Causal Categorization in Children and Adults

Brett Hayes

School of Psychology
University of New South Wales
Sydney, Australia

Bob Rehder

Department of Psychology
New York University
New York, NY

Keywords: Categorization; Causal reasoning; Cognitive development

Send all correspondence to:
Brett K. Hayes
School of Psychology,
University of New South Wales,
Sydney, Australia, 2052
Email: B.Hayes@unsw.edu.au

Abstract

Two experiments examined the impact of causal relations between features on categorization by adults and 5-6-year-old children. Participants learned about artificial categories containing instances with two causally related features and two non-causal features. They then selected the most likely category member from a series of novel test pairs. Classification patterns and logistic regression were used to diagnose the presence of independent effects of *causal coherence*, *causal status* and *relational centrality*. Adult classification was driven primarily by coherence when causal links were deterministic (Experiment 1), but showed additional influences of causal status and centrality when links were probabilistic (Experiment 2). Children's classification was based primarily on causal coherence in both cases. These results suggest that the generative model [Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1141-1159] provides a good account of causal categorization in both children and adults.

It is well established that causal knowledge plays an important role in adult categorization and property induction (see Sloman, 2005 for a review). Adults are more likely to assign an object to a category if it has the same causal features (features which causally contribute to the presence of other features) as known category members (Ahn, Kim, Lassaline, & Dennis, 2000b; Rehder & Hastie, 2001; Rehder, 2003a; Sloman, Love, & Ahn, 1998; see Ahn & Kim, 2001 and Rehder, 2010 for reviews). They often predict the features of category members on the basis of causal relations (Rehder & Burnett, 2005; see also Sloman & Lagnado, 2005; Waldmann & Hagmeyer, 2005), and use those inferences to establish category membership (Chaigneau, Barsalou, & Sloman, 2004; Hampton, Estes, & Simmons, 2007; Rehder & Kim, 2009). Finally, they inductively generalize properties from a familiar instance to a novel target item if the base and target have the same causal features (Hayes & Thompson, 2007; Lassaline, 1996; Rehder & Hastie, 2004; Rehder, 2006; 2009).

The role of causal knowledge in children's categorization decisions, however, remains less clear. The prevailing view for many decades was that young children are more likely than older children or adults to categorize on the basis of the perceptual similarity between known category members and novel objects, paying relatively little attention to complex relations like causality (Piaget & Inhelder, 1973; Sloutsky & Fisher, 2004; Springer 2001). This view has been challenged by a variety of recent developmental findings (see Hayes, 2006 for a review). For example, it has been found that when perceptually dissimilar objects have similar causal powers (i.e., give rise to similar effects) preschool children are likely to assign them to the same category (Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004). Opfer and Bulloch (2007) have shown that when salient causal relations conflict with perceptual similarity, 6-year-olds often prefer to categorize on the basis of causal relations (also see Carmichael & Hayes, 2001; Hayes, Foster, & Gadd, 2003). Finally, Ahn, Gelman, Amsterlaw, Hohenstein and Kalish (2000a) taught 7- to 9-year-olds that one category feature was the cause of two others. They told children, for example, that a fictitious animal called taliboos had promicin in their nerves, thick bones, and large eyes, and that the thick bones and large eyes were caused by the promicin (this network of causal relations is depicted in Figure 1A). Ahn et al. presented children with two animals, one missing only the cause feature (e.g., promicin) and the other missing only one of the effect features (e.g., thick bones or large eyes) and asked which was the more likely category member. Children preferred the alternative with a missing effect over the one with the missing cause (i.e., they preferred the item on the right hand side over the item on left in Figure 1B). Meunier and Cordier (2009) found a similar effect with 5-year-olds using a related design.

