Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play

Laura E. Schulz[1], Holly R. Standing[2], and Elizabeth B. Bonawitz[1]

[1]Massachusetts Institute of Technology

[2] Cambridge University

This research was supported by a McDonnell Foundation Collaborative Initiative Causal Learning grant and a John Templeton Foundation grant to L.S. Thank you to Noah Goodman, and Rebecca Saxe for helpful comments and suggestions. We are also very grateful to the Boston Museum of Science Discovery Center and participating parents and children. Address for correspondence: L. Schulz, 46-4011, MIT Department of Brain and Cognitive Sciences, 77 Massachusetts Avenue, Cambridge, MA 02139. Phone: 617-253-7957. E-mail: lschulz@mit.edu

Abstract

Previous research (e.g., Gelman & Markman, 1986; Gopnik & Sobel, 2000) suggests that children can use category labels to make inductive inferences about non-obvious causal properties of objects. However, such inductive generalizations can fail to predict objects' causal properties when A) the property being projected varies within the category; B) the category is arbitrary (e.g., things smaller than a bread box), or C) the property being projected is due to an exogenous intervention rather than intrinsic to the object kind. In four studies, we show that preschoolers (mean: 48 months; range: 42-57 months) are sensitive to these constraints on induction and selectively engage in exploration when evidence about objects' causal properties conflicts with inductive generalizations from the objects' kind to their causal powers. This suggests that the exploratory actions children generate in free play could support causal learning.

Keywords: inductive inference, preschoolers, causal learning, categorization, exploratory play

Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play

Few ideas in developmental psychology are more widely accepted than the idea that children learn through play. However, despite a broad consensus that children "construct" knowledge (and particularly causal knowledge) by actively exploring their environments (Bruner, Jolly, & Sylva, 1976; Piaget, 1952; Singer, Golinkoff, & Hirsh-Pasek, 2006), relatively little is known about how children's exploratory play might support accurate learning. The majority of research on children's play is descriptive rather than experimental and much of the seminal work on exploratory play predates contemporary analyses of children's causal reasoning (Berlyne, 1954; 1960; Bruner, 1973; Dember, 1957; Henderson & Moore, 1980; Hutt & Bhavnani, 1972; McCall, 1974; Pavlov, 1927; Piaget 1951; 1952). Moreover, there is considerable evidence that children are poor at designing informative experiments (Chen & Klahr, 1999; Inhelder & Piaget, 1958; Koslowski, 1996; Kuhn, 1989; Kuhn, Amsel, & O'Laughlin, 1988; Masnick & Klahr, 2003; Mayer, 2004). Such findings pose a challenge for accounts of cognitive development that emphasize the role of children's active exploration. If children's exploratory play is largely unsystematic, how might children generate the type of evidence that could support accurate causal learning?

Here we investigate the hypothesis that children's spontaneous exploratory play is affected by some of the same principles of rational inductive inference that support children's causal judgments. Specifically, we look at the extent to which children's exploratory play is sensitive to the relationship between an entity's category membership and its causal properties. If children selectively engage in exploratory play when evidence about objects' causal properties conflicts with inductive generalizations from the objects' kind to their causal powers, then children's exploratory actions could support causal discovery.

In everyday reasoning, we routinely use information about an object's kind to make inferences about its unobserved properties. If we think for instance, that a new device is a cell phone, we infer that it can do different things than if we think the new object is a TV remote or a garage door opener. Although both adults and children also rely on perceptual cues, including an entity's shape, color, and texture, to extend inferences about unobservable internal states and causal powers (Baldwin, Markman, & Melartin, 1993; Gelman & Markman, 1986; Gopnik & Sobel, 2000), contra Piaget (1952), very young children are not limited to making inferences based only on superficial perceptual properties; even preschoolers make inductive inferences about objects' causal properties on the basis of their kind membership. Indeed, in induction tasks, preschoolers will extend inferences about unobservable properties of objects on the basis of category labels rather than on the basis of perceptual cues, and in categorization tasks, children will use causal, rather than perceptual information to make category judgments (Gelman & Coley, 1990; Gelman & Markman, 1986; Gopnik & Nazzi, 2003; Gopnik & Sobel, 2000; Keil, Smith, Simons, & Levin, 1998; Nazzi & Gopnik, 2000; Simons & Keil, 1995). Children who are taught, for instance, that gold melts and clay burns infer that a substance that looks like clay but is called gold will melt rather than burn (Gelman & Markman, 1986). Similarly, children who know that a red block called a 'blicket' lights up a toy, will infer that a blue block called a blicket will light the toy and that a red block which is not a blicket, will not (Gopnik & Sobel, 2000).

One of the benefits of inferring causal properties from information about the objects' kind is that it obviates the need to learn the objects' causal properties by trial and error. If we know for instance, that blickets activate a toy and that these blocks are blickets, we can infer that these blocks will activate the toy without testing them. However, inductive inferences are only true in probability. If we subsequently observe evidence that these blocks *fail to* activate the toy -- that is, we get evidence against the inductive generalization – then exploratory learning may be advantageous: perhaps not all blickets activate the toy, perhaps these blocks are not really blickets, or perhaps there are some subtle features that differentiate those blickets that do activate the toy from those that do not. Further exploration may be needed to distinguish these possibilities and to discover the causal properties of any individual token of the type. Thus in specifying the conditions under which we can infer the causal properties of entities without trial and error learning, principles of inductive inference can also provide an account of the conditions under which exploration is rational. In particular, if children's exploratory play is rational, they should engage in exploratory learning when there is a conflict between their inductive predictions and observed evidence.

When might there be a conflict between inductive generalizations from an object's kind to its causal properties and subsequent evidence? As philosophers have long observed (e.g., Goodman, 1955; Hume, 1748), inductive inferences work best to the extent that nature is uniform. When all or most members of a category share properties, then knowing the causal properties of one member of the kind supports a strong inference about the properties of others. However, if category members vary widely in their causal powers (e.g., only some blickets activate toys) then knowing the causal status of one category member may not tell you much about the properties of any other. In Experiments 1 and 2, we look at whether children's exploratory behavior is sensitive to the fact that within-category variability limits the scope of inductive generalizations. We predict that children will engage in more exploration of the target property when causal properties of objects vary within a kind (i.e., there is a conflict between the inductive generalization and the observed evidence) than when causal properties vary between kinds (i.e., the evidence is consistent with the inductive generalizations).

