); the figures in green show medium correlations (r = .30 to .49), and the figures in red show large correlations (r = .50 to 1.0; Pallant, 2007). These helped to decide which order the EF covariates would be placed in the regression procedure.

Table 8. A Pearson's correlation matrix showing the colour coded degree of correlation and significance levels at * = p < .05, and ** = p < .001 – Study 5.

	1. Inductive Reasoning[1]	2. Age	3. Digit Span	4. Tower of London	5. Children's Gambling	6. Inhibition	7. Shift
1.	1.00	0.29*	0.25	0.26	-0.05	0.09	-0.06
2.		1.00	0.56**	0.32*	-0.16	-0.06	-0.32*
3.			1.00	0.37*	-0.01	0.02	-0.27*
4.				1.00	0.07	0.15	-0.05
5.					1.00	0.93**	0.01
6.						1.00	-0.04
7.							1.00

Inductive reasoning and age

A significant small correlation was observed between inductive reasoning and age, suggesting that the variance in inductive reasoning could be partly explained by age effects: the older the children, the greater number of category choices made.

Inductive reasoning and executive functions

There were no significant correlations between the EF tasks, and number of category choices made during the induction task. None of the EF tasks could successfully explain the variance in inductive reasoning. However, there were two small correlations very close to significance with inductive reasoning (correlations shown in Table 8; Digit Span, *p* = .066, and the Tower of London, *p* = .055).

---

[1] Due to the unusual distribution of inductive reasoning scores, I changed this dependent variable from continuous to a) dichotomous and b) tricotomous; neither made a difference to the normality or the correlation.

Age and executive functions

Three out of five EFs were significantly correlated with age: a strong correlation was found with Digit Span; a medium negative correlation was found with Shift, and a medium correlation was found with the Tower of London. There was no significant correlation with Children's Gambling or with Inhibition.

Executive functions

There were three significant correlations between the various EF tasks: a very strong correlation was found between Children's Gambling and Inhibition; a medium correlation was found between Digit Span and Tower of London, and a small negative correlation was found between Digit Span and Shift suggesting that as children's Digit Span improved, their scores on the Shift task decreased. All other correlations were non-significant. The strong correlation between Children's Gambling and Inhibition was so high – 0.93 – that it suggests that these two measures were tapping into the same cognitive skill. It could be that inhibition plays a large part in the Children's Gambling task; children have to inhibit the high reward but high loss cards (disadvantageous deck) and turn their attention to the low reward but low cost cards (advantageous deck) to benefit overall.

Unusual results

It seemed strange that Shift was negatively correlated with five of the six other variables, two of which were significant, and one of these was age. I thought that this unusual result may be due to the high kurtosis level observed for the Shift measure, so I carried out the more conservative Spearman's Rho correlation. Other than Shift, all the variable correlations showed the same pattern as with the Pearson's correlation, although generally the significance was slightly reduced. However, I found that using Spearman's Rho meant that none of the correlations with Shift were now significant.

I also found it unusual that, although non-significant, there was a small negative correlation between Children's Gambling and age. Kerr and Zelazo (2004) had previously shown children to increase their score with age; my findings, although non-significant, were showing the opposite. I decided to investigate further. I broke down the task into five blocks of ten card choices rather than one block of 50 card choices. I wondered whether the first block – where children were starting to explore the cards – may have affected the correlation overall. I ran the correlation again, replacing the overall score with only the last block of ten cards; children's overall preference would be evident at this stage, if they had one. The correlations showed the same pattern. At the end of the Children's Gambling task, I had asked each child if they had a favourite deck of cards and if so, what it was they liked about that deck. Reception and Year 1 made comments consistent with their preference, for example a Year 1 child who favoured the disadvantageous deck commented that "I like spots [disadvantageous deck] 'cos you win more happy stickers". However, I became aware that the common Year 2 response was different; their preference was not consistent with the deck they thought to be advantageous: "Stripes [advantageous deck] were better as you lost less, but I kept choosing spots [disadvantageous deck] 'cos you never know if you might start winning more and losing less on them". Contradictory game playing was only observed for children in Year 2. Interestingly, it was clear that the older children – Year 2s – understood the logic of the game more than the younger years (see Table 9), which made it even more surprising that they chose the disadvantageous deck more often, resulting in their reduced scores. Contradictory play of knowing the advantageous choice and choosing the disadvantageous choice in case it suddenly becomes advantageous, perhaps signifies the development of risky behaviour: "…engagement in behaviors that are associated with some probability of undesirable results" (Boyer, 2006; p. 291).

Table 9. Overall mean scores across the ages on the Study 5 Children's Gambling task, and the percentage of responses made. Correct = they could identify the advantageous deck; Incorrect = they could not identify the advantageous deck, and preferred the disadvantageous deck; No preference = they thought both decks were equally good.

	Mean score	Correct response (%)	Incorrect response (%)	No preference response (%)
Reception	13.9	55	15	30
Year 1	13.9	60	5	35
Year 2	11.63	68	11	21

Partial correlation was used to explore the relationship between inductive reasoning and the EF tasks when age is controlled for. This allowed me to investigate the unique contribution of these variables without age-effects. It is clear that age has an effect on the strength of all correlations between inductive reasoning and EF variables (see Table 10); Digit Span, Tower of London, and Children's Gambling correlations are smaller and still non-significant. The Inhibition and Switch correlations increased slightly, but are still small and non-significant.

Table 10. Study 5 EF correlations with inductive reasoning with age, and with age factored out. None of the correlations were significant.

	Digit Span	Tower of London	Children's Gambling	Inhibition	Shift
With Age	.248	.258	-.048	.085	-.064
Without Age	.110	.184	-.002	.108	.028

Regression analyses

Previous literature has suggested that category induction preference develops with age because of an accumulation of knowledge (Gentner, 1988; Sloutsky et al., 2007), and age was the only factor that showed a correlation with inductive reasoning in the correlation matrix. Therefore age will act as potentially the greatest predictor of variance between inductive reasoning scores. Inhibition and working memory (Digit Span) may play a part in children's ability to reason with complex situations (Bunge & Zelazo, 2006; Halford, 1993; Halford et al., 1998; Zelazo & Frye, 1998; Zelazo & Müller, 2002); Digit Span showed a small correlation with Inductive reasoning, although non-significant. Interestingly, Tower of London was also found to have a small, non-significant, correlation with Inductive reasoning. Therefore, these three will be the next predictor variables. Children's Gambling and Shift (once Spearman's Rho was used) showed no correlation and no significance with inductive reasoning, so shall be the final predictor variables in the regression.

A hierarchical multiple regression was used to assess the ability of 5 measures (Tower of London, Digit Span, Inhibition, Shift, Children's Gambling) to predict inductive reasoning, and more specifically category induction, after controlling for age. Age was entered first, explaining .08 (R square, 8%) of the variance of inductive reasoning. Digit Span, Tower of London and Inhibition were entered second, explaining an additional .04 (R square change, 4%) of the variance of inductive reasoning. Shift and Children's Gambling were entered third, explaining an additional .08 (R square change, 8%) of the variance of inductive reasoning. The total variance explained by the model as a whole was .20 (20%) of the variance, *F* (6, 48) = 2.03; *p* = .08. Age explained 8%, *F* (1, 53) = 4.72; *p* = .034, with a standardised coefficient beta of .29 (see Table 11). As this result was unexpected, various statistical checks were carried out[2], but the outcome was the same.

---

[2] I checked to see whether altering the model order would make a difference, by putting step 3 in second, and step 2 in third. This change made no difference and both still came out non-significant. I ran the variables as a stepwise regression and the only significant variable was age. When age wasn't included in the stepwise regression, nothing came out as a significant predictor of inductive reasoning, suggesting that all other variables have age-related effects. I made the inductive reasoning dichotomous and tricotomous and ran a logistic regression, neither made any difference and age was always the sole predictor.

Table 11. Hierarchical regression output from Study 5 showing the variance explained at each of the three steps (models).

Step 1	**R²**	**β**	**R² change**	**F change**	**Significance**
	.08		.08	4.72	.03
Age		.29			.03
Step 2					
	.12		.04	.77	.52
Age		.20			.23
Inhibition		.07			.59
Digit Span		.08			.64
Tower of London		.16			.29
Step 3					
	.20		.08	2.42	.10
Age		.11			.51
Inhibition		.86			.03
Digit Span		.14			.40
Tower of London		.10			.49
Children’s Gambling		-.84			.03
Shift		.05			.74

I was surprised to see that Inhibition and Children’s Gambling had significant standardised coefficient beta levels in step 3. Although step 3 as a model was a non-significant predictor of variance, I decided to check whether Inhibition and Children’s Gambling came out as significant predictors of inductive reasoning if all other variables were excluded. They did not: Inhibition alone *F* (1, 54) = .39; *p* = .533, explaining 0.7% of the variance of induction; Children’s Gambling alone *F* (1, 54) = .122; *p* = .73, explaining 0.2% of the variance. Therefore, these two variables showing significant beta levels must be due to a shift of balance across the measures, which can alter depending on the other measures placed in the model. They are not significant predictors of inductive reasoning variance.

2 General discussion of Study 5

I tested children aged 4-7 years old on an inductive reasoning task and a battery of EF tasks. I examined whether age could significantly predict the variance of inductive reasoning and EF. Then, once age was factored out, I assessed whether EF could be the driving mechanism behind the development of inductive reasoning from a perceptual-bias to a category-bias. I also investigated whether the EF tasks were unitary or nonunitary.

Study 5 supports my previous studies and previous research (Gentner, 1988; Gentner & Toupin, 1986; Sloutsky et al., 2007) because once again, a gradual transition from perceptual-based to category-based induction was observed. The younger children showed a preference for perceptual induction; the older children showed a preference for category induction. This study has extended these findings and demonstrated that this transition *is* significantly predicted by age. However, this study is unable to support previous work suggesting that young children's reasoning is driven by their cognitive control, mainly inhibition and working memory (Halford et al., 1998; Richland et al., 2006; Zelazo & Frye, 1998) as none of the EFs could predict inductive reasoning variance. It is also interesting to note that the majority of the EF tasks were not correlated.

2.1 Age as an influencing factor

My first prediction was that age would be a significant predictor of inductive reasoning, with older children making more category choices, and of the performance on the EF tasks, with older children scoring higher than younger children. The findings show that age *is* the most influencing factor I assessed regarding inductive reasoning. As children aged, they were more likely to show a higher percentage of category-based decisions. This provides support for my previous studies and the perceptual-bias account, which suggests that young children develop the use of category induction through time, experience and an accumulation of knowledge (Gentner, 1988; Gentner & Toupin, 1986; Inhelder & Piaget, 1958; Sloutsky et al., 2007). As I have found in previous studies, young children did not use category induction even though they were capable of completing the initial and final categorisation tasks. However, although age was correlated with all but one of the EF tasks, three were negatively correlated, which I did not predict.

The EF tasks used are standardised, and have previously been used to demonstrate age differences on these EF skills: children's performance on the specific EF tasks improve as they get older. Shift (on the Spearman's Rho correlation matrix) and Inhibition showed a small negative correlation with age. I can speculate that this is most likely to be due to the slow development of these EFs. The age range was too narrow to see a full developmental effect of Inhibition or Shift. I address this point in the future work section of Chapter 6. The Children's Gambling task showed the greatest negative correlation (although non-significant) with age. Children's responses were examined in greater detail to try and understand this unexpected finding. The explanations children gave for their own decisions suggested that although the older children were understanding the logic behind the game more than the younger children, and knew which deck of cards was the advantageous option, their risk-taking was leading them go for the disadvantageous deck of cards. They were making this choice in case their luck changed and it became the advantageous deck of cards: they risked the worse option in the hope of a greater reward. Thus signifying risky behaviour.

Although young children's decision-making and risk-taking has been investigated, the developmental trajectory is still unclear. Early theorists claimed that children are not cognitively mature enough to consider the effects of game probability until age 12 years (Piaget & Inhelder, 1951). This conflicts with more recent research that suggests children as young as 4 years old are capable of making decisions based on expected outcomes (Schlottmann, 2001), which may be interpreted as an early appreciation of risk assessment. The original Children's Gambling task was only administered to children aged 3-4 years old (Kerr & Zelazo, 2004), but they found a positive correlation with age, suggesting that as children aged, they were more likely to choose the advantageous option. Although I found this pattern for these ages, the risk-taking behaviour observed in my experiment was only in Year 2 – age 6-7 – which cannot be compared with Kerr and Zelazo's findings. However, Garon and Moore (2004) used a version of the gambling task with children aged 3-6 years and found that none of the years showed a significant preference – playing at chance – but the 6 year olds showed more understanding of the game. It seems to be agreed that as children grow up, they become more aware of the purpose of decision-making and gambling tasks, however, the age at which children start making risky decisions is still unknown, and would be interesting to investigate further. I address this in the future work section of Chapter 6.

2.2 Inductive Reasoning and Executive Function

My second prediction was that there would be a connection between the use of category induction and the performance on the EF tasks, with those children who perform better on the EF tasks, showing a greater trend towards category induction on the inductive reasoning task. Previous research has suggested that the main limiting factor for young children attempting complex tasks, or tasks requiring the processing of multiple factors, is the development of cognitive control (Bunge & Zelazo, 2006; Zelazo & Frye, 1998; Zelazo & Müller, 2002); more specifically, the development of EF tasks such as working memory (Andrews & Halford, 2002; Halford, 1993) and inhibition (Richland et al., 2006). It has been proposed that until these cognitive functions are better developed, children will continue to use simpler, more obvious solutions such as perceptual cues and item similarity (DeCaro et al., 2008; Juslin et al., 2003; von Helversen et al., 2010). My experimental findings support the theory that inductive reasoning improves as children's cognition becomes more sophisticated, however, they are unable to provide support that inductive reasoning strategy alters with the development of cognitive control and EFs. I found no significant predictor of inductive reasoning from any of the EF tasks, which does not support my prediction.

However, my correlations were two-tailed (as the study was exploratory), which is a more conservative measure, and only highlighted a correlational trend between inductive reasoning and Digit Span, and inductive reasoning and Tower of London. If I had based this work solely on previous predictions made in the literature – the relational complexity theory and the cognitive complexity and control theory – I would have used a one-tailed test, which is less conservative, and I would have observed significant correlations for inductive reasoning and Digit span (working memory): $r = .248$, $p = .033$; inductive reasoning and Tower of London: $r = .258$, $p = .027$. Although the correlations would have become significant, this would not change my conclusions since the regression demonstrated that these variables did not significantly predict the variance of inductive reasoning. These very small effects could be due to a limited sample size of 56 children. It is possible that I am experiencing a type-2 error whereby there are influencing factors, but the small sample size means the effect is not being shown. I investigated the power of my sample through G*Power (a power analysis tool). I used a linear multiple regression: fixed model R^2 increase statistical test. The power was .81 for a medium effect (defined as $f^2 = .15$), for 56 participants. However, my highest observed effect size was age, at .09, which gave me a power of .60. Clearly I did not have strong enough power to detect any effects of this size with only 56 participants. With an effect size of .09, I would need roughly 100 participants to gain a power rating of above .80; .80 is the recommended power for detecting relationships between variables (Pallant, 2007). This suggests that the power was not sufficient to detect small effects, and some effects may have become significant with a larger sample. I return to this issue in the future work section of Chapter 6.