This evidence shows that by five years of age (and perhaps earlier) children can and do use

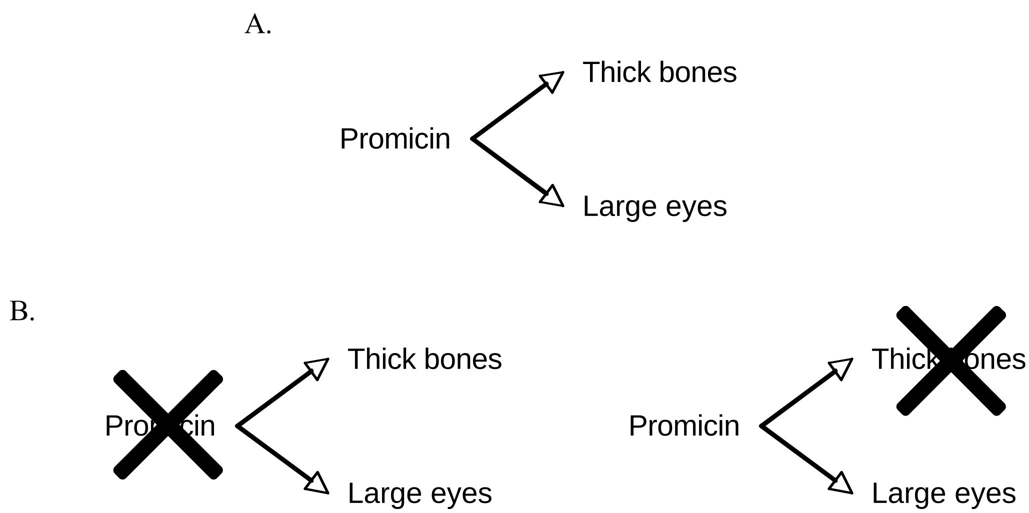

Figure 1. (A) Schematic represetation of one of the causal networks of category features taught to children in Ahn et al. (2000a). (B) One of the test trials in which children chose which of two items was a better category member. Crossed-out features were described to subjects as missing.

causal relations as a basis for categorizing. Two critical issues, however, remain unresolved. First, it is still unclear exactly how knowing about causal relations between object features affects the way that children use those features to categorize (we present several possibilities below). Second, the nature and extent of developmental change in causal categorization remains unclear. The current study addresses each of these issues.

Three Potential Effects of Causal Knowledge on Categorization

Previous work with adults suggests at least three specific ways that causal knowledge might affect categorization. First, Ahn and her colleagues have documented a *causal status effect* (Ahn, et al., 2000b; Sloman, Love & Ahn, 1998; also see Rehder, 2003a; Rehder & Kim, 2006). In this effect, causal features at lower levels of a causal chain are given more weight when deciding category membership than effect features that are dependent on the cause. For example, Ahn et al. (2000b) created novel categories consisting of instances with three typical features that were related in a causal chain (e.g., Roobans "eat fruit" which causes them to "have sticky feet" which in turn causes them to be good at "building nests"). Adults were then shown novel items with two typical and one atypical feature and asked to rate how likely it was that the item was a member of the target category. Likelihood ratings confirmed an effect of a feature's causal status; an item missing only the category's fundamental cause (e.g., eating fruit) was assigned a lower likelihood of belonging to the category as compared to an item missing only the terminal effect (building nests).

The causal status effect provides one interpretation of children's categorization decisions in Ahn et al. (2000a) and Meunier and Cordier (2009). Recall that children judged that an animal that was missing an effect was more likely to belong to the target category than one missing only the cause (they preferred the animal on the right in Figure 1B). This may have reflected their belief that the cause feature provided more evidence for category membership than the effect.

Second, Rehder and colleagues (Rehder & Hastie, 2001; Rehder, 2003a, b, Rehder & Kim, 2006) have documented a *coherence effect* in adult categorization. People are more likely to assign an item to a target category if the item manifests the pattern of correlations between cause and effect features that one expects in light of the category's causal laws. Specifically, in these studies items were more likely to be accepted as category members when a cause feature and an effect feature were either both present *or* both absent, compared with items in which a cause appeared without the expected effect or vice versa. This effect obtained above and beyond any tendency of the features considered individually to provide evidence of category membership.