Experiment 3 and 4 extend this prediction to look at whether children are sensitive to the facts that some categories, and some properties, support inductive inference better than others. In particular, natural kind categories (animals, vegetables, and minerals) permit rich inductive inferences; indeed, philosophers have suggested that natural kind categories "carve nature at its joints" (see e.g., Boyd, 1999; Kornblith, 1993; Putnam, 1975; Quine, 1969), picking out stable features of the environment that correspond to genuine causal structures in the world. Nonetheless, inductive inference is not limited to natural kinds; we can also make inductive generalizations about artifacts (Apple computers), social groups (Chicago Cubs fans), and fictional entities (unicorns). Researchers have suggested that such everyday concepts play a role in our folk theories and therefore may share with natural kind concepts, a rich, internal coherence (Bloom, 1996; Carey, 1985; Carey, in press; Gopnik & Meltzoff, 1997; Keil, 1989; Laurence & Margolis, 1999; Murphy, 2002; Murphy & Medin, 1985; Sloman, Love, & Ahn, 1998).

However, not all categories support strong inductive inferences. As noted, both adults and children use perceptual cues to make inferences about an entity's unobserved properties -- but perceptual cues are sometimes misleading. Entities that look very similar (e.g., baking soda and sugar) can belong to different categories and have different unobserved properties; conversely, entities that look quite different from one another (small and large breeds of dogs) can belong to a common category and have the same unobserved properties. Thus researchers have suggested that category groupings dependent on perceptual features (e.g., "white bears") might provide a weaker basis for inductive inference than genuine categories (e.g., "polar bears"; Carlson & Pelletier, 1995; Prasada, 2000; Prasada & Dillingham, 2006).

Moreover, arbitrary, ad hoc categories (e.g., 'non-ravens', 'things smaller than a bread box', etc.) do not seem to support any inferences at all beyond the information given; knowing that something is not a raven may not allow you to conclude much else about it. Thus although we expect members of a genuine category to have causal properties in common, we might suspend this assumption for arbitrary categories. That is, we might accept that causal properties can vary within an arbitrary category precisely because an arbitrary kind may collapse across distinct genuine kinds.

Experiment 3 looks at whether children's exploratory play is sensitive to the distinction between genuine, perceptually based, and arbitrary categories. We predict that children will explore more when there is casual variability within a genuine category than when there is causal variability within an arbitrary category (where the evidence is potentially consistent with inductive generalizations from genuine kind differences). In the case of evidence that contradicts (relatively weak) inductive generalizations from perceptually based categories, we suggest that children may explore more than they do for arbitrary kinds but less than for genuine kinds.

Finally, some properties support inductive projections better than others. Perhaps the most famous property that fails to license induction is the property of 'grue' -- the property of being green until some future time *t* and blue thereafter. As the philosopher Nelson Goodman observed, it seems rational to use the evidence of emeralds seen so far to conclude that all emeralds are green but, although equally consistent with the evidence, it does not seem rational to conclude that all emeralds are grue (1955). Philosophers (e.g., Quine, 1969) have explained this discrepancy by suggesting that properties may only be projectible to the extent that they track stable causal relations in the world (much as natural kind categories support induction to the extent that they unite entities that share real causal relationships in the world). One implication of this account is that a causal property will not be projectible if the property is not a stable property of the kind but is instead due to an arbitrary intervention. If for instance, you rub a balloon to create static cling, it will have the property of sticking to walls without licensing the inference that other (un-rubbed) balloons will stick to walls. That is, if a property is not an intrinsic property of a kind but is instead due to an arbitrary intervention, the property may not be projectible to other members of the kind. Experiment 4 looks at whether children's exploratory play is sensitive to the distinction between intrinsic and extrinsic properties of object kinds; we predict that children will explore more when an intrinsic property of a kind varies within the category (and there is thus a conflict between the inductive generalization from object kind to causal properties and the observed evidence) than when a property incidental to the kind varies within the category (and thus any discrepancy between the evidence and the inductive generalization can be explained by the arbitrary intervention)[1].

In summary, the studies to follow look at how conflicts between the predictions of inductive inference and observed evidence affect children's exploratory play (see Table 1 for an overview of the experiments and predicted results). Throughout, our dependent measure is based on children's exploratory play with objects in which the target causal property is absent. By investigating children's play with inert objects, we can eliminate the distraction of the interesting causal property itself and focus specifically on how children's predictions affect their play. We suggest that children will spontaneously engage in more exploratory actions when evidence about objects' causal properties conflicts with their inductive generalizations than when the absence of the objects' causal property is potentially consistent with such generalizations. If children do engage in this selective exploration, then children's exploratory actions could support causal discovery specifically in contexts where other inductive cues to an object's causal powers are unreliable.

Experiment 1

As noted, considerable research has looked at the relationship between children's knowledge about object categories and their inferences about unobserved properties (Disendruck & Gelman, 1999; Gelman & Markman, 1986; Gopnik & Nazzi, 2003; Gopnik & Sobel, 2000; Keil, Smith, Simons, & Levin, 1998; Nazzi & Gopnik, 2000; Simons & Keil, 1995). However, such studies have looked primarily at whether children will extend inductive inferences from one token of a type to another. Relatively few studies (though see Gelman & Markman, 1986; Gelman, Star, & Flukes, 2002 and Gelman & Raman, 2003, for exceptions) have looked at the scope of children's inferences (e.g., at the extent to which children believe the property applies generically, to *all* tokens of the type). In this experiment, we look at how children's assumption that a causal property extends across all the members of a category affects their exploratory play.

We introduce children to a set of blocks called “blickets”. Children learn that these blocks are magnetic. They are then introduced to second set of blocks. Half the children are told that the blocks in the second set are “blickets”; half are told that they are “dax”. When all of the blocks belong to the same category children should assume that the new blocks are magnetic. If the children try one block from the second set and find that it fails to stick (i.e., the evidence conflicts with their inductive inferences) they should explore the other blocks. However, when the second set of blocks belongs to a different category, finding that a single member of that category is inert should suffice for children to infer that all the members of the category are inert; the evidence is consistent with their inductive inferences and children should be less likely to explore the novel blocks.

Method

Participants

Thirty-two three and four-year-olds (47% girls; range: 42 months to 57 months, mean: 48 months) were recruited for the study. Half the children were randomly assigned to a One Kind condition, half to a Two Kind condition. Throughout these studies, children were recruited from visitors to a metropolitan science museum. Most of the children were from white, middle-class backgrounds but a range of ethnicities representing the diversity of the population was involved.