2.3 EF tasks: unitary or nonunitary

I was interested to see whether the EF tasks are unitary, or whether they have overlapping measures. My final prediction was that all of the cool EF tasks would be strongly correlated as they tap into the same cognitive mechanism, the DL-PFC. However, the cool EF tasks would not be correlated with the hot EF task as it taps into a different cognitive mechanism, the VM-PFC. In fact very few of the EF tasks were significantly correlated, which does not support my prediction. In complete contrast to my prediction, the greatest correlation was observed between the cool EF Inhibition, and the hot EF Children's Gambling. As I speculated earlier, it seems likely that these two measures are tapping into the same skill (correlation of .9). It is possible that a high level of inhibition is required during the Children's Gambling task; children must inhibit choosing the high reward-high loss cards, for the low reward-low loss cards to gain more rewards overall. The other two significant correlations were between Digit Span and Tower of London, and a negative correlation between Digit Span and Shift. The correlation between Digit Span and Tower of London suggests that working memory is required during the Tower of London task. It is possible that when children are working out the quickest strategy for completing the Tower of London tasks, they must hold this information in their memory until they have completed the task. As their working memory improves, so does their correct completion of the Tower of London tasks. The negative correlation between Digit Span and

Shift suggests that as working memory increases, Shifting decreases. This is unusual, and based on my findings, I am unable to provide an explanation for this result.

It has been suggested that 'pure' EF tasks – a task that only measures what it is supposed to – are rare, and often working memory plays a role in other simple EF tasks. For example, a high level of working memory is required during my Inhibition and Shift tasks, but the tasks claim only to measure inhibition or shift, and not working memory. However, based on my findings of very few correlations, I must conclude that overall EF tasks are unitary. There are some overlaps, but these could be due to the tasks not being 'pure' rather than the specific EFs overlapping themselves. Other factors could also be influencing results, for example differences in participants' language or visuospatial skills (Miyake et al., 2000).

2.4 Conclusion

Chapter 4 has presented age as a significant predictor of the variance of inductive reasoning, and shown that none of the EFs had any significant influence on induction. Finding that none of the EF measures were significantly correlated to inductive reasoning was surprising, so this result was checked with a multitude of statistical analyses, and all confirmed the finding. Finding that age is a significant predictor of induction is not surprising, as all of my previous studies have shown a gradual transition from perceptual to category induction with age. The influence of age on induction development has also been suggested by previous researchers (Gentner, 1988; Sloutsky et al., 2007). However, this is the first time that this relationship has been shown statistically. Considering the multitude of previous work into the influence of EF on reasoning (Bunge & Zelazo, 2006; Halford, 1993; Halford et al., 1998; Richland et al., 2006; Zelazo & Frye, 1998; Zelazo & Müller, 2002), it is surprising that none of the EFs could significantly predict the variance of induction. However, my findings showed that working memory (Digit Span) and planning (Tower of London) were close to a significant correlation with inductive reasoning, which suggests that they may support induction. It is clear though that EFs on the whole do not drive the developmental trajectory. These findings suggest that it is highly unlikely that young children's natural bias is held back by performance factors (Gelman, 2003; Goswami, 1992, 2001). Instead, it is more likely that the transition from a perceptual-bias to a category-bias with age that I observed in Study 1, was genuine. However, it is still unknown whether the transition is due to a cognitive maturation whereby children must go through fixed developmental stages (Piaget, 1964), or whether the transition is caused by an accumulation of knowledge and experience, the effect of which is not limited by a cognitive maturation (Gentner, 1988; Sloutksy et al., 2007).

Chapter 5: Can interactive educational sessions with animals boost young children's inductive reasoning?

In Chapters 2 and 3, I tested children aged 3-9 years old. I explored how methodological manipulations can affect the developmental trajectory of induction, and may have caused the contradictory findings supporting each of the two broad accounts of induction. In Chapter 2, I discovered a gradual transition from a perceptual strategy preference to a category strategy preference that could not be due to maturity groupings or the inability to inhibit highly perceptual distracters – both of which had previously been suggested by researchers in support of the category-bias account. In Chapter 3, I found that when children were presented with more complex relational category structures and rules, their use of category induction was delayed by a year. This pattern appears to be domain-general. I also found that labels during the induction trial significantly increased the use of category induction. In Chapter 4, I turned my attention to the development of inductive reasoning, and possible mechanisms affecting the process. I investigated the relationship between inductive reasoning and EF by asking children to complete an inductive reasoning task as well as a battery of EF tasks. Although I found that age could significantly predict some of the variance of induction, I found that when age was factored out, none of the EF tasks were significant predictors. This led me to conclude that although EF may support inductive reasoning, the development of EF is unlikely to be the driving force behind the transition from perceptual to category induction.

In this chapter I concentrate on whether the use of category induction can be enhanced through an accumulation of domain knowledge, or whether the development is fixed due to a stage-like cognitive maturation process. My findings so far suggest that the transition from perceptual to category induction is caused by an accumulation of relevant domain knowledge with age. If this is the case, young children will be able to correctly differentiate between categories of items, but only with time and a greater appreciation of the importance of category membership (through an accumulation of knowledge) will they begin to switch their focus from perceptual cues to relational cues. However, although this currently seems like the most plausible explanation, the transition could be due to another type of cognitive maturation besides EF. I designed Study 6 to explore whether an interactive educational session could boost young children's category induction due to an increase in domain knowledge. If the transition is caused by a gradual accumulation of knowledge, then I would expect to see that children taking part in an interactive educational session would show an increase in category induction. Whereas, if the transition is caused by a cognitive maturation, then I would expect that although the educational session would increase children's knowledge, their induction would not alter as they would be restricted by the stage of their cognitive maturation.

1 Study 6: The influence of interactive educational zoo sessions

In the introduction I discussed that although it has been found that children are able to learn from engaging media such as television (Troseth et al., 2006), they appear to retain the information better through direct experience (DeLoache & Chiong, 2009; Strouse & Troseth, 2008). Current research has encouraged learning through interaction, and has shown this technique to improve children's ability to absorb, understand and apply newly taught information to different situations (Au et al., 2008; Ellefson et al., 2008; Ganea et al., 2010). The evidence suggests that it should be possible to enhance inductive reasoning: if children are able to take newly learned information and apply it to different situations, I hypothesised that it would be possible to boost children's category induction through interactive educational sessions targeting the relevant domain. In my study, I focused on whether interactive educational sessions with animals boosted category induction for the whole domain.

Study 6 has two aims, firstly to explore whether category induction preference *can* be enhanced through educational sessions and increased knowledge (Gentner, 1988; Sloutsky et al., 2007) or whether a stage-like transition prevents cognitive maturation (Piaget, 1962). Secondly, if category induction can be enhanced, does it affect the whole domain (see Gentner, 1988) or only the specific items trained within the domain (see Goswami, 1992)? To investigate these questions, I administered four inductive reasoning tasks to every child and compared the performance of children in an intervention condition, who also participated in an interactive educational session at Twycross Zoo, and children in a comparison condition. Children in both conditions took part in all four tasks, two before and two a week after the zoo visits. To test whether domain-specific training would enhance category induction for the whole domain or only for those items seen during the educational sessions, two of the tasks used images of the types of animals seen in the educational session, and two used novel, biologically plausible images.

If children from the intervention condition show an increase in category choices during the second testing session, after their interactive educational session at the zoo, then this provides support for a more flexible cognitive development, whereby increased knowledge and experience are needed to alter children's outlook on reasoning (Gentner, 1988; Sloutsky et al., 2007). However, if children from the intervention condition show no significant improvement during induction, then this will lend support to a more rigid, stage-like cognitive maturation (Piaget, 1964). The category-bias account would predict that if one observes an increase in category choices after the education session, it would only affect the specific items seen during the session (Goswami, 1992), and not the novel items. However, the perceptual-bias account would predict that if the educational session does improve induction, it would affect the biological domain as a whole, rather than specific items, and one would see an increase of category choices with familiar and novel test items.

1.1 Method

1.1.1 Pre-study participants

Adults:

Fifteen adults (mean age = 19.5 years, range 18-21; 2 males and 13 females) participated in the first stimuli similarity pre-test and 26 adults (mean age = 19.7 years, range 18-25; 4 males and 22 females) participated in the second stimuli similarity pre-test following image adjustments. Twenty-five adults (mean age = 19.2 years, range 18-25; 5 males and 20 females) participated in a domain identification and stimulus naming pre-test, and 24 adults (mean age 19.2 years, range 18-24; 8 males and 16 females) participated in the pre-test of the induction tasks.

Children:

Thirty-one primary school children (11 Reception, 10 Year 1 and 10 Year 2; range 4.02-7.01; 15 males and 16 females) participated in a stimuli similarity pre-test and 32 primary school children

(12 Reception, 10 Year 1 and 10 Year 2; range 4.02-7.00; 16 males and 16 females) participated in a domain identification and stimulus naming pre-test.

1.1.2 Main study participants

Two hundred and fifty-two primary school children participated: 126 Year 1 (5-6 years) and 126 Year 2 (6-7 years); range 5.07-7.09 years; 124 males and 128 females. One hundred and twenty-three children participated in the comparison condition, 129 children participated in the intervention condition. Both conditions completed all four tasks however the intervention group also took part in an interactive educational session at Twycross Zoo. The schools were matched for their overall performance in Mathematics, Science and English, the percentage of EAL (English as an Additional Language), SEN (Special Educational Needs), pupil absence, free school dinners (as an indication of socio-economic status), and the Ofsted report and grading (see Appendix E).

1.1.3 Stimuli

Computer aided designs (CAD, designed using GMax or Microsoft Word) of 392 images were used, of which 196 were taken from Study 1 (46 juvenile bugs, 58 adult bugs and 92 transitional bugs) creating Sandbugs and Rockbugs, and 196 were taken and adapted from Study 2 (46 juvenile bugs, 58 adult bugs and 92 transitional bugs) creating Ground weevils and Forest weevils. Three hundred real images were used (92 juveniles, 92 transitional images and 116 adults) with equal numbers of Chinchillas, Squirrels, Tree frogs and Common frogs; see Appendix A for examples of all categories of stimuli. Due to difficulties finding suitably matched transitional images – to ensure the transition looked realistic – the real stimuli transitions had one image which increased in size rather than two separate images, as with the CAD images. I felt it more important to make the transitions look realistic, than to have four separate images overall to match the novel kinds. All transitions, real and novel, were highly realistic and convincing, see Figures 20 and 21.

Figure 20. Transformation from a juvenile Forest weevil to an adult Forest weevil used in Study 6.

Figure 21. Transformation from a juvenile Common frog to an adult Common frog used in Study 6.

1.1.4 Pre-tests

Stimuli similarity:

During each of the four induction tasks in the main experiment, children saw 12 triads consisting of an adult target, a same-category but perceptually dissimilar juvenile, and a different-category but perceptually similar adult (see Figures 22 and 23).

Figure 22. An example of a novel natural kind induction triad used in Study 6. The juvenile is transformed into an adult target (transformation), then the adult target is shown with a category choice and a perceptual choice (induction triad).

Figure 23. An example of a real natural kind induction triad used in Study 6. The juvenile is transformed into an adult target (transformation), then the adult target is shown with a category choice and a perceptual choice (induction triad).

Children saw all the triads for the 4 tasks, 12 in each task, so 48 triads overall. To make certain that the target item was perceptually more similar to the perceptual distracter test item than the same-category test item, the similarity between test and target items was validated using adult and child ratings. Children chose the most similar test item (perceptual choice or category choice) to the target on all 48 triads. Adults rated the perceptual similarity of each of the 96 pairs: there were 48 triads, each of which created two slides, one with the target and perceptual choice, and one with the target and category choice.

Adults Adults rated the 96 pairs for perceptual similarity on a 5-point Likert scale. The scores were then divided into eight category pairings: firstly split into the four tasks, then split into each task's triad pairings (each triad created two slides). For example, the Rockbug-Sandbug task would be split into Rockbug Target and Sandbug Perceptual choice pairs (first category pairing) and the Rockbug Target and Rockbug Category choice pairs (second category pairing). The scores from these would be compared to see which of the two test items (the perceptual choice or the category choice) were seen as more perceptually similar to the target. Paired sample t-tests were conducted and a bonferroni correction was set at $p < .025$ as each of the target images were used twice, once with the perceptual choice test item and once with the category choice test item.

In the first testing session, six out of eight category pairings were significantly different, however the 'Squirrel Target – Chinchilla Perceptual' was not significantly different from the 'Squirrel Target – Squirrel Category': t (14) = -1.07; p = .305, and the 'Chinchilla Target – Squirrel Perceptual' was not significantly different from the 'Chinchilla Target –Chinchilla Category': t (14) = .96; p = .353.

Some of the real image perceptual distracters (chinchillas and squirrels) were not showing enough perceptual similarity to the target. Therefore, the perceptual distracters that were rated as less perceptually similar to the target than the category choice were substituted for new images. The substitutions were successful and adults rated items designed to be higher in perceptual similarity to the target – the perceptual choice – as perceptually more similar to the target than the category choice, see Table 12.

Table 12. Statistics confirming that adults were pairing items designed to be higher in perceptual similarity for all stimuli used in Study 6.

Target - Perceptual	*Mean score (out of 5)*		*Target - Category*	*Mean score (out of 5)*	*Paired-t-test*
Squirrel - Chinchilla	2.88	vs.	Squirrel - Squirrel	1.90	*t* (25) = -6.39; *p* < .001
Chinchilla - Squirrel	2.76	vs.	Chinchilla - Chinchilla	2.05	*t* (25) = -5.35; *p* < .001
Tree frog - Common frog	3.15	vs.	Tree frog - Tree frog	2.36	*t* (25) = -5.25; *p* < .001
Common frog - Tree frog	3.16	vs.	Common frog - Common frog	1.99	*t* (25) = -6.79; *p* < .001
Rockbug - Sandbug	4.33	vs.	Rockbug - Rockbug	1.66	*t* (25) = -14.85; *p* < .001
Sandbug - Rockbug	4.32	vs.	Sandbug - Sandbug	1.56	*t* (25) = -14.10; *p* < .001
Forest weevil - Ground weevil	3.61	vs.	Forest weevil – Forest weevil	2.29	*t* (25) = -7.12; *p* < .001
Ground weevil - Forest weevil	3.63	vs.	Ground weevil - Ground weevil	2.24	*t* (25) = -8.11; *p* < .001

Children The use of a Likert scale was considered too complex for young children, therefore a simpler technique was used. On the 48 triads (12 from each task), children had to choose which test item looked most like the target item. One-sample t-tests confirmed that in all triads the perceptual choice distracter was seen as more perceptually similar to the target than the category choice, see Table 13.