Importantly, coherence provides an alternative interpretation of children's categorization patterns in Ahn et al. (2000a) and Meunier and Cordier (2009). The item missing only an effect feature in the right hand side of Figure 1B may have been chosen as a better category member because it violated only one expected correlation (between promicin and thick bones) whereas the left hand item violated two (between promicin and thick bones, and promicin and large eyes). Hence children's preferences in these studies could have reflected their sensitivity to coherence rather than causal status.

Finally, Rehder and Hastie (2001) proposed another possible effect of causal knowledge, namely, that a feature's importance to categorization increases as a function of the number of causal relations it is involved in, regardless of whether it plays the role of cause or effect. This *relational centrality effect* provides yet another interpretation of children's categorization patterns in Ahn et al. (2000a) and Meunier and Cordier (2009). Because the missing feature in the left-hand side of Figure 1B was involved in more causal links (two) than the missing feature in the right-hand side (one) then relational centrality suggests that the right-hand items is more likely to be a category member. Note that adult studies have not consistently found an effect of relational centrality. Whereas Kim and Ahn (2002a; 2002b) found that features involved in one causal link were rated as more important than those involved in no links, Rehder and Kim (2006) found that a feature involved in three relations was no more important than a feature involved in just one. Nonetheless, subjects in the Ahn et al. and Meunier and Cordier studies may have been more likely to exhibit an effect of relational centrality because of the use of a more sensitive measure (forced-choice) as compared to the adult studies (ratings scales).

In summary, although current empirical evidence clearly establishes that causal knowledge affects children's categorization, it is unclear how it does so. The empirical findings reviewed thus far may reflect either a causal status effect, a coherence effect, or a relational centrality effect. Indeed, they may reflect some combination of all three effects. The question of how causal knowledge affects children's categorization is important, because the answer has implications for whether theoretical accounts of causal-based categorization proposed to account for adult performance can be generalized to children. For example, Sloman et al.'s (1998) *dependency model* predicts a causal status effect but not effects of coherence. Thus, evidence for (against) the latter effect would constitute evidence against (for) the dependency model. In contrast, unlike the dependency model, Rehder's (2003a; b; Rehder & Kim, 2006) *generative model* predicts coherence effects, and thus evidence for that effect would support for that model. We will describe both of these accounts in more detail later. But it should be clear that our current ignorance regarding the specifics of children's causal categorization makes it impossible to tell whether either of these theoretical accounts could apply to children.

Evaluating Effects of Causal Knowledge on Categorization

Accordingly, the aim of the current studies was to test for the presence of the causal status, coherence and relational centrality effects in 5-6 year old children and adults, and thereby acquire data that will further constrain theoretical accounts of the development of causal categorization. To this end, our experiments used both a design and a method of statistical analysis that were novel in that they allowed us to test for the presence of *all three causal effects* independently, as now described.

Children and adults were instructed on artificial categories with four typical features. To convey which features were typical, subjects were presented with two exemplars from an artificial category (see Table 1 for an example). One instance showed features that were described as typical of most category members. The other (the atypical instance) had opposite values on each feature dimension, found only in a minority of category members. In both types of instances, two of the four features were described as causally related. For example, subjects were told that most Rogos can stay underwater a long time *because* they have big lungs. The other two features (have long ears and sleep during the day) were causally unrelated to any feature. In Table 1 and throughout this article, typical feature values are indicated with a '1' and atypical values are indicated with a '0'. In addition, the labels C, E, N_1, and N_2 are used to denote the cause, effect, and two neutral feature dimensions, respectively.

Table 1

Example of typical and atypical study instances from the Rogos category

Typical Instance (1111) "Most ROGOS have…"				Atypical Instance (0000) "Some ROGOS have..."			
Dimension C	Dimension E	Dimension N_1	Dimension N_2	Dimension C	Dimension E	Dimension N_1	Dimension N_2
Have big lungs	Stay underwater for a long time	Have long ears	Sleep during the day	Have small lungs	Stay underwater for a short time	Have short ears	Sleep at night

After learning a category's typical and atypical features and the causal linkages, participants were presented with the seven test items described in the left-hand side of Table 2. Each test item consisted of two objects, labeled X and Y in Table 2, and subjects were required to choose which was more likely to be a member of the target category. For each object in a pair, Table 2 indicates the value on each of four feature dimensions. An 'x' indicates that no information about the dimension was provided for that item. For example, '10xx' denotes an instance that was described as having the typical cause feature but missing the typical effect feature (i.e., it had the atypical feature on that dimension).