Materials

Two sets of cylindrical purple blocks (10 cm tall; 6 cm diameter) with yellow ends were used. One set consisted of five blocks on which the yellow end of each block was magnetic. The second set consisted of ten inert blocks. A 30 cm by 30 cm magnet board was set upside down on four legs (60 cm tall) so that if the magnetic blocks were stuck to the bottom of the board, they would hang suspended. Inert blocks would of course fall from the board.

Procedure

The children were tested individually in a quiet corner of the museum. The experimenter placed the first set of five blocks on the table. (See Figure 1a for stimuli and design). She said, "See these? These are blickets. Watch this!" She then stuck one of the blocks to the magnet board so it hung suspended. The child was then allowed to try. After the child had stuck all five blocks to the board, the experimenter poured the second set of ten blocks onto the table. In the "One Kind" condition she said, "And here are some blickets"; in the "Two Kind" condition she said, "And here are some dax". She then told the child, "I have to go write something down for a minute. You go ahead and play." The experimenter walked out of the child's sight. Children were allowed to play freely with both sets of blocks for up to 60 seconds or until they turned away from the play table. The free play session was videotaped for coding. Only children who played for at least 30 seconds were included in the study; all children met the inclusion criteria. At the end of the trial, the experimenter returned, thanked the child, and terminated the study.

Results and Discussion

Throughout, all values reported as significant have p-values of .05 or less; two-tailed tests are used except when noted; results reported as non-significant are non-significant by both one and two-tailed tests. We analyzed children's play in three ways. We looked at children's overall play time, the number of blocks from the second (inert) set that children tried to stick to the board, and the number of children who tried to stick more than a single block from the inert set to the board. Because we could not be sure that the number of blocks children would try would be normally distributed, we use non-parametric (Mann-Whitney U tests) rather than parametric tests (t-tests) throughout, providing a conservative test of the hypotheses. Children played for the same amount of time in both conditions (One Kind, mean: 52.06 seconds; Two Kind, mean: 54.56 seconds, $t(30) = 1.01$, $p = ns$)[2]. However, children tried to stick significantly more of the blocks from the second, inert set to the board in the One Kind condition than in the Two Kind condition (Mann-Whitney U = 52, N = 32). None of the children in the One Kind condition tried only a single block from the second set whereas half of the children (8 of 16) in the Two Kind condition tried only a single block from the second set (χ^2 (1, N = 32) = 10.67). See Figure 2.

These results suggest that children readily accept that members of different categories can have different causal properties but explore extensively when causal properties vary within a category. Children were much more likely to try to stick individual, as yet untested, blocks to the board when they believed that all the blocks were 'blickets' than when they believed that the blocks in the first set were 'blickets' and the blocks in the second set were 'dax'.

Note that on a process level, such differential exploration could occur for at least four different reasons: 1) children might selectively explore other blickets because the evidence of the inert blocks is insufficient to overturn their expectation that other blickets will be magnetic; 2) children might selectively explore other blickets because the evidence of the inert blocks indicates that the causal status of each blicket is uncertain, thus children might be motivated to discover the causal properties of each token of the type; 3) children might selectively explore other blickets because they want to discover the factors that distinguish blickets that share the target property from those that do not, or 4) children might not explore other dax because the evidence of the inert blocks is sufficient for them to infer to infer that all the other dax will lack the target causal property. These motivations are not mutually exclusive and this experiment does not dissociate them. Here we offer a computational level account of how inductive inference might affect exploration (i.e., addressing the goals and logic of the behavior), rather than an account at the level of the representational algorithm (i.e., how this logic might be implemented; Marr, 1982). Critically, note that from a computational perspective, all four motivations result in an equivalent, adaptive outcome: if children spontaneously explore more individual objects when causal properties vary within a category than when they vary between categories, then children will be more likely to discover the causal properties of individual objects in cases where the objects' causal properties cannot be otherwise inferred.

Experiment 2

Experiment 1 suggests that children's exploratory play is affected by the assumption that members of a common category will share causal properties. However, all the children were initially introduced to five members of a category, all of which did share a causal property; that is, all five blocks in the first set were magnetic. Thus it is possible that rather than starting with the assumption that category membership is a reliable cue to the extension of causal properties, children developed this expectation during the initial phase of the experiment. In Experiment 2, we replicate Experiment 1 but start by showing children only a single magnetic block. If children assume (rather than learn in the course of the study) that category members share causal properties, the results of Experiment 1 should replicate in Experiment 2.

Additionally, children in the One Kind condition of Experiment 1 heard a single category label, "blicket", whereas children in the Two Kind condition heard two labels – "blicket" and "dax". Moreover, children in the Two Kind condition were asked to accept that these two labels applied to perceptually identical objects. To eliminate the possibility that these additional processing demands might have interfered with, and limited, children's tendency to explore the novel, second set of objects in the Two Kind condition, in Experiment 2, we introduce all the children to two labels for the objects. Finally, in Experiment 2 we refer to "blickets" and "feps" rather than "blickets" and "dax". Although in Experiment 1, the blocks were always referred to collectively as "some dax" (i.e., suggesting that a single block would be "a dack"), in principle "dax" could be both singular and plural. We thus substituted the word "feps" to eliminate any concern that the grammatical ambiguity might have confused the children.

Method

Participants

Thirty-two three and four-year-olds (56% girls; range: 42 months to 55 months, mean: 48 months) were recruited for the study. Half the children were randomly assigned to a Both Kinds condition, half to a Different Kinds condition.

Materials

The same materials used in Experiment 1 were used in this experiment except that only a single magnetic block from the first set of blocks was used.

Procedure

The procedure replicated the procedure in Experiment 1 with the following changes. In both conditions, the experimenter introduced the child to a single purple block. She said, “See this? This is a blicket. Watch this.” She then stuck the block to the magnet board so it hung suspended. The block was removed and the child was allowed to try. The experimenter then poured the second set of ten blocks onto the table. In the “Both Kinds” condition she said, “Here are some blickets and here are some feps”; in the “Different Kinds” condition she said, “Here are some feps”.

Results and Discussion

Children played for the same amount of time in both conditions (Both Kinds, mean: 54.25 seconds; Different Kinds, mean: 54.87 seconds, $t(30) = .24$, $p = ns$). However, children tried to stick significantly more blocks from the inert set to the board in the Both Kinds condition than in the Different Kinds condition (Mann-Whitney U = 223, N = 32). Children were significantly more likely to try only a single block from the second set in the Different Kinds condition than in the Both Kinds condition (χ^2 (1, N = 32) = 7.38). See Figure 2.