Table 13. Statistics confirming that children were pairing items designed to be higher in perceptual similarity for all stimuli used in Study 6. The t-values show that the target and the perceptual choice were paired frequently, which was significant.

Triad	Perceptual choice was chosen (%)	Statistics
Squirrel as target	80%	*t* (30) = -161.49; *p* < .001
Chinchilla as target	78%	*t* (30) = -148.03; *p* < .001
Common frog as target	77%	*t* (30) = -149.75; *p* < .001
Tree frog as target	77%	*t* (30) = -148.03; *p* < .001
Forest weevil as target	95%	*t* (30) = -218.00; *p* < .001
Ground weevil as target	95%	*t* (30) = -206.67; *p* < .001
Rockbug as target	97%	*t* (30) = -409.37; *p* < .001
Sandbug as target	95%	*t* (30) = -225.15; *p* < .001

Domain identification and stimulus naming:

Adults A domain categorisation pre-test was conducted to check the stimuli were considered to be biologically plausible. Adults were shown 32 images individually (2 juveniles and 2 adults of each category) and asked to state whether the item was a living or a non-living thing, and what they believed it to be.

- Adult and juvenile Chinchillas were labelled as living kinds 100% of the time, with common responses being ‘chinchilla’ and ‘mouse’.
- Adult and juvenile Squirrels were labelled as living kinds 100% of the time, with the only response being ‘squirrel’.
- Adult and juvenile Tree frogs and Common frogs were labelled as living kinds 100% of the time, with the only response being ‘frog’.

- The adult Forest weevils were labelled as living kinds 80% of the time and the adult Ground weevils were labelled as living kinds 82% of the time, with common responses for both being ‘scorpion’ and ‘ant’; juvenile Forest and Ground weevils were labelled as living kinds 78% of the time, with common responses being ‘bird’ and ‘mouse’.

- Adult Sandbugs and Rockbugs were labelled as living kinds 100% of the time, with common responses being ‘beetle’ and ‘bug’; juvenile Sandbugs were labelled as living 87% of the time, and juvenile Rockbugs were labelled as living 91% of the time, with common responses being ‘worm’ and ‘maggot’.

Children The same 32 images were shown.

- Adult Chinchillas were labelled as living kinds 94% of the time and the juvenile Chinchillas were labelled as living kinds 100% of the time, both with common responses being ‘mouse’ and ‘rat’.

- Adult and juvenile Squirrels were labelled as living kinds 97% of the time, both with the only responses being ‘squirrel’.

- Adult Tree frogs and Common frogs were labelled as living kinds 97% of the time, juvenile Common frogs were labelled as living kinds 97% of the time; juvenile Tree frogs were labelled as living kinds 100% of the time, all with the only responses being ‘frog’.

- Adult Forest and Ground weevils were labelled as living kinds 95% of the time, with common responses being ‘insect’ and ‘beetle’; juvenile Forest and Ground weevils were labelled as living kinds 94% of the time, with common responses being ‘ant’ and ‘worm’.

- Adult Sandbugs and Rockbugs were labelled as living kinds 100% of the time, with common responses being ‘spider’ and ‘beetle’; juvenile Sandbugs were labelled as living 78% of the time, and juvenile Rockbugs were labelled as living 62% of the time, with common responses for both being ‘fish’ and ‘worm’.

Induction task:

Adults completed the induction tasks to ensure they were easy to understand and complete before they were given to children. Adults scored highly and above chance on the initial and final categorisation tasks: Chinchilla-Squirrel Initial *M* = 95% (*SD* = 6.00), *t* (23) = 36.91; *p* < .001; Final *M* = 99% (*SD* = 3.05), *t* (23) = 78.28; *p* < .001. Tree frog-Common frog Initial *M* = 99% (*SD* = 3.05), *t* (23) = 78.28; *p* < .001; Final *M* = 96% (*SD* = 6.58), *t* (23) = 34.38; *p* < .001. Sandbug-Rockbug and Ground weevil-Forest weevil were all at 100%. Adults made significantly more category choices than expected by chance in all tasks: Chinchilla-Squirrel *M* = 97% (*SD* = 6.44), *t* (23) = 35.44; *p* < .001; Tree frog-Common frog *M* = 97% (*SD* = 11.91), *t* (23) = 19.29; *p* < .001; Sandbug-Rockbug *M* = 95% (*SD* = 20.54), *t* (23) = 10.68; *p* < .001; Ground weevil-Forest weevil *M* = 96% (*SD* = 18.78), *t* (23) = 12.04; *p* < .001.

1.2 Design and Procedure

There were two between-subject factors: year (with 2 levels: Years 1 and 2) and condition (with 2 levels: intervention and comparison): no child took part in both conditions or was in more than one year group. Children in the intervention condition were from schools that had already contacted Twycross Zoo for an educational session; children in the comparison condition were from schools closely matched to the intervention schools. There was one within-subject factor: stimuli (with 2 levels: novel vs. real). The dependent variable was the number of correctly categorised items during the initial and final categorisation tasks, and the number of perceptual or category choices made during the induction tasks. The breakdown of participant allocation is detailed in Table 14. The 129 children participating in the intervention condition also took part in an interactive educational session at the zoo. Both conditions completed all four tasks: two a week before (one real and one novel) and two a week after the educational zoo session (one real and one novel). All tasks were counterbalanced.

Table 14. The split of children into each condition for Study 6.

Year	**Intervention**	**Comparison**	**Total**
1	67	59	**126**
2	62	64	**126**
Total	**129**	**123**	**252**

Children were taught a rule to categorise and differentiate between two item kinds – in this case either Rockbugs from Sandbugs (novel CAD images), Ground weevils from Forest weevils (novel CAD images), Chinchillas from Squirrels (real images) or Tree frogs from Common frogs (real images) – based on a feature:

> *"**Sandbugs** live in the sand, and have round heads for making soft burrows; **Rockbugs** live rocks, and have angled heads for digging; **Forest weevils** live in trees, and have long tails for good balance; **Ground weevils** live in the earth, and have short tails to fit into small holes; **Tree frogs** live in trees, and have rounded suckers on their feet for climbing; **Common frogs** live in ponds, and have straight webbed feet for swimming; **Squirrels** live in trees, and have long bushy tails for balancing, **Chinchillas** live in mountains, and are round and fluffy for keeping warm".*

The procedure followed the same four-fold methodology as previous studies (1) category learning, 2) initial categorisation, 3) induction task, 4) final categorisation; see Appendix F for all hidden properties), with 12 trials in each task, and four tasks in total.

1.2.1 Interactive educational sessions

Each interactive educational session could accommodate around 30 children, and was carried out in the educational centre at Twycross Zoo. The sessions were designed and presented by members of the educational team at Twycross Zoo, and lasted 30 minutes; at no point was I involved in the session, but I observed each one. The session was called 'Animal adaptations to habitats', and comprised children seeing and touching small animals, such as a tree frog and a chinchilla (see Figure 24). During the interaction, children had to answer questions and were taught about the importance of animal adaptations to their environments. For example, children were asked why they thought tree frogs needed suckers on their feet, the reason – in this case to climb up the trees to where they live – was emphasised and reiterated. The main focus of the educational session was to highlight the importance of each animal's features, and show that every feature has a purpose.

Figure 24. The animals seen during the Twycross Zoo interactive educational session – Study 6. From top left to bottom right: a White's Tree Frog, a Leopard gecko, a Madagascan hissing cockroach, a Chinchilla, and a Mexican red knee tarantula. (The images are presented with permission from photographer Deborah Bardowicks, Twycross Zoo).

1.3 Results

1.3.1 Categorisation Performance

Only children who performed significantly above chance in both initial and final categorisation tasks were included in the final sample. According to a binomial test, scores of 10/12 (83%) and above were significantly above chance (proportion = 0.5, *p* = .04). Being a repeated measures study, children had to pass the categorisation trials for all four tasks to be included. Seventeen children (12 Year 1, 6 intervention and 6 comparison; 5 Year 2, 2 intervention and 3 comparison) scored at or below chance and were removed. The remaining 235 children (114 comparison, 53 Year 1, 61 Year 2; 121 intervention, 61 Year 1, 60 Year 2) scored highly and above chance on the initial and final categorisation trials: **Real (first testing session):** Initial *M* = 96% (*SD* = 5.27), *t* (234) = 133.05; *p* < .001; Final *M* = 97% (*SD* = 5.08), *t* (234) = 141.52; *p* < .001. **Real (second testing session):** Initial *M* = 94% (*SD* = 6.09), *t* (234) = 111.58; *p* < 001; Final *M* = 95% (*SD* = 5.60), *t* (234) = 124.19; *p* < .001. **Novel (first testing session):** Initial *M* = 98% (*SD* = 4.02), *t* (243) = 184.09; *p* < 001; Final *M* = 98% (*SD* = 4.08), *t* (234) = 181.33; *p* < .001. **Novel (second testing session):** Initial *M* = 98% (*SD* = 4.38), *t* (243) = 167.47; *p* < 001; Final *M* = 98% (*SD* = 3.91), *t* (234) = 189.16; *p* < .001.

1.3.2 Induction Performance

The percentage of category choices made by each participant in the induction task was examined, with participants making at least 10/12 category choices were considered to have a category bias (binominal test proportion 0.5, *p* = .04). Participants who made less than 2/12 category choices were considered to have a perceptual-bias (binominal test proportion 0.5, *p* = .04).

A mixed GLM was conducted with two within-subjects factors: session (with two levels: first testing session vs. second testing session) and stimuli type (with two levels: real vs. novel), and two between-subjects factors: year group (with two levels: 1 vs. 2), and condition (with two levels: intervention vs. comparison).

1.3.3 Main effects

There was no overall difference in inductive strategy preference between the intervention and comparison groups: *F* (1, 231) = 2.02; *p* < .156, showing that the groups were well matched. Children in Year 2 made significantly more category induction choices than Year 1: *F* (1, 231) = 16.63; *p* < .001 (see Figure 25), and a category bias presented more with real kinds than novel kinds *F* (1, 231) = 24.20; *p* < .001. Overall, more category choices were made during the second testing session compared to the first testing session: *F* (1, 231) = 22.84; *p* < .001, which suggests practice effects.

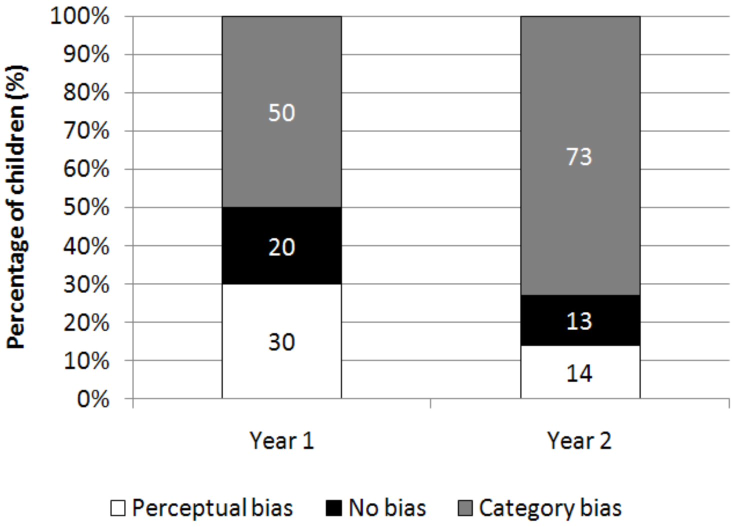


Figure 25. The percentage of children showing each bias in Study 6. This is collapsed across conditions and sessions.

1.3.4 Interactions

There was a significant interaction between condition and session: *F* (1, 231) = 6.55; *p* = .011, with post-hoc t-tests showing that the increase in category choices made during the second testing session (after the educational session), compared to the first testing session, was only significant for children in the intervention condition: intervention condition *t* (120) = -5.22; *p* < .001; comparison condition *t* (113) = -1.67; *p* = .10, (see Figure 26). All other interactions were non-significant. In particular, there was no three-way interaction between condition, session and stimuli type, suggesting that the increase in number of category choices made during the intervention condition's second testing session was similar for the real and novel kinds: *F* (1, 231) = 2.28; *p* = .133. I wanted to check that I had enough power to observe this interaction, if there had been one, so I conducted a post-hoc power analysis using G*Power, ANOVA: repeated measures, within-between interaction statistical test. The recommended power for detecting relationships is .80 (Pallant, 2007), and the post-hoc power analysis showed that this interaction had a power of .91; if there had been an interaction, it would have been observed. This suggests that the educational sessions affected the whole domain, not just the items experienced. There was no significant interaction between intervention, session and year, suggesting that both years were equally affected by the interactive educational session: *F* (1, 231) = .301; *p* = .584.

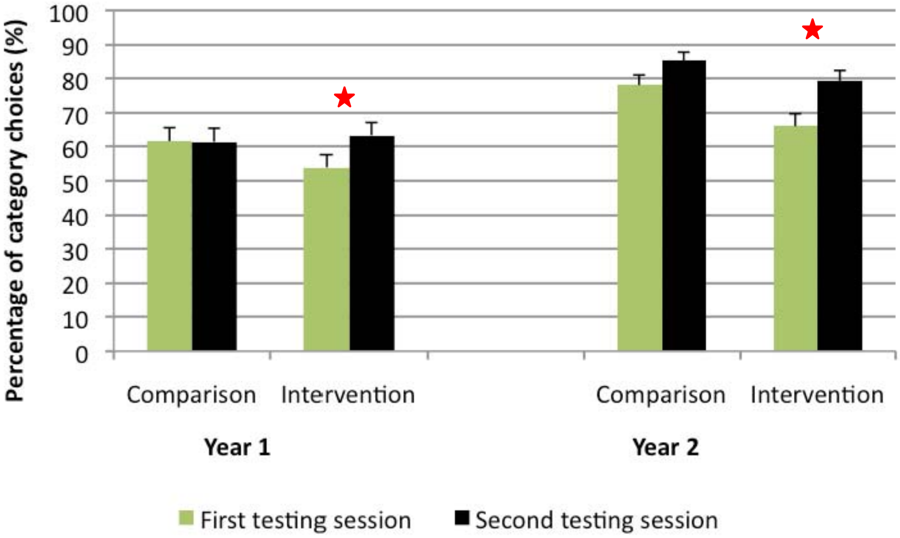


Figure 26. The mean percentage of category choices made by the Year 1s and Year 2s, split into the intervention and comparison conditions, and the first and second testing sessions – Study 6. This

figure is collapsed across the stimuli type as no significant difference was found between the choices made for real versus novel kinds. It also highlights the stable effect for year.