The seven test items in Table 2 were chosen because they allow for the identification of multiple effects of causal knowledge on classification. First, note that only two of the test items compare instances with different numbers of typical features. Items T_A and T_B compared instances with two (X) versus zero (Y) typical features and were included as a manipulation check to determine if subjects attended to the information about which features were typical of the category (they should choose alternative X). In contrast, the remaining five test items (T_C–T_G) compare instances with the same number of typical features. Of course, if category choices are determined only by the number of typical features, they will be at chance on these five items. But by changing the importance of features, and feature combinations, the interfeature causal relation that subjects were taught provides a basis for distinguishing between X and Y in items T_C–T_G. In particular, responses to those items can be used to determine whether subjects distinguish the items in those pairs on the basis of causal status, coherence, relational centrality, or some combination of all three.

Table 2

Test items, predictions for the causal status, coherence, and relational centrality effects, and empirical results from Experiments 1 and 2.

				choice$_k$ (X, Y) [Preferred Alternative]							
				Typicality Only	Typicality + Causal Status Effect Only	Typicality + Coherence Effect Only	Typicality + Relational Centrality Effect Only	Experiment 1 (Preference for X)		Experiment 2 (Preference for X)	
				$w_C = 1$ $w_E = 1$ $w_N = 1$ $w_H = 0$	$w_C = 2$ $w_E = 1$ $w_N = 1$ $w_H = 0$	$w_C = 1$ $w_E = 1$ $w_N = 1$ $w_H = 1$	$w_C = 2$ $w_E = 2$ $w_N = 1$ $w_H = 0$				
Test Item	Choice X	Choice Y	*diff*$_k$ (X, Y)					Adults	5-6 Year Olds	Adults	5-6 Year Olds
T_A	11xx	00xx	$2(w_C + w_E)$	.98 [X]	1.0 [X]	.98 [X]	1.0 [X]	.95**	.80**	.96**	.80**
T_B	xx11	xx00	$4w_N$	.98 [X]	.98 [X]	.98 [X]	.98 [X]	.99**	.75**	.94**	.71**
T_C	10xx	01xx	$2(w_C - w_E)$	.50 [=]	.88 [X]	.50 [=]	.50 [=]	.52	.51	.71**	.48
T_D	10xx	xx10	$w_C - w_E - w_H$	.50 [=]	.73 [X]	.27 [Y]	.50 [=]	.30**	.48	.34**	.28**
T_E	01xx	xx01	$-w_C + w_E - w_H$	.50 [=]	.27 [Y]	.27 [Y]	.50 [=]	.32**	.43	.37**	.42
T_F	11xx	xx11	$w_C + w_E - 2w_N + w_H$	.50 [=]	.73 [X]	.73 [X]	.88 [X]	.70**	.68**	.78**	.57
T_G	00xx	xx00	$-w_C - w_E + 2w_N + w_H$	.50 [=]	.27 [Y]	.73 [X]	.12 [Y]	.62*	.55	.57	.54

Note. Predicted values for *choice*$_k$ (X, Y) are generated from the values of parameters w_C, w_E, w_N, and w_H. Whether object X or Y is preferred is shown in brackets ("=" means no preference). Empirical choice probabilities are tested against .5. * $p < .05$. ** $p < .01$