These results corroborate the results of Experiment 1, suggesting that children’s exploratory play is affected by their inferences about the relationship between object kinds and object properties. Indeed in this experiment, because children are told that the second set of blocks contains a mixture of both the old and new kinds, we might expect them to be particularly motivated to discover which objects are members of which kind and/or what features distinguish the kinds. The results are consistent with this. Having been shown only that a single blicket was magnetic, children seemed to expect other blickets to be magnetic; given evidence that violated this inference, children who believed that the second set of blocks contained blickets tested extensively for other magnetic blocks. However, children who believed the second set of blocks were members of a different category seemed to accept that this new category might have different causal properties; they did not engage in this exploration. This suggests that children's spontaneous exploratory play is affected by their inductive inferences about how category membership affects the extension of causal properties.

Experiment 3

Children's performance in Experiments 1 and 2 is consistent with previous work (Gelman & Markman, 1986; Gopnik & Sobel, 2000) suggesting that children will treat perceptually similar stimuli differently depending on how they are labeled. In general however, perceptual cues (like an object's shape and color) are informative about kind membership; in the absence of differential object labels, children might infer that all the purple, cylindrical blocks are members of a common kind – and therefore that they also have common causal properties. This suggests that the results of the "One Kind" and "Both Kinds" conditions of Experiments 1 and 2 should replicate even if the objects are referred to only with a descriptive category label: "purple ones".

However, as noted, perceptual cues are sometimes misleading about kind membership: entities that look alike can belong to different categories and have different unobserved properties and entities that look different can belong to a common category and have the same unobserved properties. Thus perceptually based categories, like "purple ones", might be a weaker basis for inductive inference than categories based on object labels, like "blickets". Moreover, if the perceptual categorization captures only arbitrary similarities between objects and fails to respect genuine category boundaries, inductive inferences based a perceptual category might be altogether unwarranted. Given for instance, that pigs like to play in the mud, it might be reasonable to infer that "the muddy ones" will say, "oink"; however, it would not be reasonable to make this inference after throwing mud on a chicken coop. To test whether children are sensitive to gradations in the inductive power of genuine, perceptually based, and arbitrary categories, in Experiment 3, we compare the results of the One Kind condition of Experiment 1 with two new conditions. In one condition, the Arbitrary Perceptual Kind condition, the descriptive label is clearly not informative about a genuine category distinction, whereas in the other, the Genuine Perceptual Kind condition, it is potentially informative.

Specifically, in the Arbitrary Perceptual Kind condition, children are shown a first set of purple blocks and a second set of orange blocks. The orange blocks are then covered with purple tape and reintroduced as "purple ones". Children are then shown that the first set of blocks is magnetic. Although all the objects are now perceptually indistinguishable from each other, children have reason to believe that the common appearance and common descriptive label are arbitrary; the shared perceptual properties can be ascribed to an exogenous intervention and thus do not provide inductive support for a common underlying category with common causal properties. Indeed in this condition, the arbitrary category "purple ones" now collapses across two potential genuine categories: real "purple ones" and real "orange ones". Because the causal variability between the first and second set of blocks in Experiment 3 can be ascribed to a difference in genuine kinds (and the common label disregarded as an arbitrary category), the evidence about the inert blocks should not violate the children's inductive inferences. We predict that children will explore fewer of the inert blocks in the Arbitrary Perceptual Kind condition of Experiment 3 than in the in the One Kind condition of Experiment 1.

We also compare these results with a Genuine Perceptual Kind condition, in which all the objects really are purple. Here the perceptual information is potentially informative about the objects' shared category membership and therefore about their shared unobserved properties. In this condition, children may be more likely to infer that the objects have causal properties in common than were children in the Arbitrary Kind condition; however, because perceptual cues are potentially misleading about kind membership, children may be less likely to infer that the objects share causal properties than were children in the One Kind condition. Thus we predict that the results for the Genuine Perceptual Kind condition will be intermediate between the results for the One Kind and the Arbitrary Perceptual Kind condition.

Method

Participants

Thirty-two three and four-year-olds (41% girls; range: 41 months to 57 months, mean: 48 months) were recruited for the study; half were randomly assigned to an Arbitrary Perceptual Kind condition, half to a Genuine Perceptual Kind condition.

Materials

The same materials used in Experiment 1 were used in this experiment. An additional set of 10 orange cylindrical blocks (identical to the other blocks except for color) and a roll of purple tape were also used.

Procedure

The procedure replicated the procedure in Experiment 1 with the following changes (see Figure 1b). In the Arbitrary Perceptual Kind condition, the experimenter poured 5 purple magnetic blocks on the table and then poured the 10 orange blocks into a different pile. She looked at the orange blocks and said, "Oh look." She pulled out the roll of purple tape and told the child, "I have some purple tape here. I'm going to make these look like those. Go ahead and play for a minute." The child was allowed to stack and sort the five purple blocks (the child did not yet know the blocks or board were magnetic and no child ever used the blocks in conjunction with the magnet board). While the child was playing, the experimenter pretended to cover the orange blocks with tape. (In fact, the experimenter simply substituted the orange blocks for the inert purple blocks, concealed under the table.) When she finished, the experimenter pointed to both sets of blocks and said, "See these? These are purple ones. Watch". She took one of the blocks from the first set and stuck it onto the magnet board so it hung suspended. The child was then allowed to try. After the child had stuck all five blocks from the first set to the board she said, "I have to go write something down for a minute. Go ahead and play." The children were allowed to play freely with all 15 blocks for 60 seconds.

In the Genuine Perceptual Kind condition, the experimenter poured the 5 purple magnetic blocks on the table and then poured the 10 purple inert blocks into a different pile. She pointed to both sets of blocks and said, "See these? These are purple ones. Watch". The experiment then proceeded as in the Arbitrary Perceptual Kind condition.

Results and Discussion

The planned comparisons of interest were the pair-wise comparisons between the One Kind condition of Experiment 1 and the two conditions of this experiment. Children played for the same amount of time in the One Kind condition of Experiment 1 (mean: 52.06 seconds) and the Arbitrary Perceptual Kind condition (mean: 53.81 seconds): $t(30) = 0.86$, *ns,* but played longer in the Genuine Perceptual Kind condition (mean: 55.56 seconds) than the One Kind condition, $t(30) = 2.14$, $p = 0.04$, and there was a trend for children to play more in the Genuine Perceptual Kind condition than the Arbitrary Kind condition: $t(30) = 1.85$, $p = 0.07$). The slight difference in overall playtime is likely due to the fact that children in the Arbitrary Kind condition had some prior exposure to the first set of blocks (i.e., while the second set was being 'taped').