2 General Discussion of Study 6

In this chapter, I investigated the natural strategy bias of induction: whether the ability to use category induction can be improved, and whether experience affects reasoning for the domain as a whole or only the specific items experienced. I used the triad paradigm to disambiguate category-based and perceptual-based induction choices in all four induction tasks, two of which were of familiar animal types seen during the educational zoo session, and two which were novel, yet biologically plausible kinds for which children would have no prior knowledge. Children were either in an intervention condition, and visited Twycross Zoo for a 30-minute interactive educational session about animal adaptations to habitats, or children were in a comparison condition, and did not receive an educational session.

Year 1 children made fewer category choices overall compared to the Year 2 children, suggesting a transition from perceptual to category choice with age. This confirms my previous studies, and goes some way to supporting the theory that an early bias for perceptual, surface features is most likely (Gentner, 1988; Piaget, 1964; Sloutsky et al., 2007). The increase in category choices in the second testing session of the intervention condition highlights that experience can boost category induction, and it appears that experience with animals enhances category induction for the domain as a whole, not just the specific animals experienced.

2.1 Development of induction and the effect of experience

Children in Year 2 made more category choices during the induction tasks than children in Year 1, again supporting my previous studies that inductive preference gradually transitions from perceptual to category induction with age. Both Year 1 and Year 2 children in the intervention condition showed an increase in category choices during the second testing session, suggesting that this transition cannot be due to a fixed cognitive maturation. Strikingly, the observed increase of category induction for the intervention condition occurred after interactive educational sessions, where the focus was *not* on learning how to complete inductive reasoning tasks. This shows that the mechanism driving the development of category induction is very likely to be due to an accumulation of experience and knowledge for a domain. Children were not taught how to reason, nor were they trained to a specific task; their change in strategy preference was due to an increased understanding and appreciation of the importance of category membership. The educational session was designed by Twycross Zoo to highlight the importance of animals' adaptations to their environment. The educational team not only wanted to show children how to identify animals by subtle differences in their features, but also that these are important for their survival.

Although everyday interaction with animals – pets, for example – are good for informing children about categories, and provide a basic knowledge, an educational session like the one at Twycross Zoo adds depth to this knowledge. The educational session highlighted the significance and importance of categories and their features, and explained how vital the category features are for survival. The information showed children that subtle cues are just as important predictors of category, as the overall appearance. This level of informative knowledge is likely to be the reason for the increase in category choices during the second testing session. This approach to learning not only provided children with information, but also *explained* the importance and application of this information through real-life interactions. This level of information is generally not gained through interaction with pets or general visits to the zoo.

2.2 Improvements in the targeted domain

The experience and interaction with the animals boosted the transition from perceptual to category induction, at least for the targeted biological domain. Children were not trained to use category induction, neither were they given any education about the novel kinds. Yet strikingly, children still significantly increased the number of category choices during the unknown novel trials as well as the familiar real-kind trials.

I considered whether domain-specific experience would enhance category induction in the whole domain (Gentner, 1988) or only in the specific items experienced within the domain, akin to Goswami's prediction that once children understand the specific relations involved, they will be biased to use relational information (Goswami, 1992). My research directly pitted targeted real items – real kind item types seen at Twycross Zoo – with non-targeted novel items, and found that although the increase in category choices was slightly greater in the real kind targeted category, there was also a significant improvement in the novel non-targeted category. There was no significant difference between the increase of category choices made in the real and novel tasks by children in the intervention condition, during the second testing session. This suggests that when a child begins to understand the importance of category membership during induction, this strategy preference is applied to the whole domain, in this case, the biological domain. This directly conflicts with the hypothesis that young children are capable of applying relational knowledge as long as they understand the specific relational connections between items in a particular context (Goswami, 1992; 2001). If this hypothesis were correct, training would only improve reasoning on experienced items (i.e., the real kinds). In contrast, I found that children who had experience with the real kinds used in the Twycross Zoo educational sessions, transferred this understanding onto the real and novel kinds used in my induction tasks, and increasingly made category induction decisions for both stimuli types. Nevertheless, it is impossible to say whether this effect is domain-specific (just within the biological domain) or domain-general. The unresolved issue will be discussed in Chapter 6, when I outline my ideas for future work.

2.3 The developmental mechanism behind induction

My findings reported in previous chapters demonstrated a transition from perceptual to category induction, suggesting that the natural default for induction is perceptual. In the introduction to this chapter, I discussed the two possible explanations for the transition. Firstly, the accumulation of knowledge theory suggests that the transition occurs as children accumulate relevant knowledge about the importance of category membership, and begin to appreciate the significance of category membership during induction (Gentner, 1988; Sloutsky et al., 2007). Secondly, the cognitive maturation theory suggests that the transition occurs as children’s overall cognition matures, which allows them to understand and apply increasingly abstract concepts (Piaget, 1964); an accumulation of knowledge would not necessarily result in an increase in category induction.

The findings from Study 6 suggest that the transition in inductive preference is unlikely to be due to a cognitive maturation, as the children in the intervention condition showed a significantly greater increase in the number of category induction choices made after the interactive educational session, relative to children in the comparison condition. With the testing sessions only a few weeks apart, these children would not have had the time to mature cognitively. This suggests that their altered strategy preference from perceptual to category was not due to a fixed stage-like cognitive development (Piaget, 1964), and instead the educational experience and accumulation of relevant knowledge encouraged them to appreciate and consider the importance of category membership. It also shows that knowing the category of items was not enough to use category-based induction: only the children who could correctly categorise the items above 83% were included, yet not all of these children showed a category-bias. This suggests that although children knew the specific category information enough to correctly categorise the items, they were lacking in the general relational understanding needed to fully appreciate the importance of this category information during induction. After the intervention condition received the interactive educational session, their appreciation for the importance of category information increased for all items. These data support the argument of a gradual transition from perceptual to category induction through experience and an accumulation of knowledge (Gentner, 1988; Sloutsky et al., 2007), and are less consistent with stage-like theories of cognitive development (Piaget, 1964) as well as the natural bias being category-induction (Gelman, 2003).

2.4 Conclusion

Study 6 has shown that a fundamental cognitive skill such as reasoning can be improved through interactive educational experience. The 30-minute interactive session helped children to see beyond the obvious, and encouraged them to consider the importance of features crucial to specific categories: they were able to apply their new knowledge to tasks disassociated to their zoo visit, rather than simply training them to improve on specific tasks only. This inductive improvement suggests that children’s performance was not constrained simply by their stage of cognitive maturation, but rather by their level of knowledge. Further research is needed to test whether experience and improvement in one domain, in this case biological, would cause a knock-on improvement in other domains.

In conclusion, children’s category induction can be boosted through interactive educational sessions, and children apply this newly learned information onto other items within the same domain.

These findings are consistent with the theory of a gradual transition from perceptual to category induction through experience and an accumulation of knowledge (Gentner, 1988; Sloutsky et al., 2007), and are less consistent with stage-like theories of cognitive development (Piaget, 1964), as well as the natural bias being category induction (Gelman, 2003).

Chapter 6: General discussion and future work

1 Summary

My overall aim was to investigate the development of young children's inductive reasoning. I had two main interests based on the literature I reviewed in Chapter 1. Firstly, I wanted to investigate the contradictory findings arising from research supporting the two broad accounts of inductive development, and examine whether methodological differences may explain these discrepant findings. Secondly, I wanted to explore the possible driving mechanism behind the development of inductive reasoning, and whether the development of category induction could be boosted with interactive educational sessions. Six studies, four chapters, covered these two main interests. Before I summarise my experimental work and address these two main interests, I shall describe the methodological manipulations I implemented to create a more powerful and valid test of children's default induction strategy.

1.1 Improving the methodology and stimuli

On examining previous research into inductive reasoning, it not only became apparent that there were conflicting accounts of induction development, but also that there were potential reasons for the contradictory findings supporting these accounts. Specifically, there were a number of weaknesses and differences in the methodologies.

In the pioneering inductive reasoning research, real biological kinds were used as examples in the induction tasks. Since the examples were familiar, children did not need to be taught the categories presented (Gelman & Markman, 1986). However, a disadvantage was that their induction strategy may have been influenced by previous knowledge and experience of the items. In 2007, Sloutsky et al., published a four-fold methodology: (1) category learning, 2) initial categorisation, 3) induction, 4) final categorisation), which enabled the use of novel kinds, and avoided the influence of previous knowledge. Although the methodology was well designed, they were heavily criticised for their stimuli, which were seen as arbitrary and unrealistic (Gelman & Waxman, 2007). Researchers in support of the category-bias account believed that children would not identify these artificial novel items as living kinds, and would therefore not apply the same reasoning as for real kinds. Another concern of previous research regarded linguistic information. Researchers in support of the category-bias account often used labels to highlight category membership during induction (Gelman & Markman, 1986; see also Gelman, 2003). This technique was criticised by researchers in support of the perceptual-bias account, who believed that labelling added an extra feature of similarity between items, causing label-based induction rather than category-based induction (Sloutsky & Fisher, 2004; Sloutsky, et al., 2001). Sloutsky et al., (2007) avoided using labels by presenting children with long explanations about the relational connection between items. It occurred to me that these explanations required a good level of language comprehension and attention, which may have been beyond the level of the majority of 4-year olds.

The four-fold methodology effectively eliminated the influence of previous knowledge, but in the Sloutsky et al., design there were only 8 categorisation trials, therefore it was impossible to judge whether an individual child was categorising the items above chance according to a binomial test. Therefore, I increased the number of categorisation trials to 12, since according to a binomial test, scores of 10/12 (83%) and above were significantly above chance (proportion = 0.5, *p* = .04). This made the task more stringent, and provided a check that children fully understood how to differentiate between the novel categories before they attempted the induction task. I wanted to avoid the potential priming aspect of labelling, but was uncertain about the lengthy relational explanations. Previous research demonstrated that young children understand the ageing process (see Inagaki & Hatano, 1996 for a review) therefore I designed a relational transformation of a juvenile growing up into an adult. These transformations preceded every induction triad, and formed the basis of the categorisation tasks (see Chapter 2, Study 1). The transformation successfully highlighted the relational connection without the need for language. Confirmation was found from children's comments: "Well, it's still a Sandbug 'cos it's just grown up"; "It was a Rockbug when it was a baby, and now it's grown up, it's still a Rockbug". This also meant that I was able to manipulate the perceptual similarity between the target and category choice in the induction triad: the category choice was always a small, pale, legless juvenile; the target was always a fully-grown dark adult. My final methodological improvement regarded the hidden properties. I based my hidden properties on work by Gelman and Markman (1986), and ensured they were all biologically plausible, non-leading, realistic and simple to understand for young children (see Appendices A, B and E for all examples).

Having strengthened the methodology and resolved certain outstanding concerns, I proceeded to strengthen the stimuli. While the four-fold methodology is perfectly designed for use with novel kinds, and novel kinds are perfect for eliminating varying levels of previous knowledge, I wanted to avoid very artificial novel kinds, like those used in Sloutsky et al., (2007) and ensure the items were biologically plausible. Once I had created my stimuli, I conducted numerous pre-tests to ensure that children saw these novel kinds as biological, living and realistic. I assigned them biologically plausible names, such as ‘Rockbug’ and ‘Sandbug’, and relevant behaviours for their identifying feature: “Rockbugs live in rocks and have sharp pointy heads for digging”. No previous induction study had created such biologically plausible novel kinds.

Finally, to fully investigate the development of induction, I needed to examine a wider age range of children. Previous research primarily focused on children aged 4-5 years old. However, focusing on such a narrow age range would make it impossible to identify any developmental changes. Therefore I tested children aged 3-9 years old, and for the first time in the literature, the developmental trajectory of inductive reasoning was observed.

1.2 Factors affecting the trajectory of inductive reasoning

Having addressed concerns about previous methodology and stimuli used to test inductive reasoning, I had created a strong base to investigate which manipulations were able to affect the trajectory of induction. It seemed remarkable that there were contradictory findings from researchers in support of the two broad accounts of inductive reasoning – the perceptual-bias and category-bias – when both sets of researchers were working with the same aged children and using the same triad paradigm. The paradigm directly pits perceptual choice against category choice, yet the results observed were polar opposites. Chapters 2 and 3 explored whether differences in methodology and stimuli affected the developmental trajectory of induction, which in turn caused the contradictory results. I wanted to investigate whether certain manipulations made inductive reasoning, and specifically category induction, easier for children. Throughout Studies 1-4, I observed a clear gradual transition from perceptual to category induction consistent with age. Interestingly, I also identified which factors were able to alter this trajectory.

Researchers supporting the category-bias account have often used real images which mean that although the target and perceptual choice are perceptually more similar than the target and category choice (Gelman & Markman, 1986), they are not as closely matched as the artificial stimuli often used by researchers supporting the perceptual-bias account, which can be manipulated to an exact degree (Sloutsky et al., 2007). I began by looking at the influence of the degree of perceptual similarity between the target and perceptual distracter in the induction triads (HSDs vs. LSDs; Study 1), and found that all years were equally more distracted by the high similarity distracters (HSDs) than the low similarity distracters (LSDs). This finding is consistent with claims made in previous literature (Osherson, et al., 1990; Sloman, 1993) suggesting that humans are highly influenced by similarity. However, I uncovered a novel finding that although the degree of perceptual similarity affected children's choice, the gradual transition in inductive strategy preference could *not* be due to the influence of similarity. I observed no interaction between the level of similarity and year group; the effects of age and similarity were independent of one another. Although more category choices were made as children were getting older, all year groups were equally distracted, and made fewer category choices when the perceptual choice was an HSD. Therefore, the transition towards category induction is unlikely to be caused by an increased ability to inhibit HSDs. Nevertheless, there was an alternative explanation for young children's preference for the perceptual distracter: specifically, the perceptual distracter matched the target in its maturity (Taylor and Gelman, 1993). Before I could make a firm conclusion I needed to investigate whether the young children's inductive reasoning responses had been influenced by generalisation based on the maturity of the item: whether there are developmental changes in the types of category information children consider to be important (maturity categories vs. taxonomic categories).