Because these effects have not previously been assessed in the context of forced-choice judgments, to present our predictions rigorously (and provide quantitative examples), we introduce the following definitions. Define w_C, w_E, and w_N as the evidentiary weights provided by the cause feature, the effect feature, and the neutral features, respectively. We assume that an object's degree of category membership is increased by w_i if a feature on dimension type i is present and decreased by w_i if it is absent. In addition, define w_H as the weight associated with whether the object exhibits coherence: An object's degree of category membership is increased by w_H if the cause and effect features are both present or both absent and decreased by w_H if one is present and the other absent. In other words, the degree of evidence that an object i belongs to the category k is defined as,

$$ev_k(i) = w_C f_{i,C} + w_E f_{i,E} + w_N f_{i,N_1} + w_N f_{i,N_2} + w_H h_i \quad (1)$$

where $f_{i,j}$ is an indicator variable reflecting whether the typical feature on dimension j is present (+1), whether the atypical feature is present (–1), or whether no information about dimension j is provided (0). In addition, h_i is defined such that $h_i = f_{i,C} f_{i,E}$ and as a consequent indicates whether i is coherent (+1), incoherent (–1), or neither (0). Table 3 presents how ev_k simplifies for each of the eight instances that appear in the seven test items in Table 2. For example, ev_k (10xx) = $w_C - w_E - w_H$, because instance 10xx has the cause feature (w_C), is missing the effect feature ($-w_E$), and is incoherent because the cause is present but effect absent ($-w_H$). Table 3 also presents how ev_k evaluates for a number of theoretically-important parameter sets, as discussed below.

Given the definition for an individual object's degree of category membership in Equation 1, we can specify which of two instances should be preferred when they are paired together in a forced-choice judgment. Specifically, the probability of choosing X over Y is a function of the difference between ev_k (X) and ev_k (Y),

$$\begin{aligned} diff_k(X,Y) &= ev_k(X) - ev_k(Y) \\ &= \left(w_C f_{X,C} + w_E f_{X,E} + w_N f_{X,N_1} + w_N f_{X,N_2} + w_H h_X\right) - \\ &\quad \left(w_C f_{Y,C} + w_E f_{Y,E} + w_N f_{Y,N_1} + w_N f_{Y,N_2} + w_H h_Y\right) \\ &= w_C\left(f_{X,C} - f_{Y,C}\right) + w_E\left(f_{X,E} - f_{Y,E}\right) + w_N\left(f_{X,N_1} - f_{Y,N_1}\right) + \end{aligned}$$

$$w_N\left(f_{X,N_2} - f_{Y,N_2}\right) + w_H\left(h_X - h_Y\right)$$

$$= w_C m_{XY,C} + w_E m_{XY,E} + w_N m_{XY,N_1} + w_N m_{XY,N_2} + w_N m_{XY,h} \quad (2)$$

where $m_{XY,j}$ are *match variables* indicating whether X and Y match on dimension *j*. For example, $m_{XY,C} = 2$ if the cause feature is present in X but absent in Y, –2 if it is present in Y but absent in X, and 0 if X and Y both have the cause feature or both don't have it. Table 2 presents how $diff_k$ (X, Y) simplifies for each of the seven test items depending on whether the two objects in those items have matching or mismatching values on each dimension.

Table 3
Simplified expressions of ev_k for each test object and values of ev_k for parameter values that illustrate the causal status, coherence, and relational centrality effects.

Object *i*	ev_k (*i*)	Typicality Effect Only	Typicality + Causal Status Effect Only	Typicality + Coherence Effect Only	Typicality + Relational Centrality Effect Only
		$w_C = 1$ $w_E = 1$ $w_N = 1$ $w_H = 0$	$w_C = 2$ $w_E = 1$ $w_N = 1$ $w_H = 0$	$w_C = 1$ $w_E = 1$ $w_N = 1$ $w_H = 1$	$w_C = 2$ $w_E = 2$ $w_N = 1$ $w_H = 0$
11xx	$w_C + w_E + w_H$	2	3	3	4
10xx	$w_C - w_E - w_H$	0	1	–1	0
01xx	$-w_C + w_E - w_H$	0	–1	–1	0
00xx	$-w_C - w_E + w_H$	–2	–3	–1	–4
xx11	$2w_N$	2	2	2	2
xx10	0	0	0	0	0
xx01	0	0	0	0	0
xx00	$-2w_N$	–2	–2	–2	–2