As predicted, children in the One Kind condition of Experiment 1 tried to stick more blocks from the second, inert set to the board than children in the Arbitrary Perceptual Kind condition (one-tailed Mann-Whitney U = 174, N = 32)[3]. There was a trend for children to be more likely to try only a single block from the second set in the Arbitrary Perceptual Kind condition than the One Kind condition (χ^2 (1, N = 32) = 3.31, p = .07). See Figure 2.

By contrast, children's performance in the Genuine Perceptual Kind condition was not significantly different either from the One Kind condition or from the Arbitrary Perceptual Kind condition (Genuine Kind v. One Kind: Mann-Whitney U = 145.5, N = 32, p = *ns*; Genuine Perceptual Kind v. Arbitrary Perceptual Kind: Mann-Whitney U = 101, N = 32, p = *ns*). Children were no more likely to try a single block from the second set in the Arbitrary Perceptual Kind condition than in the Genuine Perceptual Kind condition (χ^2 (1, N = 32) = 1.03, p = ns).

These results suggest that preschoolers are sensitive to gradations in the inductive strength of different cues to kind membership. In all three conditions, preschoolers were allowed to play freely with identical objects identified by a single label. However, the children seemed to recognize both A) that common category labels generally support inductive inference about shared causal properties and B) that this inference can be suspended for arbitrary descriptive labels. Thus when all the objects were "blickets", children extensively explored variability in causal properties within the category; when all the objects were "purple ones" (but were purple only due to an exogenous intervention), the children were less likely to explore the within-category variability.

However, although children's performance in the One Kind condition and the Arbitrary Perceptual Kind condition were different from one another, children's performance in the Genuine Perceptual Kind condition was not significantly different either from their performance in the One Kind condition or the Arbitrary Perceptual Kind Condition. Although one must be cautious about drawing conclusions from this data, the results provides suggestive evidence that children might treat categories defined by perceptual cues as intermediate in inductive strength between categories designated by object labels and arbitrary categories. Such an intermediate status for perceptual cues to kind membership would be consistent with previous research, which has suggested both that children use perceptual information to extend inferences about unobserved properties (Baldwin, et al., 1993; Becker & Ward, 1991; Bowerman, 1996; Graham & Poulin-Dubois, 1999; Imai & Gentner, 1997; Jones, Smith, & Landau, 1991; Keil, 1994; Landau, et al.,1988; 1992; Smith, Jones & Landau, 1992; Soja, 1992; Soja, Carey & Spelke, 1991) -- and that children will override perceptual cues in order to extend inductive inferences on the basis of category labels (e.g., Gelman & Markman, 1986; Gopnik & Sobel, 2000).

Experiment 4

Experiment 3 looked at how children's exploratory play was affected by the inductive strength of different cues to category membership. The results suggested that children's exploratory play is rational with respect to the inference that perceptual properties intrinsic to an object can support generalizations about kind membership (and therefore about unobservable properties) and that perceptual properties that objects acquire through exogenous interventions do not support such generalizations. Experiment 4 addresses a similar question but focuses not on whether the perceptual cues to category membership are intrinsic or extrinsic to the objects, but on whether the causal property being projected is intrinsic or extrinsic to the objects.

We predict that intrinsic properties of objects will be readily projected to other members of the kind but that arbitrary causal properties (properties that attach to the objects only through exogenous intervention) will not. To test the prediction that the distinction between intrinsic and extrinsic properties of objects affects children's exploratory play, we replicate the One Kind condition of Experiment 1 but before children see that the first set of blocks is magnetic, we introduce a sham intervention on the blocks. Because children in this condition can potentially construe the magnetism as an effect of the intervention (rather than as a property of the objects themselves), they should be less likely than the children in Experiment 1 to infer that other blickets will be magnetic.

Method

Participants

Seventeen three and four-year-olds (31% girls; range: 43 months to 55 months, mean: 48 months) were recruited for the study.

Materials

The same materials used in Experiment 1 were used in this experiment. A 20 cm x 14 cm x 11 cm box was used for the sham intervention. The top of the box had a wire coat hanger "antenna" and a rotating dial (made from a Master's lock) and one side of the box opened so that objects could be placed inside.

Procedure

The procedure replicated the procedure in Experiment 1 with the following change (see Figure 1c). The experimenter placed the first set of five blocks on the table. She said, "See these? These are blickets. I'm going to put these blickets in this special box. Let's turn the dial three times and count to ten." She then introduced the box and placed all five blocks inside the box. She rotated the dial three times, and counted to 10, saying "Ping!" at the end. She then removed the blocks from the box and removed the box from the table. Although nothing was said to the children explicitly, the sham manipulation introduced the possibility that the intervention, rather than the object kind, was responsible for the magnetic properties of the first set of objects. That is, there is a pragmatic implication that the magnetism is an effect of the intervention with the box. Following the sham intervention, the experiment proceeded exactly as in the One Kind condition of Experiment 1.

Results and Discussion

We compared the results of Experiment 4 with the One Kind condition of Experiment 1. Children played for the same amount of time in both conditions (Experiment 4, mean: 55.38 seconds; Experiment 1, One Kind condition, mean: 52.06 seconds, $t(31) = 1.46$, $p = ns$). There was a trend for children to be more likely to try only a single block from the second set in Experiment 4 than in the Experiment 1, One Kind condition (χ^2 (1, N = 33) = 3.31, p = .07) and children stuck more blocks from the second, inert set to the board in the Experiment 1, One Kind condition than in Experiment 4 (one-tailed Mann-Whitney U = 186, N = 33)[4]. See Figure 2.

These results suggest that children are sensitive to constraints on projecting properties to other members of a kind. Although the results were partially a trend and therefore must be interpreted with caution, children's tendency to engage in such differential play is perhaps particularly noteworthy given the subtlety of the manipulation: children were not told that the magnetism was the result of the sham intervention, and except for the sham intervention, the One Kind condition of Experiment 1 and Experiment 4 were identical. Nonetheless, when the magnetism was unambiguously a property of the blickets, children expected other blickets to be magnetic; when the evidence was inconsistent with this prediction, they engaged in extensive exploration. However, when the magnetism was potentially due to an external manipulation, children were less likely to explore the within-kind variability. Children's exploratory play seems to be rational with respect to the inference that members of a category will share an object's intrinsic causal properties but will not necessarily share properties due to an external intervention.