I designed a new study (Study 2), which not only directly pitted perceptual choice against category choice, but also included a mixture of maturities: juvenile and adult category choices. If the data showed the same pattern regardless of the item maturity, then the interpretation of induction being a gradual transition from perceptual to category preference would be more convincing. I observed an identical gradual transition from perceptual to category preference regardless of whether the category choice was a juvenile or an adult. The finding reassured me that children's choices were based on either a preference for perceptual or category information, and not based on the maturity of the stimuli presented.

Studies 1 and 2 showed a gradual transition in young children's inductive strategy preference that could not be due to either the degree of perceptual similarity or the maturity of the stimuli. It did however highlight that when the perceptual choice was higher in perceptual similarity to the target (as is common in research supporting the perceptual-bias account), children made fewer category choices compared to when the perceptual choice was lower in perceptual similarity to the target (as is common in research supporting the category-bias account).

I continued this methodological exploration by focusing on the effect of domain and category structure (Study 3). The literature suggests that young children have an essentialist bias to use category induction with natural kinds, a bias not found with artefacts (Gelman, 2003; Keil, 1989; Markman, 1989). However, this had not been tested in a well-matched, across-domain study. I also noticed that researchers supporting the two accounts often used different category structures during testing. Researchers in support of the perceptual-bias account tended to use complex relational category structures (see Sloutsky et al., 2007 images in Figure 27), whereas researchers in support of the category-bias account tended to use examples which, although complex overall (as is the nature of real kinds), they could be distinguished on the basis of individual features, for example, birds have beaks (see Gelman & Markman, 1986 images in Figure 27).

Figure 27. Differences between stimuli used in research by Gelman and Markman (1986; the left-side images; bottom bird = target, left bird = category choice, and top bat = perceptual choice) and Sloutsky et al., (2007; the right-side images; top ziblet, with more fingers than buttons = target, bottom right ziblet = category choice, and bottom left flurp, with more buttons than fingers = perceptual choice).

I wondered whether this difference could also be a factor in the conflicting results. I manipulated category structure (featural vs. relational) and domain (natural kind vs. artefact) orthogonally using well-matched artificial categories. Every child participated in either the featural or relational condition, and completed both a natural kind and an artefact task within their given condition. If an essentialist bias is present for natural kinds, one would expect to see higher levels of category induction for natural kinds compared to artefact kinds. Equally, if featural conditions are less demanding on cognition, one would assume that more category choices would be made during the featural condition compared to the relational condition. Interestingly, my findings showed no differences in inductive strategy preference between the domains of natural kind and artefact, suggesting that the development of induction is domain-general. However, category induction in children completing the relational category structure condition was delayed by approximately a year; this delay could be due to the greater demands on cognitive processing (Halford et al., 1998; Richland et al., 2006; Thibaut et al., 2010; Zelazo & Frye, 1998).

Drawing on the two literatures, I was left with one final methodological discrepancy: labelling during the induction triads (Study 4). There is an on-going debate as to whether children need linguistic information to use category induction, i.e., whether labels are proxies for categories (Gelman & Markman, 1986; Waxman, 1999; Waxman & Booth, 2003) or whether labels are features for categories (Sloutsky & Fisher, 2004; Sloutsky, et al., 2001). I wanted to explore whether using labels during my induction triads would alter children’s strategy preference. Children completed the exact same task, but were divided into either a ‘with labels’ condition or a ‘without labels’ condition. I found that when labels were used, children of all ages were significantly biased to use category induction, compared to the gradual transition observed in the ‘without labels’ condition. One interpretation of this finding is that children used the labels as a reminder of the category membership of the items, and then went on to use this category information to inform their induction decisions (supporting Gelman & Markman, 1986; Waxman, 1999; Waxman & Booth, 2003). However, based on this study, it is not possible to rule out the alternative interpretation: that children simply regarded the labels as an additional, highly salient feature, and therefore the category choice shared more features with the target than the perceptual choice. In fact, since children in both conditions could accurately identify the category membership of the items shown both before and after the induction task, it is unlikely that they had forgotten the category membership during the induction task. I would argue that it is more likely that the labels added an extra shared feature between target and category choice, on which children in the ‘with labels’ condition based their decisions. Because of the possible priming effect of labels, I decided not to include labels in my main thesis studies, and instead focused on addressing the question of when children begin to use category induction, unprompted.

Studies 3 and 4 showed a gradual transition in young children’s inductive strategy preference that did not differ for domain. However, this transition was delayed by a year when children had to reason using more complicated category structures, and was enhanced by the use of labels during induction. This highlighted that when simpler featural category structures were used (as by researchers supporting the category-bias account), children made more category choices compared to when more complex relational category structures were used (as by researchers supporting the similarity-bias account). It also shows that children made more category choices when labels were used during induction (as by researchers supporting the category-bias account) compared to when labels were avoided (as by researchers supporting the perceptual-bias account).

1.3 Mechanisms driving developmental changes in inductive strategy preference

In the first two experimental chapters I successfully explored various methodological differences that could be behind the contradictory findings. I found that young children gradually transition from perceptual to category induction, in support of the perceptual-bias account. The transition occurred earlier when the tasks used featural stimuli, labels, and LSDs, all of which were favoured more by researchers in support of the category-bias account. Having explored the primary differences between previous research methodologies, I continued by investigating the underlying mechanism driving this gradual transition. Previous literature and observations from Studies 1-4 highlighted two potential driving forces: 1) the development of cognitive control (EFs), and 2) an accumulation of knowledge and understanding. These two potential driving forces were considered in Chapters 4 and 5. Chapter 4 examined the relationship between inductive reasoning strategy and the development of cognitive control (EFs). Chapter 5 examined the effect of interactive educational sessions on inductive strategy preference. I wanted to investigate whether certain extraneous factors were affecting the development of inductive strategy, and whether this strategy preference could be altered.

Drawing on the EF and reasoning literature (Bunge & Zelazo, 2006; Halford, 1993; Halford et al., 1998; Richland et al., 2006; Thibaut et al., 2010; Zelazo & Frye, 1998; Zelazo & Müller, 2002), I began by selecting three simple EF tasks to measure some of the most important EFs for cognition: working memory, shifting, and inhibition (Baddeley, 1996; Logan, 1985; Miyake, Friedman, Emerson, Witzki & Howerter, 2000; Rabbitt, 1997). I also included one cool complex EF task, and one hot complex EF task which tap into the three simple EFs, as well as other cognitive skills such as planning and decision-making. All of these EF tasks were completed alongside one of my inductive reasoning tasks (Study 5). The EF tasks are standardised: Digit Span (working memory), the Shape School part two (inhibition), the Shape School part three (shifting), the Tower of London (strategic planning and problem-solving), and the Children's Gambling task (affective decision-making). Using a selection of EF tasks allowed me to directly compare children's inductive reasoning strategy against different aspects of simple and complex EFs.

Age, unsurprisingly, was significantly correlated with inductive reasoning, and was the only significant predictor of the variance in inductive reasoning. This finding supports the perceptual-bias account and is consistent with the gradual transition I observed in Studies 1-4: the number of category choices made increased with age. When using a two-tailed exploratory analysis, none of the

EF tasks reached correlational significance. Equally, none of the EF tasks could significantly predict the inductive variance before or after age had been removed during the multiple regression. When using one-tailed analysis, Digit Span (simple EF, working memory) and Tower of London (cool complex EF, planning and problem solving) were significantly correlated with inductive reasoning. This suggests that perhaps working memory, planning and problem solving play a part in inductive reasoning, but since the correlations are relatively small, it is unlikely that these are crucial factors driving the development of inductive reasoning.

The data from Study 5 is more consistent with the perceptual-bias account. Age proved to be a significant predictor of the preferred induction strategy, unlike the EF tasks, which were unable to significantly predict induction variance. I would conclude that EF may play a part in inductive reasoning, but it is unlikely to be the driving force behind the transition. Although I observed a transition, and know age to be a significant predictor of induction, I was unable to say whether this was due to a cognitive maturation (Piaget, 1964) or an accumulation of knowledge and experience (Gentner, 1988; Sloutksy et al., 2007). Before I could show any firm support for either theory, I explored this matter in greater depth in Chapter 5.

In the final experimental chapter, Chapter 5, I divided children into two conditions – comparison and intervention – and tested them all on four inductive reasoning tasks over two sessions. In between the two sessions the children in the intervention condition visited Twycross Zoo for the day, and took part in an interactive educational session about animal adaptations to habitats. The educational session was designed and run by the educational team at Twycross Zoo. Children were allowed to touch or hold small animals whilst being told about the importance of feature adaptations to habitats. The children were unaware that the zoo visit was connected to my visits in school. Two of the inductive reasoning tasks used real animals, for example, Chinchillas, and two of the tasks used novel animals, for example, Forest weevils. I found that the number of category choices significantly increased during the second testing session for those in the intervention condition (who had taken part in the interactive educational session at Twycross Zoo), but did not significantly increase for those in the comparison condition. This suggested that category induction could be enhanced through interactive experience. Interestingly, this significant increase was observed for the novel kinds as well as the real kinds seen during the educational session. This implies that domain-specific experience (in this case biological) enhances category induction for the domain as a whole (Gentner, 1988), and challenges previous research that has suggested that reasoning improves only for the specific items experienced (Goswami, 1992). Goswami (1992; 2001) claimed that young children are capable of reasoning analogically, although they can be constrained by their knowledge of the relevant relations. Therefore, once a child has had explicit knowledge about the relational connections between specific items used in a reasoning task, then they will use relational information by default. However, if the child has not had specific relational knowledge about the items shown, then they will not know the relational connections, and will be more likely to revert to using perceptual cues. My findings make this theory unlikely for induction.

The data from Study 6 is consistent with researchers who suggest that induction can be enhanced through experience and an accumulation of knowledge (Gentner, 1988; Sloutsky et al., 2007). My findings do not support researchers who suggest that reasoning cannot be enhanced until children's cognition has matured into the next developmental stage (Piaget, 1962). Importantly, the education and experience does not have to be directly connected to the reasoning tasks for children to improve. In this case, children learning about adaptations to habitat during an educational session at the zoo, were able to apply this newly learned information onto my inductive reasoning tasks in school.

My studies have added to the current literature; have addressed some key concerns regarding methodological standardisation, and have started the exploration into the driving force behind the induction strategy transition. Studies 1-4 allowed me to highlight and identify problems with directly comparing previous inductive reasoning research, and I was able to begin to explain why previous research has yielded such opposite results. In particular I was able to provide more support to the perceptual-bias account by observing a gradual transition under a multitude of methodological manipulations. The gradual transition from perceptual to category induction guided my focus towards understanding the process behind this transition. Studies 5 and 6 allowed me to begin the search for the driving mechanism behind induction, based on previous literature as well as the data from Studies

1-4. Although I have already identified variables that may or may not drive induction, there are many ways to extend this research to get a clearer idea of the driving force and its trajectory.

2 Towards a theory of category induction development

The majority of inductive reasoning research fits neatly into two broad accounts of induction development. Firstly, researchers in support of the perceptual-bias account state that children have a natural bias to use perceptual induction, and only at a later stage do they begin to understand the importance of category membership and apply it during induction (Gentner, 1988; Sloutsky, et al., 2007). Secondly, researchers in support of the category-bias account state that children have a natural bias to use category induction when the relational connections between items are clear (Gelman, 2003; Gelman & Markman, 1986; Opfer & Bulloch, 2007). In the previous section I explained how I improved current methodologies and stimuli, and addressed outstanding concerns about whether methodological manipulations could influence the trajectory of induction. These changes and investigations have enabled me to make firmer conclusions about the development of induction, and what can affect the developmental trajectory. In all of my studies, and with a variety of different stimuli, I found a gradual transition from perceptual induction to category induction with age. Although I did not exactly replicate results previously reported from researchers in support of either account, the youngest children in my studies were predominantly perceptually-biased, which is consistent with the perceptual-bias account. I found that an increased category preference was observed only when the perceptual distracter was an LSD, the category structure was simple featural, and when labels were used; all of these manipulations are common in studies supporting the category-bias account. Although children taking part in these cognitively simpler manipulations showed an increase in category induction, the youngest children still favoured perceptual cues, and a transition was still observed, which supports the perceptual-bias account.

Therefore, overall, my findings are more consistent with, and lend support to, the perceptual-bias account. The account states that young children are initially perceptually bound, and only over time begin to shift to using relational information. My findings and support of the perceptual-bias account are particularly striking considering all of the methodological improvements made, including making the inclusion criteria stricter, and using highly plausible biological novel kinds.

In the next section I draw from my experimental findings to develop a theory of the driving force behind induction.

2.1 What is the driving force behind children developing category induction?

Assuming that the natural default for induction is perceptual, the next step is to develop a theory to explain the subsequent development of a category induction strategy. To do this, I shall draw inspiration from a broader research base, contrasting theories based on cognitive maturation versus an accumulation of knowledge.

There are two broad types of cognitive maturation accounts that relate to my findings. Firstly, those who believe that children's cognition develops through a qualitative change in the nature of children's thinking, due to a stage-like shift of maturity (Inhelder & Piaget, 1958; Piaget, 1964), whereby time and increasing sophistication of cognition allows children to move sequentially through pre-set cognitive stages. Secondly, those who hypothesise that children's cognitive capabilities improve as their cognitive control (EF) develops (Andrews & Halford, 2002; Halford, 1993; Morrison, Doumas & Richland, 2011; Richland et al., 2006; Thibaut et al., 2010). In this case children can only successfully complete reasoning tasks when their cognitive control has matured. In contrast, what I refer to as the accumulation of knowledge account, hypothesises that the transition to using category induction is driven by the level of knowledge and understanding about the relational importance during induction (Gentner, 1988; Gentner & Toupin, 1986; Inhelder & Piaget, 1958; Sloutsky et al., 2007). This theory places no emphasis on induction being restrained by cognitive stages or the development of cognitive control, and instead suggests that this transition occurs as soon as children have accumulated enough information to understand the relational significance. In the following section I shall discuss how some of my experimental work can be used to choose between these two broad types of account, and based on this, I can develop a hypothesis as to the driving force of induction.

2.1.1 Support for cognitive maturation

Two experimental sections of my thesis can be used to investigate the theory that cognitive maturation drives the ability to use category induction. I shall begin by discussing Study 3, Chapter 3: category structure, then move on to discussing Study 5, Chapter 4: EFs.