General Discussion

Collectively, these studies suggest that there is a systematic relationship between rational principles of inductive inference and children's spontaneous exploratory play. Preschoolers were most likely to explore individual objects when there was a conflict between the evidence they observed and their inductive generalizations from the object's kind to the objects' casual properties. That is, children differentially engaged in targeted exploratory play in contexts where active exploration was critical to causal discovery. Such selective exploration could support learning. Children's tendency to explore objects when they cannot predict the extension of the objects' properties gives children an opportunity to discover the properties of individual objects in precisely those contexts where such properties cannot be otherwise inferred. By flexibly shifting from inductive inference to trial and error learning, children seem to be able to adapt their exploratory play to optimize their chances of discovering novel information.

The current findings are consistent with recent research suggesting that children's exploratory play is driven not only by perceptual properties of stimuli (e.g., salience and novelty) but also by formal properties of the evidence children observe. Thus for instance, preschoolers engage in more exploratory play when given confounded evidence about the causal structure of a toy than when given unconfounded evidence (Schulz & Bonawitz, 2007). In the current study, children also engaged in selective exploration of ambiguous evidence. When inductive inferences about an object's properties based on its kind membership conflicted with observed data about the object preschoolers seemed to be sensitive to the ambiguous status of other tokens of the type and extensively tested individual objects. This suggests that children are not only sensitive to the distinction between evidence that does and does not disambiguate causal relationships (see e.g., Masnick & Klahr, 2003 and Sodian, Zaitchik, & Carey, 1991), but actively explore ambiguous evidence.

Do children actually learn the relationship between object kinds and causal properties from the evidence of their own interventions? Our experiment does not address this directly. As noted, there are many different things children might discover through trial and error exploration (e.g., that not all blickets activate the toy or that there are subtle features distinguishing blickets that do activate the toy from those that do not). In simple cases like those described here, it seems probable that children do learn for instance, at least that not all members of the kind have the causal property. However, we do not want to suggest that in all cases children's free play reliably leads to accurate learning. Informative interventions might occur relatively infrequently in the course of play and children might fail to attend to the evidence they generate. Nonetheless, children's tendency to engage in selective exploration when their inductive inferences fail could be advantageous; whether or not children learn from their explorations in any particular instance, overall, children would be more likely to investigate where there is something to be learned.

Children's exploratory behavior could be informative even if it were simply an index of children's causal judgments. However, the current studies suggest that children's spontaneous exploratory actions are closely linked to rational processes of inductive inferences. Such exploratory actions provide not only useful indicators of what children already know, but also a powerful way for children to gain new knowledge. If, even when children are "just playing", they are sensitive to and selectively explore evidence that challenges inductive generalizations, then children's spontaneous play could provide an important route to causal discovery.

References

Atran, S. (1998) Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, 21, 547–68.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants ability to draw inferences about non-obvious object properties - evidence from exploratory play. *Child Development*, 64(3), 711-728.

Becker, A. H., & Ward, T. B. (1991). Children's use of shape in extending novel labels to animate objects: identity versus postural change. *Cognitive Development*, 6 (1), 3–16

Berlyne, D.E. (1954). A theory of human curiosity. *British Journal of Psychology*, 45(3), 180-191.

Berlyne, D.E. (1960). *Conflict, Arousal and Curiosity*. New York: McGraw-Hill.

Bloom, P. (1996) Intention, history, and artifact concepts. *Cognition,* 60, 1–29.

Bowerman, M. (1996). The origins of children's spatial semantic categories: cognitive versus linguistic determinants. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–176). Cambridge, MA: Cambridge University Press.

Boyd, R. (1999). Kinds, Complexity and Multiple Realization, *Philosophical Studies*, 95, 67-98.

Bruner, J.S. (1973). Organization of early skilled action. *Child Development*, 44(1), 1-11.

Bruner, J.S., Jolly, A. & Sylva, K. (1976). *Play: It's role in development and evolution*. New York: Basic Books.

Carlson, G. N. and Pelletier, F. J. (eds.) (1995). *The Generic Book*. University of Chicago Press, Chicago.

Carey, S. (in press). *The origin of concepts.* Cambridge, MA: MIT Press.

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

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098-1120.

Dember, W. N. (1957). The relation of decision-time to stimulus similarity. *Journal of Experimental Psychology*, 53(1). 68-72.

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

Gelman, S. A. (2004). Psychological essentialism in children. Trends in Cognitive Sciences, 8, 404-409.

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 (5), 796–804

Gelman, S. A., & Hirschfeld, L. A. (1999). How biological is essentialism? In S. Atran, & D. Medin (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23(3), 183-209.

Gelman, S.A. & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, 74(1), 308-325.

Gelman, S. A., Star, J., & Flukes, J. (2002). Children’s use of generics in inductive inferences. *Journal of Cognition and Development*, 3, 179 – 199.

Goodman, N. (1955). The new riddle of induction. In: *Fact, Fiction and Forecast*. Cambridge, MA: Harvard University Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. Child *Development*, 71(5), 1205-1222.

Graham, S. A., & Poulin-Dubois, D. (1999). Infants’ reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language*, 26 (2), 295–320.

Gopnik, A. & Nazzi, T. (2003). Words, kinds and causal powers: A theory theory perspective on early naming and categorization. In D. Rakison, & L. Oakes (Eds.) *Early categorization*. Oxford: Oxford University Press

Henderson, B. & Moore, S.G. (1980). Children's responses to objects differing in novelty in relation to level of curiosity and adult behavior. *Child Development*, 51(2), 457-465.

Hume, D. (1999). *An enquiry concerning human understanding*. Oxford: Oxford University Press. (Original work published 1748).

Hutt, C. & Bhavnani, R. (1972). Predictions from play. *Nature*, 237, 171-172.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. *Cognition*, 62, 169–200.

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

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

Keil, F.C., Smith, W.C., Simons, D.J., & Levin, D.T. (1997). Two dogmas of conceptual empiricism: implications for hybrid models of the structure of knowledge. *Cognition*, 65, 103-135.

Kornblith, H. (1993). *Inductive inference and its natural ground: an essay in naturalistic epistemology*. Cambridge, MA: MIT Press.

Koslowski, B. (1996). *Theory and Evidence: The Development of Scientific Reasoning*. Cambridge, MA: MIT Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4), 674-689.

Kuhn, D., Amsel, E., & O'Laughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. Cognitive Development, 3 (3), 299–321

Landau, B., Smith, L., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, 31, 807-825.

Laurence, S. & Margolis, E. (1999). “Concepts and Cognitive Science”, In: S. Laurence and E. Margolis (eds.), *Concepts, Core Readings*. Cambridge, MA: MIT Press, 3-81.

Marr, D. (1982). Vision: A computational investigation in the human representation of visual information. San Francisco, CA: Freeman.

Masnick, A. M., & Klahr, D. (2003). Error Matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development*, 4(1), 67-98.

Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition*, 78, 1–26.

Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59, 14–19.

McCall, R.B. (1974). Critique of a field (review of The Study of Behavioral Development by Joachin F. Wohlwill). *Science*, 184, 673-674.

Murphy, G.L. (2002). *The Big Book of Concepts*. Cambridge, MA: The MIT Press.

Murphy, G.L. & Medin D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 94(3), 289-316.

Nazzi, T. & Gopnik, A. (2000). A shift in children’s use of perceptual and causal cues to categorization. *Developmental Science*, 3(4), 389-396.

Pavlov, I.P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press.

Piaget, J. (1951). *Play, dreams and imitation in childhood*. New York: W. W. Norton.

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: International Universities Press.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Science*, 4, 66–72.

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

Putnam, H. (1975) *Mind language and reality; Philosophical papers, volume 2*, Cambridge: Cambridge University Press.

Quine, W. V. O. 1969. Natural kinds. In: *Ontological reality and other essays*. New York: Columbia University Press.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition*, 91, 113–153.

Schulz, L.E. & Bonawitz, E. B. (i2007) Serious Fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology 43(4), 1045-1050*.

Simons, D.J. & Keil, F.C., 1995. An abstract to concrete shift in the development of biological thought: the insides story. *Cognition* 56, 129–163.

Singer, D.G., Golinkoff, R.M. & Hirsh-Pasek, K. (2006). *Play=learning: How play motivates and enchances children's cognitive and social-emotional growth*. New York: Oxford University Press.

Sloman, S., Lombrozo, T. & Malt, B. (2007). Ontological commitments and domain-specific categorization. In M. J. Roberts (Ed.). *Integrating the Mind*. Hove, UK: Psychology Press.

Sloman, S.A., Love, B.C., Ahn, W (1998). Feature centrality and conceptual coherence. *Cognitive Science*, 22, 189-228.

Smith, L.B., Jones, S.S., & Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology*. 28,273-286.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62, 753-766.

Soja, N. N. (1992). Inferences about the meanings of nouns: the relationship between perception and syntax. *Cognitive Development*, 7, 29±45.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. *Cognition*, 38 (2), 179–211

Strevens, M. (2000). The essentialist aspect of naïve theories. *Cognition*, 74, 149-175.

Footnotes

[1.] This distinction between intrinsic and extrinsic properties is similar to the distinction some psychologists have made between essential and non-essential properties of kinds (e.g., Atran, 1998; Bloom, 1996; Gelman & Hirschfeld, 1999; Gelman, 2003; 2004; Keil, 1989; Matan & Carey, 2001). We refer to intrinsic versus extrinsic rather than essential versus non-essential because as some philosophers and psychologists have observed (Strevens, 2000; Rheder & Hastie, 2004; Sloman, Lombrozo, & Malt, 2007), the relevant distinction between projectible and non-projectible properties can be inferred by general causal reasoning processes rather than a belief in essentialism per se.

[2.] Children's overall play time reflects not only the time they spent trying to stick the inert blocks to the board but all their play (i.e., time playing with the magnetic blocks and time spent with the blocks *as* blocks -- stacking and building with them).

[3.] This result was a trend with a two-tailed Mann-Whitney U test (p = .08).

[4.] This result was a trend with a two-tailed Mann-Whitney U test (p = .07).

Table 1. Inference, evidence, and prediction for each of the experimental conditions.

Experiment and condition	Infer that new blocks will be magnetic?	Evidence	Selective exploration?
Exp 1: One kind condition	Yes	New blocks are not magnetic	Yes
Exp 1: Two kind condition	No		No
Exp 2: Both kinds condition	Yes		Yes
Exp 2: Different kinds condition	No		No
Exp 3: Genuine perceptual kind condition	Yes (weakly)		Yes (weakly)
Exp 3: Arbitrary perceptual kind condition	No		No
Exp 4: Extrinsic property condition	No		No

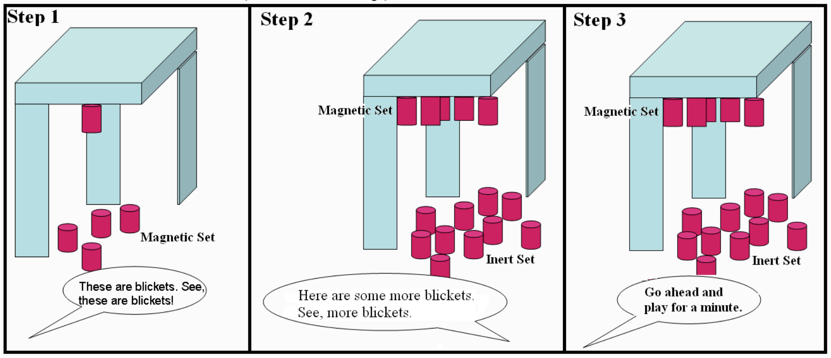
Figure Captions

Figure 1. Sketch of stimuli and procedure in Experiments 1, 3, and 4. (No figure is included for Experiment 2 or for the Genuine Perceptual Kind condition of Experiment 3 because these procedures are similar to those in Experiment 1. See text for details).

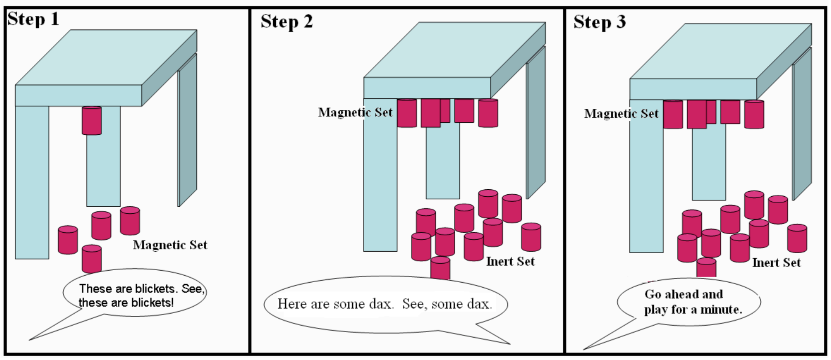
Figure 2. Mean number of inert blocks and number of children trying exactly one block in each experimental condition.