My category structure study (Study 3) used the same task to compare the effects of two different category structures on inductive strategy preference. Half of the children completed the task with a simple featural category structure (categories based on a single feature); the other half completed the task with a complex relational category structure (categories based on the relationship between two features). Although I observed a gradual transition from perceptual to category induction for both conditions, the transition was delayed by approximately a year for those children using the complex relational category structure. Children in the relational category structure condition made fewer category choices than those of the same age in the featural category structure condition. This finding provides strong support for the cognitive maturation theory, which suggests that underdeveloped cognitive control can limit children's reasoning. These results suggest that the complex category structure placed a greater demand on cognitive control because it required processing of multiple factors, and the operation of various EFs, such as working memory and inhibition. Although children in both conditions could correctly categorise the items, and could understand the relational connections, those in the complex relational condition (where a developed EF is needed), were unable to apply category induction at such a young age. The greater demand on EFs resulted in children reverting back to using simpler perceptual cues during the induction task. This theory would suggest that as a child's EFs develop and improve, so would their ability to reasoning on more complex tasks.

Interestingly however, when I investigated whether there was a direct relationship between the variance in inductive reasoning and the most important EFs (Study 5), I found that none of the EFs measured could significantly predict inductive strategy preference. Age *was* a significant predictor of EF performance, with older children performing better than younger children, and age *was* a significant predictor of inductive reasoning, with older children making more category choices. However, there was no relationship between performance on the inductive reasoning task and performance on the EF tasks once age was removed. This finding fails to support previous literature suggesting that the development of EF would have a direct impact on children's reasoning (Bunge & Zelazo, 2006; Halford, 1993; Halford et al., 1998; Richland et al., 2006; Zelazo & Frye, 1998; Zelazo & Müller, 2002). It also fails to support my earlier assumption from Study 3.

Although young children clearly struggle more when presented with multiple task requests or complex information, the development of cognitive control and EFs is unlikely to be the driving force behind induction.

2.1.2 Support for accumulation of knowledge

Two experimental sections of my thesis can be used to investigate the theory that an accumulation of knowledge drives induction, and more specifically, the ability to use category induction. I shall begin by discussing Study 3, Chapter 3: domain, then move on to discussing Study 6, Chapter 5: educational sessions.

My domain study (Study 3) divided children into two conditions to directly compare the domains of natural kind and artefact based on the same inductive reasoning task. I found that domain had no effect on the developmental trajectory of induction: as children transition from perceptual to category induction in one domain, they transitioned at the same rate in the other domain. This suggests a domain-general effect of induction, and fails to support previous literature suggesting that young children will have an essentialist bias to use category induction for the natural domain before other domains (Gelman, 2003; Goswami, 2001; Keil, 1989; Wellman & Gelman, 1998). This finding would suggest that as children accumulate relational knowledge and understanding in one domain, they transfer the concept of relational significance to other domains, rather than having to learn this information for each specific domain. Study 3 showed that when children understood to use category membership, then they did so for both domains. This was also directly linked with age, with older children making more category choices. As children get older they accumulate more information, and gradually transition their strategy from perceptual to category induction with increased knowledge and understanding.

The accumulation of knowledge theory, and my assumptions based on Study 3, can be supported by my study into the influence of interactive educational sessions at Twycross Zoo (Study 6). Children were divided into a comparison condition and an intervention condition. Although they completed the same four inductive reasoning tasks across two testing sessions, the intervention group also took part in an interactive educational session with small animals in between the two testing sessions. Only the children taking part in the interactive educational session made significantly more category choices in their second testing session, suggesting that the ability to use category induction develops with an accumulation of relevant knowledge. Importantly, the significant increase was not only observed for the real kinds seen during the interaction, but also for the novel artificial kinds. Once again, these findings suggest that the children in the intervention condition were transferring the concept of the importance of relational information during induction, and in this case it was from familiar to unfamiliar kinds. This goes against previous literature suggesting that domain information is insufficient, and children may be held back by performance constraints, such as not knowing the relational connections between specific items (Goswami, 1992). Instead, this data supports the hypothesis that at as soon as children have learnt the importance of relational information within a domain, they can apply that knowledge to all items within that domain (Rattermann & Gentner, 1998).

In conclusion, my studies have shown that it is very likely that the natural bias for inductive reasoning is perceptual and, children gradually make the transition to using category induction over time. My data has also shown that certain methodological manipulations can affect the developmental trajectory, which can lead to contradictory results. These methodological differences are likely to be the cause of different theories about the development of inductive reasoning. Taking inspiration from the broad accounts of reasoning development, and from my data, I can speculate on a theory about the driving mechanism of inductive reasoning development. I have shown that EFs are unlikely to be the driving force behind induction; however, I do suspect that they support induction, and can affect the developmental trajectory. For example, if the induction task presented uses complex information, or multiple cognitive requests, then children's focus on category information will be reduced due to the greater demand on EF. It is undeniable that a certain level of cognitive maturation is necessary for retaining, and applying knowledge about category relations. However, it is unlikely that children's inductive reasoning progression is driven predominantly by a stage-like maturation, or the development of EF. Based on my experimental work, I speculate that the driving force behind induction, and specifically category induction, is an accumulation of knowledge. My work has shown that children made more category induction choices after receiving an interactive educational session about biological kinds. Therefore, I believe that as children gain a greater level of relevant knowledge regarding the importance of category membership, they will begin to appreciate the role of category membership during induction. Once children have knowledge enough to understand the relevance of category membership, they transfer this conceptual understanding to other domains.

3 Future work

In this section I present two areas for future work. The first area develops and extends the work from my thesis, specifically inductive reasoning and EFs; inductive reasoning and risk taking, and exploring inductive reasoning in more depth within domains. The second area discusses my ideas for new studies to extend our knowledge about inductive reasoning. Firstly, I shall propose a longitudinal study, and secondly I shall suggest ways to investigate geographical or environment influences.

3.1 Extending my current work

This section is a mixture of minor amendments to my current work, and new studies based on my current work.

3.1.1 Inductive reasoning and executive functions

Based on the results of Study 3, and the previous literature (Halford, 1993; Richland et al., 2006; Thibaut et al., 2010; Zelazo & Frye, 1998), the null effect of EF on inductive reasoning variance was unexpected. Although, based on my findings, I have concluded that EF seems to support, but not drive inductive reasoning development, I believe that some extra work on this area would strengthen this conclusion.

My aim for this work extension would be to include purer EF tasks, an extended age range, and a larger sample size to increase power. The EF and inductive reasoning tasks would initially remain the same as Study 5 to allow for direct comparison with the already collected data. Study 5 consisted of three year groups and 60 children, 20 children from each year group. Four of these children were unable to pass the strict criteria for inductive reasoning and were therefore removed, leaving the data of 56 children to be analysed. Although this sample size was sufficient to observe medium to large effects, it did not provide sufficient power to observe more subtle influences. Thus, it is possible that I experienced a type-2 error, whereby I have had to accept my null hypothesis (that EF will have no relationship with inductive reasoning), even though it is false. Based on this possibility, it would be interesting to see whether an increase in sample size and therefore power, would enable us to detect significant, albeit small effects of EF on induction preference. A total sample size of 90 children would provide power of .80 to detect effects of size I observed in Study 5, so to ensure I had sufficient power, I would have a minimum of 30 children from each year group participating.

Since my previous studies showed that most children make the transition from a perceptual to a category bias sometime between Reception and Year 2 (4-7 years old), I focused on these three age groups for my EF research. However, it is possible that this narrowed age range was too compressed to see the full developmental change of EF, and therefore reduced my chances of observing its influence. It is well established that the development of EF shows the greatest increase during the pre- and early school years (see Espy et al., 2006), so including children aged 3-9 years old would allow me to directly compare the full developmental transition for inductive reasoning (as shown in Study 1), against the full developmental trajectory of EF. It may be that any driving force on inductive reasoning from EF development occurs at an earlier age than I observed in Study 5.

Finally, although the EF tasks administered were commonly used, standardised tests of EF, it is possible that the simple EFs were not pure tasks. It has been suggested that various EF tasks claiming to tap into one factor of EF, in fact tap into more than one (Pasalich, Livesey & Livesy, 2010). For example, the Shape School task assessed inhibition and shifting, however during testing it became apparent that this task was also examining working memory: children had to remember the naming rules, for example, '…the children with hats on, their name is their shape'. Due to the simple nature of Digit Span (working memory), one would assume that this would be a pure EF task and only tap into working memory, however other researchers have claimed that even a task like this could involve inhibitory processes (Hala et al., 2003). Future work could replicate Study 5 by using the same inductive reasoning task, and the same complex tasks, however the non-pure simple EF tasks would be substituted for pure tasks. There are some researchers that believe that a pure EF task is unlikely to exist because there is always going to be some overlap in the cognitive processes required (Burgess, 1997; Carlson, 2003). However, using pure, or at least purer, EF tasks would allow for a closer comparison between induction and the three most important EFs. It would be interesting to see whether using purer EF tasks would reveal a clearer pattern of inter-relationships in my correlation and regression analyses.

3.1.2 Inductive reasoning and risk taking

A second area of interest would be to further examine the negative (although non-significant) correlation between inductive reasoning and the Children's Gambling task, as this was unexpected. This was the largest negative correlation with age. I acknowledge that the following proposed work extension goes beyond the initial aims of the thesis, however, my EF and inductive reasoning study has opened up new questions regarding the EF tasks. Although my experience shows that children enjoy receiving stickers, it is possible that the level of reward (and therefore emotional impact) was not high enough for children to focus simply on winning as many stickers as possible.

My new proposed procedure would assess children's performance when faced with a range of different rewards. I would include four different conditions varying on degree of emotional impact, where children would be rewarded differently depending on their condition. Condition One would be told that they would be awarded one sticker for playing: low emotional impact. Condition Two would be told that the more stickers they win, they more they get to take home (directly replicating Study 5). Condition Three would be told that the more stickers they win, the more chocolates or sweets they get to take home (adapted from Kerr and Zelazo's use of M&M chocolates): children may find chocolates more appealing than stickers. Condition Four would be told that their score will appear on a board, and only the top scoring children will be awarded a prize: high emotional impact as children are openly pitted against one another. Clearly new ethical considerations would have to be made for Condition Four, and perhaps all children, unbeknown to them, would be awarded a prize at the end. However, schools often have systems of team points or sticker charts which clearly show the high achieving children, or sports days which highlight the fast versus slow children, so Condition Four may not be so unethical. Creating the four conditions would allow me to see whether the level of emotional incentive alters the risky behaviour; whether children are less likely to risk take when they stand to lose something more important to them.

As well as creating four conditions, it would be interesting to widen the age range upwards, and include adults. This would allow me to see whether the risky behaviour continues into adulthood. At this point I would be able to see whether there were any relationship between the participants expressing risky behaviour, with those showing a 'no bias' during the induction tasks. It could be that these 'no bias' participants *are* in fact biased, but have other variables at play, for example, risk taking. A participant may know that the category choice is the correct choice, but would occasionally choose the perceptual choice to see whether they are still rewarded with a sticker (children are rewarded with a sticker for playing the inductive reasoning game).

3.1.3 Inductive reasoning across domains

A third area for development would be to expand the domain research. In Study 3, I investigated the developmental trajectory of induction across the natural domain and the artefact domain. A null effect was found for domain, suggesting a domain-general effect of induction. However, it occurred to me that perhaps this comparison was not the most informative. Although the pre-tests confirmed that in the majority of cases children saw the artefacts as nonliving, there may be an extraneous variable influencing the results. Throughout the generations, children's television has personified inanimate objects, for example, Thomas the Tank Engine, or the machines in Bob the Builder. Although children may have recognised my items as 'nonliving' in the sense that they are machines, they may have assigned them a living status, and transferred their natural kind induction knowledge. Although this seems unlikely, it would be interesting to adapt my induction task to test other less personified artefacts. I could use different item types within the domains: within the biological domain I could compare animals with natural substances, such as honey and stone (as used in Gelman & Markman, 1986). Within the artefact domain I could compare machines with household items, such as a chair and a hammer. It would be interesting to see whether the same perceptual to category transition was observed for the items that would not as easily be given the status of 'living'.

It would also be interesting to see whether children's category induction could be enhanced for the artefact domain, through interactive educational sessions with nonliving items. In this case, the methodology would replicate Study 6, but the natural kind tasks and educational session could be replaced by artefact kinds, and an educational session providing information about the importance of features and uses of nonliving kinds.

3.2 New studies

3.2.1 Longitudinal study

Although a developmental transition from perceptual to category induction has been repeatedly observed with age, this has only ever been observed cross-sectionally. To fully understand the developmental process one needs to observe a child's progress repeatedly over a few years. Previous research has conducted longitudinal studies into social reasoning (Damon, 1989) and moral reasoning (Walker, 1989; Walker & Taylor, 1991), however a longitudinal study has never been conducted into inductive reasoning. Previously, longitudinal research into reasoning has worked with participants from the age of 4 years, over a two-year time frame (Damon, 1989; Walker, 1989; Walker & Taylor, 1991), interviewing children either annually or bi-annually. The largest limitation of this longitudinal research was the drop-out rate, usually due to participants moving away. Walker (1989) tested 240 participants during the first testing session, but was only able to test 233 of these during the second testing session. Therefore a large sample size would be required to start.

My experimental work shows a gradual transition from a very clear perceptual induction preference at age 3, to a very clear category induction preference by age 9. However, I am unable to determine whether each individual child's change in strategy preference is gradual or sudden. Firstly, it is possible that every child gradually transitions with age, as the overall across-age transition in my experimental work would suggest. This could be due to a gradual appreciation of the importance of category induction, through an accumulation of knowledge. Secondly, it is possible that each individual child transitions quite suddenly, but due to the inter-child variability, the overall transition appears gradual. Each child could be accumulating knowledge, but after a certain level of cognitive maturation, the child would abruptly understand the importance of category induction.

Based on my experimental findings, I would ideally work with a group of children as soon as they turn 3 years (predominantly perceptual induction), and monitor their strategy preference until they were age 7 years (predominantly category induction). This age bracket highlights the greatest transition from perceptual to category induction preference. Although previous research collected data annually or bi-annually, I would hope to collect data at least every six months to monitor whether individual transitions are gradual or sudden: one or two years between testing sessions would not be accurate enough for determining the rate of transition. Monitoring each child individually, over a four-year period would enable me to see whether the transition was gradual – as the overall developmental trajectory would suggest – and would support the idea of an accumulation of knowledge, or whether it was a more complete change of strategy preference, which would perhaps give more support to a stage-like shift.

Testing children so frequently would result in having to use different stimuli for each induction task so that they do not become too familiar with one category of stimuli. To begin with I would work solely with the biological domain. Although a longitudinal study into the developmental trajectory of inductive reasoning has never been conducted, it has the potential to more accurately identify the true developmental process.