1a) Stimuli and procedure in Experiment 1

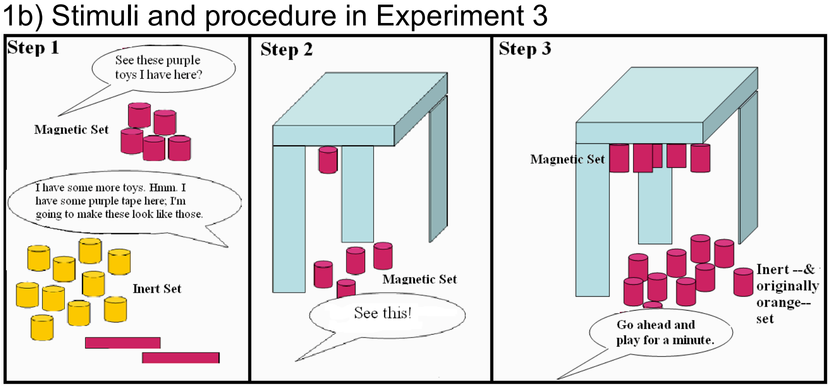
One Kind Condition (Blicket Only)



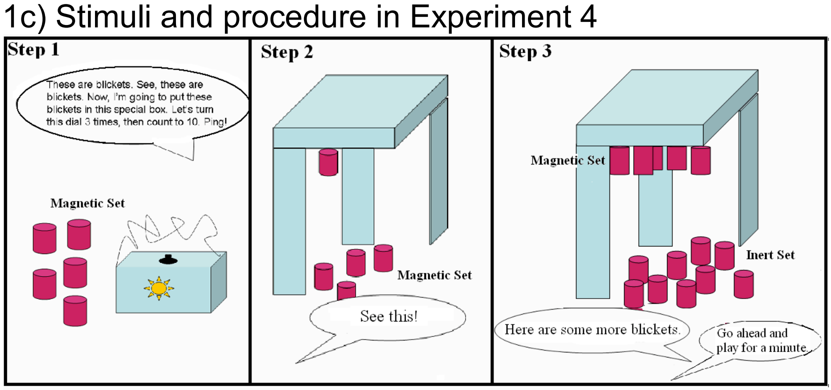
Two Kind Condition (Blicket/Dax)

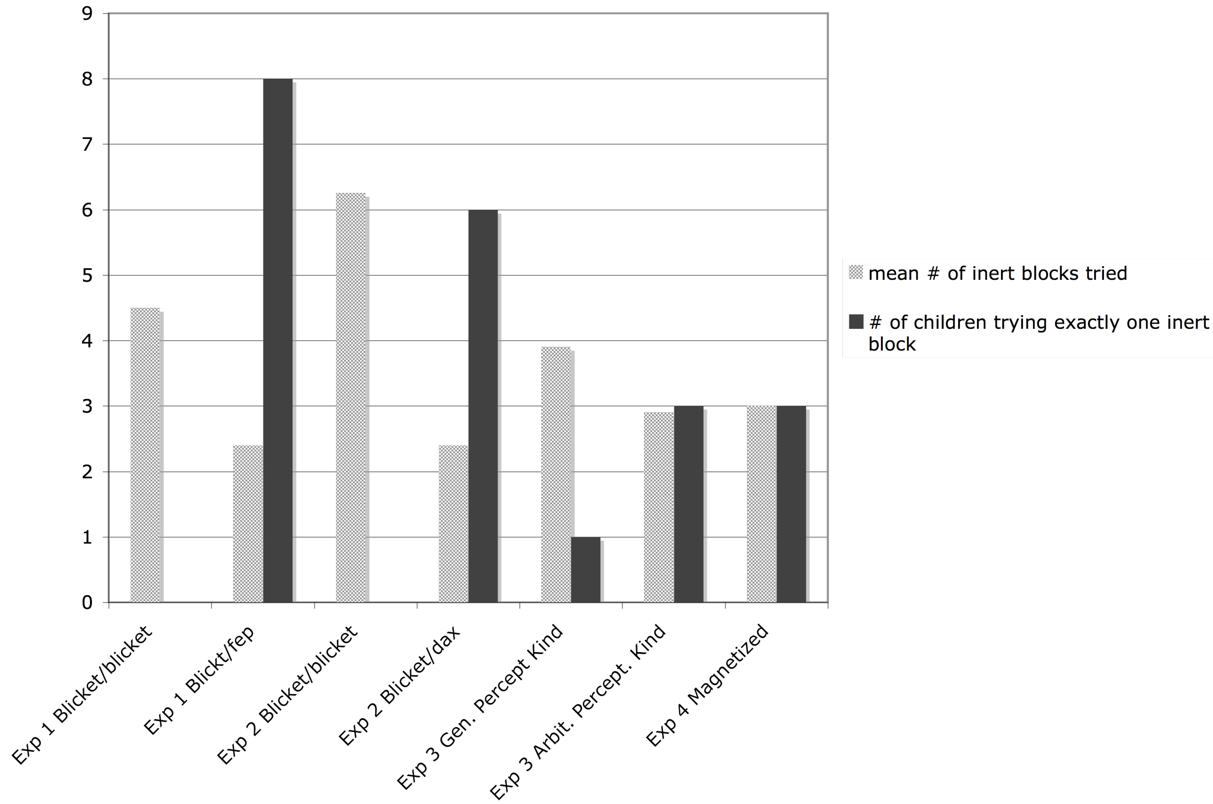


1b) Stimuli and procedure in Experiment 3



1c) Stimuli and procedure in Experiment 4




9
8
7
6
5
4
3
2
1
0
Exp 1 Blicket/blicket
Exp 1 Blickt/fep
Exp 2 Blicket/blicket
Exp 2 Blicket/dax
Exp 3 Gen. Percept Kind
Exp 3 Arbit. Percept. Kind
Exp 4 Magnetized
mean # of inert blocks tried
# of children trying exactly one inert block