3.2.2 Geographic and cultural effects

My final conclusion, based on my experimental work, was that the natural default for induction is perceptual, and that with an accumulation of knowledge children begin to transition to using category induction. Study 6 showed that for some children in my study, a 30-minute interactive educational session with animals was enough to boost this transition. The schools chosen for my experimental work were deliberately matched across a variety of parameters to ensure roughly the same socio-economic status, intelligence, and availability of educational materials. However, it occurred to me that introducing geographical location as an unmatched variable could yield some interesting findings. If a 30-minute interactive session could boost category induction, I wondered whether children from locations with differing levels of biological experience, would show different developmental trajectories. It would be interesting to investigate whether those children who are brought up in a more rural environment, exhibit higher levels of category induction from an earlier age. If this were the case, it would provide extra support to the accumulation of knowledge theory, showing that cognitive maturation cannot be the driving force. However, if it was found that children surrounded by biological experiences did not show a transition until a similar age as children growing up without any significant level of biological experience, then this would provide support for the cognitive maturation theory: children cannot begin to understand the importance of induction until they have matured cognitively. Unfortunately, data from a limited population is often generalised to other populations and cultures, but is not often tested and verified across these other populations.

Lopez, Atran, Coley, Medin and Smith (1997) tested two very different groups on inductive reasoning tasks: undergraduates from America, and Itzaj Maya adults who have extensive biological knowledge. Lopez et al., found significant differences in the way that these two groups responded to induction questions about mammals. The American undergraduates tended to use diversity-based induction (induction based on how distant two animals are from each other in terms of folk taxonomy: the greater the distance, the greater the diversity). Whereas the Itzaj Maya adults tended to use ecological-based induction (induction based on the likelihood of needing an ecological agent, e.g., if two animals are diverse – tapirs and squirrels – then they are more likely to need an ecological agent, such as being bitten by a bat, to pass on infection, compared to two animals that are closer in folk taxonomy – rats and mice – who are more likely to pass on the infection without the ecological agent). However, there was a multitude of differences between these groups, some of which would be impossible to control for, like education, literacy, language and livelihood. To test notably different cultures guarantees more group differences; these differences cannot be controlled for because they are part of the group formation. In 2002, Bailenson, Shum, Atran, Medin and Coley attempted to overcome some of these conflicting group differences by using the ‘triangulation strategy’, whereby they tested the two extreme groups, as well as a third group which resembled each group in different significant ways. For example, 1) undergraduates brought up in the United States, 2) Itzaj Maya adults who have extensive biological knowledge, and 3) bird specialists (extensive bird knowledge) brought up in the United States. In this instance, if the third group performs in a similar way to one of the two original groups, then the researchers can assume the shared variables to be influencing the performance.

My first proposed study would investigate the difference in inductive reasoning strategy between children from rural versus urban locations. To maintain validity, it would be essential that the children were tested in the same environment as they were brought up in, to ensure that their biological input had been consistent. For example, if a child was tested in a rural location, and therefore classified as a 'rural participant', but they had only moved there recently from an urban location, then their upbringing would have been different compared to the children who had always lived in the rural location. The difference could cause misleading results. My second proposed study would investigate differences between extreme cultures, as in Lopez et al., (1997; Bailenson, 2002). For example, in some cultures children grow up tending to, and living in close proximity to the animals. It would be interesting to examine whether these children have a category bias earlier than children who grow up in an inner city, and who have never had the opportunity to interact with animals. Depending on the context, the inductive reasoning task may have to be adapted to ensure it is appropriate across languages and for younger ages. If children growing up surrounded by biological interaction do show a strong category bias at a young age, then it would be worth adapting the task to test children as young as 2 years old to look for a floor-effect.

The main aims for this research would be firstly, to examine whether biological experience is enough to trigger category induction, or whether experience must be coupled with education. Secondly, whether children can transition to using category induction at any age, given sufficient experience and education, or whether the transition is dependent on a cognitive maturation. One possible way to address these aims would be to create a triangulation study similar to Bailenson et al., (2002). My conditions would be as follows: 1) rural and school educated (i.e., children who receive a formal school education as well as an animal-specific education through helping to look after animals, such as on a farm), 2) urban and school educated (i.e., children who receive a formal school education, but have no interaction with animals), and 3) rural and no education (i.e., children who, due to cultural reasons, receive no formal school education, but have substantial animal-specific education through tending to, and living in close proximity to animals). All conditions would complete my induction tasks. If condition one – rural and educated – performed in a similar way to one of the other conditions, then I might be able to identify which variable – school education or animal-specific experience – would be the most influencing factor for determining category induction.

4 Final conclusion

My empirical findings have broadly supported the perceptual-bias account, which suggests that the natural default for inductive reasoning is perceptual, and over time children transition to using category induction. Thus, I have shown that an early bias for category induction is unlikely. My work has shown the first developmental trajectory of induction across six year groups, and has provided explanations for the contradictory findings between the two broad accounts of induction. Importantly, it has also explored the influence that certain methodological manipulations have on the developmental trajectory of induction, and whether these need to be standardised for future research.

What is most intriguing is the driving force behind this developmental progression from a perceptual-bias to a category-bias. My findings have shown that children can categorise the items very accurately from a young age (Studies 1-6), yet only begin to use this category knowledge in induction from around 6-7 years old. I've shown that this inability to focus on category knowledge is not because the younger children are more distracted by highly similar alternative choices (Study 1), or because they are focusing on differing types of category knowledge (e.g., maturity categories; Study 2). It is also not the case that children progress at different rates depending on the domain (Study 3), or are unable to transfer their knowledge onto unfamiliar items (Study 6). Instead, there appears to be something that children are unable to understand at an early age about the importance of category membership, or the relevance of it during induction. Although I do think that an underdeveloped PFC can affect the apparent trajectory of induction, for example, children's capabilities on completing complex induction tasks (Study 5), I believe that the development of EFs support but do not drive the development of induction. Instead, my work shows the importance of an accumulation of educational knowledge for enhancing category induction (Study 6). This cannot be explained by a fixed, stage-like cognitive maturation, and instead, is more consistent with the broad accumulation of knowledge theory (Gentner, 1988; Sloutsky et al., 2007). I propose that cognitive maturation supports the development of category induction, but children must first understand the relevance of category membership during induction through an accumulation of knowledge and experience. This suggests that all children begin perceptually-biased, and may make the transition to using category induction at an earlier age if they have the opportunity to accumulate category knowledge through experience and interaction. It is not enough to simply know the relational connections between items; children must understand the *importance* of these relational connections. This can only occur when children learn to appreciate the deeper significance of items, as was taught at Twycross Zoo (Study 6).

As I have discussed, there is much research still to be conducted, which would allow me to make firmer conclusions about the developmental trajectory of induction, and the driving force behind the transition to category induction. I have proposed ways in which I would like to further this research by extending my current work, and by designing new studies. I have set the stage for further exploration into the driving force behind induction, and whether certain early-life influences can affect each child's inductive reasoning capabilities.

References

Alexander, P. A., Wilson, A. F., White, C. S., Willson, V. L., Tallent, M. K., & Shutes, R. E. (1987). Effects of teacher training on children's analogical reasoning performance. *Teaching and Teacher Education, 3,* 275-285.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45,* 153-219.

Au, T. K., Chan, C. K. K., Chan, T., Cheung, M. W. L., Ho, Y. Y. S., & Ip, G. W. M. (2008). Folkbiology meets microsbiology: A study of conceptual and behavioral change. *Cognitive Psychology, 57,* 1-19.

Backscheider, A., Shatz, M., & Gelman, S. A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development, 64,* 1242-1257.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49,* 5-28.

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84,* 1-53.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64,* 3-26.

Bechara, A., Damasio, A., Damasio, H., & Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7-15.

Bliesnezer, R., Wills, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology, 2,* 247-265.

Boyer, T. W. (2006). The development of risk-taking: A multi-perspective review. *Developmental Review, 26,* 291-345.

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 (1),* 114-122.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15,* 118-121.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methdology of frontal and executive function.* Hove, East Sussex: Psychology Press.

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

Carlson, S. M. (2003). Executive function in context: Development measurement, theory, and experience. *Monographs of the Society for Research in Child Development, 68,* 138-151.

Damon, W. (1989). Patterns of change in children's social reasoning: A two-year longituidinal study. *Child Development, 51,* 1010-1017.

Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development, 5,* 151-176.

Deák, G., & Bauer, P. J. (1996). The dynamics of preschoolers' categorization choices. *Child Development, 67,* 740-767.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107,* 284-294.

DeLoache, J. S., & Burns, N. N. (2004). Early understanding of the representational function of pictures. *Cognition, 52,* 83-110.

DeLoache, J. S., & Chiong, C. (2009). Babies and baby media. *American Behavioral Scientist, 52,* 1115-1135.

Denckla, M. B. (1994). Measurement of executive function. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp.117-142). Baltimore: Brookes.

Dennis, M. (1991). Frontal lobe functioning in childhood and adolescence: A heuristic for assessing attention regulation, executive control, and the intentional states important for social discourse. *Developmental Neuropsychology, 7,* 327-358.

Ellefson, M. R., Blagrove, E., & Espy,K. A. (In preparation). The Shape School extended for older children, adolescents and adults.

Ellefson, M. R., Brinker, R. A., Vernacchio, V. J., & Schunn, C. D. (2008). Design-based learning for biology. *Biochemistry and Molecular Biology Education, 36(4),* 292-298.

Elliott, C. D. (1996). *The British Ability Scales* (2nd edition). Windsor: NFER-Nelson Publishing Company Ltd.

Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology, 13,* 495-499.

Espy, K. A., Bull, R., Martin, J., & Stroup, W. (2006). Measuring the development of executive control with the shape school. *Psychological Assessment, 18,* 373-381.

Fisher, A. V. (2010). What’s in a name? How rocks and stones are different from dogs and puppies. *Journal of Experimental Child Psychology, 105,* 198-212.

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.

Fisher, A. V., Matlen B. J., & Godwin, K. E. (2011). Semantic similarity of labels and inductive generalization: Taking a second look. *Cognition, 118,* 432-438.

Ganea, P. A., Allen, M., Butler, L., Carey, S., & DeLoache, J. S. (2009). Toddlers’ referential understanding of pictures. *Journal of Experimental Child Psychology, 104,* 283-295.

Ganea, P. A., Ma., L., & DeLoache, J. S. (2010). Young children’s learning and transfer of biological information from picture books to real animals. *Child Development.*

Garon, N., & Moore, C. (2004). Complex decision-making in early childhood. *Brain and Cognition, 55,*158-170.

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., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: categories and inferences in 2-year old children. *Developmental Psychology, 26,* 796-804.

Gelman, S. A., & Gottfried, G. (1996). Causal explanations of animate and inanimate motion. *Child Development, 67,* 1970-1987.

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. *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.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10,* 277-300.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 26,* 295-320.

Goswami, U. (1992). *Analogical reasoning in children.* Hillsdale, NJ: Lawrence Erlbaum.

Goswami, U. (2001). Analogical reasoning in children. In D. Gentner, K. J. Holyoak & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 437-470). Cambridge, MA: MIT Press.

Goswami, U., & Brown, A. L. (1990). Higher-order structure and relational reasoning: Contrasting analogical and thematic relations. *Cognition, 36,* 207-226.

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., Kilbreath, C. S., & Welder, A. N. (2001). Words and shape similarity guide 13 month-olds' inferences about nonobvious object properties. In J. D. Moore and K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 352-357). Hillsdale, NJ: Erlbaum.

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.

Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: two tasks are harder than one. *Journal of Cognition and Development, 4,* 275-298.

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.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50,* 171-188.

Hatano, G., & Inagaki, K. (1999). A developmental perspective on informal biology. In D. L. Medin & S. Atran (Eds.), *Folkbiology.* (pp.321-354). Cambridge, MA, US: The MIT Press.

Hatano, G., & Inagaki, K. (2002). Domain-specific constraints of conceptual development. *International Journal of Behavioral Development, 24(3),* 267-275.

Hayes, B. K., & Thompson, S. P. (2007). Causal relations and feature similarity in children's inductive reasoning. *Journal of Experimental Psychology*, *136 (3),* 470-484.

Heit, E., & Hayes, B. K. (2005). Relations between categorization, induction, recognition, and similarity: Comment on Sloutsky and Fisher (2004). *Journal of Experimental Psychology: General, 134,* 596-605.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2,* 411-422.

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental neuropsychology, 28,* 617-644.

Inagaki, K., (1990). The effects of raising animals on children's biological knowledge. *British Journal of Developmental Psychology, 8,* 119-129.

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development, 67(6),* 2823-2840.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Jipson, J., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development, 78,* 1675-1688.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 924-941.

Jones, S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8,* 113-139.

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.

Kerr, A., & Zelazo, P.D. (2004). Development of "hot" executive function: The children's gambling task. *Brain and Cognition, 55,* 148-157.

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.

Kripke, S. (1971). Identity and necessity. In M. K. Munitz (Ed.), *Identity and individuation.* New York: New York University Press.

Kripke, S. (1972). Naming and necessity. In D. Davidson and G. Harman (Eds.), *Semantics of natural language.* Dordrecht: D. Reidel.

Lezak, M. D. (1993). Newer contributions to the neuropsychological assessment of executive functions. *Journal of Head Trauma Rehabilitation, 8,* 24-31.

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica*, 60, 193-210.

Lopez, A., Atran, S., Coley, J., Medin, D., & Smith, E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32,* 251-295.

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

Markman, E. M. (1989). *Categorization and naming in children: problems in induction.* Cambridge: Bradford Book/MIT Press.

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

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4,* 310-322.

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

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive.* London: Longmans.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Morrison, R. G., Doumas, L. A. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science, 14,* 516-529.

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development, 75,* 1850-1870.

Newman, G. E., & Keil, F. C. (2008). Where is the essence? Developmental shifts in children's beliefs about internal features. *Child Development, 79(5),* 1344-1356.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339-363.

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

Opfer, J. E., & Gelman, S. A. (2001). Children's and adults'models for predicting teleological action: The development of a biology-based model. *Child Development, 72,* 1367-1381.

Osborne, J. (2002). Notes on the use of data transformations. *Practical Assessment, Research & Evaluation, 8.* Retrieved June 15th, 2011 from http://PAREonline.net/getvn. asp?v=8&n=6.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97,* 185-200.

Osherson, D. N., Stern, J., Wilkie, O., Stob, M., & Smith, E. E. (1991). Default probability. *Cognitive Science, 15,* 251-269.

Pallant, J. (2007). SPSS survival manual. U.K.: Open University Press.

Pasalich, D. S., Livesey, D. J., & Livesey, E. J. (2010). Performance on stroop-like assessments of inhibitory control by 4- and 5-year-old children. *Infant and Child Development, 19,* 252-263.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching, 2(3),* 176-186.

Piaget, J., & Inhelder, B. (1951). *The origin of the idea of chance in children.* New York: Norton.

Preissler, M., & Bloom, P. (2007). Two-year-olds appreciate the duel nature of pictures. *Psychological Science, 18,* 1-2.

Preissler, M. A., & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development, 5,* 185-212.

Putnam, H. (1970). Is semantics possible? In H. E. Kiefer and M. K. Munitz (Eds.), *Language, belief, and metaphysics.* State University of New York Press.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P.Rabbitt (Ed.), *Methodlogy of frontal and executive function* (pp. 1-38). Hove, UK: Psychology Press.

Rakison, D. H. (2000). When a rose is just a rose: The illusion of taxonomies in infants' categorization. *Infancy, 1,* 77-90.

Rakison, D. H., & Hahn, E. (2004). The mechanisms of early categorization and induction: smart or dumb Infants? In R. Kail (Ed.), *Advances in child development and behavior* (pp. 281-322). New York: Academic Press.

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.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94,* 249-273.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665-681.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its changes in the course of development. *Child Development, 75,* 1387-1401.

Rosengren, K. S., Gelman, S. A., Kalish, C. W., & McCormick, M. (1991). As time goes by: Children's early understanding of growth in animals. *Child Development, 62,* 1302-1320.

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72,* 103-122.

Shallice, T. (1982). Specific impairments of planning. *Transactions of the Royal Society of London ,298,* 199-209.

Sikora, D. M., Haley, P., Edwards, J., & Butler, R. W. (2002). Tower of London test performance in children with poor arithmetic skills. *Developmental Neuropsychology, 21,* 243-254.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231-280.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science, 34,* 1244-1286.

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., & Fisher, A. V. (2005). Similarity, Induction, Naming, and Categorization (SINC): Generalization or verbal inductive reasoning? Response to Heit and Hayes. *Journal of Experimental Psychology: General, 134,* 606-611.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18 (2),* 179-185.

Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology, 35,* 1478-1492.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels and the development of inductive inference. *Child Development, 72,* 1695-1709.

Sloutksy, V. M., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74,* 822-833.

Springer, K. (1992). Children’s awareness of the biological implications of kinship. *Child Development, 63,* 950-959.

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

Strouse, G. A., & Troseth, G. L. (2008). “Don’t try this at home”: Toddlers’ imitation of new skills from people on video. *Journal of Experimental Child Psychology, 101,* 262-280.

Tarlowski, A. (2006). If it’s an animal it has axons: Experience and culture in preschool children’s reasoning about animates. *Cognitive Development, 21,* 249-265.

Taylor, M. G., & Gelman, S. A. (1993). Children’s gender and age-based categorization in similarity and induction tasks. *Social Development, 2,* 104-121.

Thibaut, J-P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology, 106,* 1-19.

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children’s use of video as a source of socially relevant information. *Child Development, 77,* 786-799.

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology, 46,* 220-229.

Walker, L. J. (1989). A longitudinal study of moral reasoning. *Child Development, 60,* 157-166.

Walker, L. J., & Taylor, J. H. (1991). Stage transitions in moral reasoning: A longitudinal study of development processes. *Developmental Psychology, 27,* 330-337.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds’ expectations for novel words. *Cognition, 70,* B35-B50.

Waxman, S. R., & Booth, A. E. (2003). The origins and evolution links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science, 6,* 130-137.

Waxman, S. R., Lynch, E. B., Casey, K. L., & Baer, L. (1997). Setters and Samoyeds: The emergence of subordinate level categories as a basis for inductive inference. *Developmental Psychology, 33,* 1074-1090.

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72,* 1653-1673.

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.

Welsh, M. C., Pennington, B. F., & Grossier, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children? *Developmental Neuropsychology, 7,* 131-149.

White, C. S., & Alexander, P. A. (1986). Effects of training on four-year olds' ability to solve geometric analogy problems. *Cognition and Instruction, 3,* 261-268.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85,* 223-250.

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

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.

Appendices

Appendix A. Examples of images used in Studies 1-6.

Example images used in Studies 1 and 2

Sandbugs and Rockbugs:

Example images used in Studies 3, 4 and 5

Rocky bugs and Desert bugs (Studies 3 and 4); Sandbugs and Rockbugs (Study 5):

Country trudges and City trudges (Study 3):

Example images used in Study 6

Sandbugs and Rockbugs:

Forest weevils and Ground weevils:

Chinchillas and Squirrels:

Common frogs and Tree frogs:

Appendix B. The hidden properties used in Experiments 1, 2, 4 and 5.

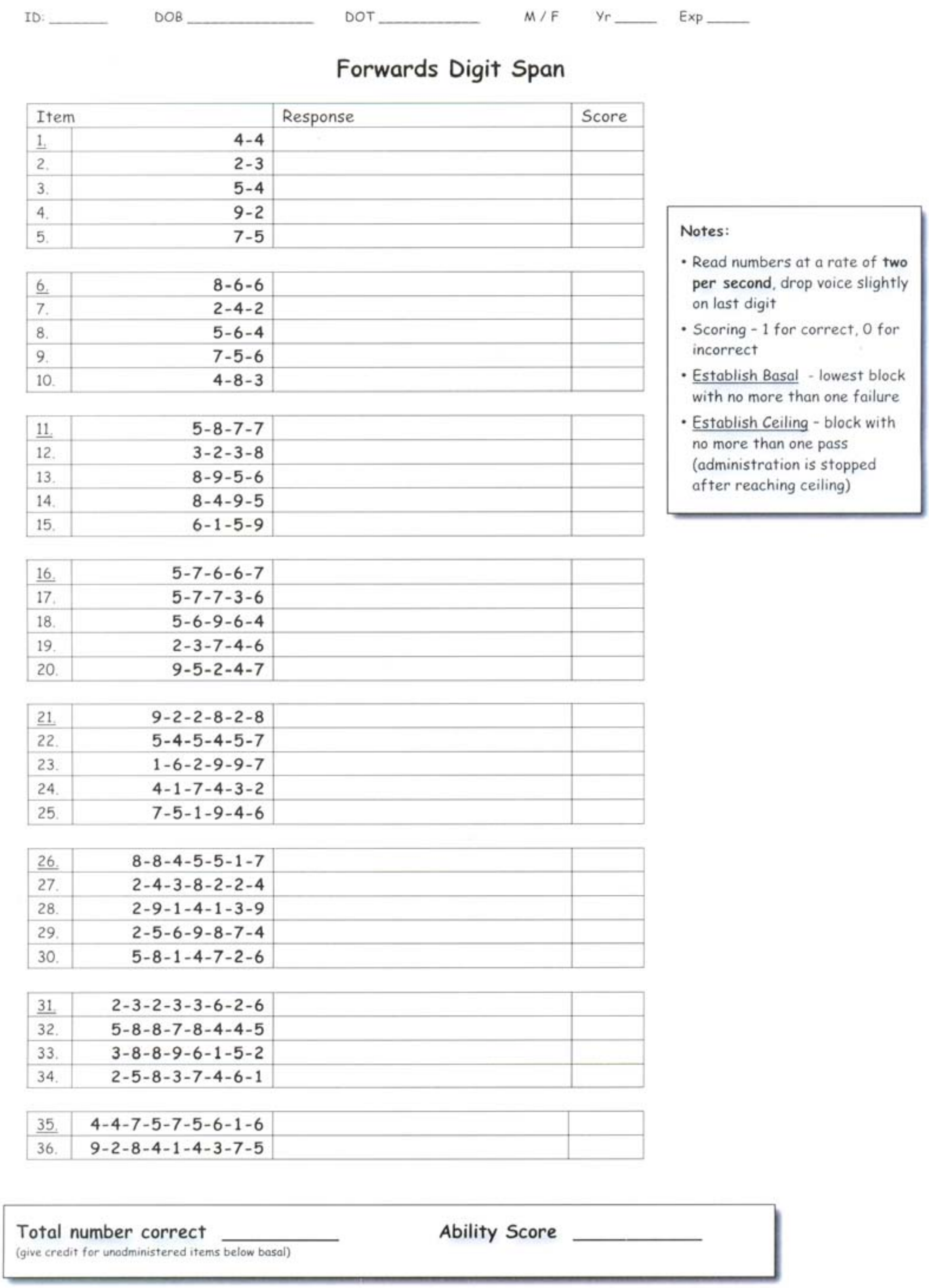
NATURAL KINDS
Has thick blood
Eats flies
Comes from America
Has cold blood
Has soft bones
Moves slowly
Sleeps during the day
Has a good sized heart
Likes to be warm
Doesn’t eat in winter
Lays eggs
Has a diminutive stomach

Appendix C. The hidden properties used in Experiment 3.

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

Appendix D. The forwards and backwards Digit Span digits



ID: _______ DOB _______________ DOT ____________ M / F Yr _____ Exp _____

Forwards Digit Span

Item		Response	Score
1.	4-4		
2.	2-3		
3.	5-4		
4.	9-2		
5.	7-5		
6.	8-6-6		
7.	2-4-2		
8.	5-6-4		
9.	7-5-6		
10.	4-8-3		
11.	5-8-7-7		
12.	3-2-3-8		
13.	8-9-5-6		
14.	8-4-9-5		
15.	6-1-5-9		
16.	5-7-6-6-7		
17.	5-7-7-3-6		
18.	5-6-9-6-4		
19.	2-3-7-4-6		
20.	9-5-2-4-7		
21.	9-2-2-8-2-8		
22.	5-4-5-4-5-7		
23.	1-6-2-9-9-7		
24.	4-1-7-4-3-2		
25.	7-5-1-9-4-6		
26.	8-8-4-5-5-1-7		
27.	2-4-3-8-2-2-4		
28.	2-9-1-4-1-3-9		
29.	2-5-6-9-8-7-4		
30.	5-8-1-4-7-2-6		
31.	2-3-2-3-3-6-2-6		
32.	5-8-8-7-8-4-4-5		
33.	3-8-8-9-6-1-5-2		
34.	2-5-8-3-7-4-6-1		
35.	4-4-7-5-7-5-6-1-6		
36.	9-2-8-4-1-4-3-7-5		

Notes:

- Read numbers at a rate of **two per second**, drop voice slightly on last digit
- Scoring - 1 for correct, 0 for incorrect
- Establish Basal - lowest block with no more than one failure
- Establish Ceiling - block with no more than one pass (administration is stopped after reaching ceiling)

Total number correct ___________ Ability Score ___________
(give credit for unadministered items below basal)

ID: _______ DOB _______________ DOT ____________ M / F Yr _____ Exp _____

Backwards Digit Span

Item		Response	(correct)	Score
Example A	5-6		6-5	
1.	5-2		2-5	
2.	4-6		6-4	
3.	5-8		8-5	
4.	1-3		3-1	
5.	7-6		6-7	
6.	4-5-5		5-5-4	
7.	1-7-6		6-7-1	
8.	8-4-5		5-4-8	
9.	8-2-4		4-2-8	
10.	3-2-9		3-2-9	
11.	6-4-4-3		3-4-4-6	
12.	2-6-1-9		9-1-6-2	
13.	6-9-8-6		6-8-9-6	
14.	9-7-5-8		8-5-7-9	
15.	3-4-5-2		2-5-4-3	
16.	7-7-9-3-8		8-3-9-7-7	
17.	2-4-1-1-5		5-1-1-4-2	
18.	1-4-5-9-8		8-9-5-4-1	
19.	3-5-2-6-9		9-6-2-5-3	
20.	4-8-6-2-3		3-2-6-8-4	
21.	6-7-7-8-1-1		1-1-8-7-7-6	
22.	4-9-7-3-2-6		6-2-3-7-9-4	
23.	5-3-2-9-1-3		3-1-9-2-3-5	
24.	2-4-3-2-9-8		8-9-2-3-4-2	
25.	7-1-5-6-9-6		6-9-6-5-1-7	
26.	4-9-2-2-3-1-7		7-1-3-2-2-9-4	
27.	9-4-3-2-8-6-1		1-6-8-2-3-4-9	
28.	8-3-9-7-1-5-4		4-5-1-7-9-3-8	
29.	4-1-2-7-3-1-6		6-1-3-7-2-1-4	
30.	2-5-9-3-4-9-5		5-9-4-3-9-5-2	

Notes:

- Read numbers at a rate of **two per second**, drop voice slightly on last digit.
- Scoring - 1 for correct, 0 for incorrect
- Establish Basal - lowest block with no more than one failure
- Establish Ceiling - block with no more than one pass (administration is stopped after reaching ceiling)

Total number correct ___________ **Ability Score** ___________

(give credit for unadministered items below basal)

Appendix E. The factors considered for matching schools

School, Group	Ofsted grade	English average: national average = 80%	Math average: national average = 79%	Science average: national average = 88%	EAL	SEN	Ethnicity	Free school dinners	Achieve at least expected (out of 300)	Absence: national average = 5.5%
1, 1	2	88%	84%	100%	-	Well below average	Majority white British	Well below average	272	5.7%
2, 1	1	92%	87%	97%	Very few	Below average	Majority white British	Below average	277	4.6%
3, 1	1	86%	81%	95%	-	Average	-	Well below average	262	3.8%
4, 2	1	93%	93%	100%	-	Below average	Majority white British	-	287	4%
5, 2	1	91%	81%	93%	Very few	Below average	Majority white British	Below average	265	3.5%
6, 2	2	91%	91%	96%	-	Average	Majority white British	Below average	278	4.3%

* Group 1 = Comparison; Group 2 = Intervention

** Ofsted 1 = outstanding; Ofsted 2 =good

*** English, Mathematics and Science: results achieved by schools on the Key Stage 2 national curriculum tests. Children achieving their expected grade.

Appendix F. The hidden properties used in Experiment 6.

Sandbugs
Eats flies
Comes from Europe
Has cold blood
Has soft bones
Moves slowly
Has a diminutive stomach

Rockbugs
Has thick blood
Sleeps during the day
Has an ample sized heart
Likes to be warm
Doesn’t eat in winter
Lays eggs

Forest weevils
Likes hot weather
Eats wood from trees
Is strong
Comes from Canada
Moves fast
Lays 50 eggs at a time

Ground weevils
Can swim
Only lives one year
Communicates through vibrations
Lays eggs in nests
Eats grubs
Sleeps during the winter

Tree frogs
Hunts through the night
Have nine vertebrae
Can be toxic
Lays eggs in clusters
Changes colour to camouflage
Comes from Australia

Common frogs
Has weak teeth
Doesn’t drink
Has no diaphragm
Three-chambered heart
Has no ribs
Breathes through its skin when under water

Chinchillas
Sleeps during the day
Comes from America
Lives 10-15 years
Doesn’t drink much water
Has dust-baths
Doesn’t smell

Squirrels
Can run fast
Is blind when young
Is active during the day
Has a great sense of smell
Can jump long distances
Sleeps during the winter