The probability of choosing X over Y can now be obtained by passing $diff_k$ (X, Y) through a logistic,

$$choice_k(X,Y) = \frac{1}{\left(1 + e^{-diff_k(X,Y)}\right)} \quad (3)$$

Equation 3 predicts a choice probability in favor of X of close to 1 when $diff_k$ (X, Y) >> 0, close to 0 when $diff_k$ (X, Y) << 0, and close to .5 when $diff_k$ (X, Y) ≅ 0.

Finally, note that the three potential effects of causal knowledge on classification can be defined in terms of the parameters w_C, w_E, w_N, and w_H. A causal status effect obtains when $w_C > w_E$, that is, when the cause feature has a greater effect on categorization judgments than the effect feature. A coherence effect obtains when $w_H > 0$, that is, when objects are better category members if cause and effect are both present or both absent. A relational centrality effect obtains when $w_E > w_N$, that is, when the neutral feature has less of an effect on categorization judgments than the effect feature. The relational centrality effect is defined as the importance of *N* relative to *E* rather than *C* because, due to the possible presence of a causal status effect, *E* is likely to be the least heavily weighed feature that is involved in a causal relationship.

With these definitions established, we can now specify how subjects will respond to the seven test items in Table 2. We consider the cases where there is an effect of feature typicality only and then, in turn, an additional effect of causal status, coherence, or relational centrality.

Typicality effect only. First consider how subjects will respond to the test items if only a typicality effect obtains; that is, when each feature provides evidence for category membership ($w_i > 0$ each feature weight *i*) but there is no causal status effect ($w_C = w_E$), no coherence effect ($w_H = 0$), and no relational centrality effect ($w_E = w_N$). As an example, Table 3 presents a set of weights that exemplifies these constraints, namely, $w_C = 1$, $w_E = 1$, $w_N = 1$, and $w_H = 0$. For these weights, it also presents values of ev_k corresponding to each test object, from which values of $diff_k$ and $choice_k$ can be computed from Equations 2 and 3 for each test item. For example, for test item T_A, ev_k (11xx) = 2 and ev_k (00xx) = –2, which results in $diff_k = 4$. Applying Equation 3, this value of $diff_k$ yields a choice probability $choice_k = .98$; that is, X should be strongly preferred over Y. This reflects the (obvious) fact that objects with more typical features are better category members. In addition, there should be no preference between objects with the same number of features. For example, for test item T_C, ev_k (10xx) = ev_k (01xx) = 0, $diff_k = 0$ and thus $choice_k = .50$. Table 2 presents the predicted values of $choice_k$ for each test item for these parameter values, and indicates in brackets whether X or Y should be preferred or whether there should be no preference ("="). The overall response profile {X, X, =, =, =, =, =} for the seven test items reflects our claim, made earlier, that a lack of any effect of causal knowledge should manifest itself in chance performance on all items except T_A and T_B. Note that these predictions and those presented below hold not just for the example parameter values shown in Table 3 but for any parameters that satisfy the constraints.

Typicality + causal status effect only. Now consider how subjects will respond to the seven test items if a causal status effect also obtains, that is, if $w_C > w_E$. These predictions are generated still assuming the absence of both a coherence effect ($w_H = 0$) and a relational centrality effect ($w_E = w_N$). Of course, subjects should still choose alternative X on test items T_A and T_B because it has the greater number of features. More interesting is test item T_C which involves a choice between 10xx (cause present but effect absent) and 01xx (cause absent but effect present). Although these instances have the same number of typical features, 10xx has the heavily weighed feature (the cause) but is missing the less heavily weighed feature (the effect). In contrast, 01xx has the less heavily weighed effect but is missing the heavily weighed cause. Thus, if subjects exhibit a causal status effect, they should choose alternative X (10xx) on test item T_C.

Again, to provide a quantitative example, Table 3 presents the values of ev_k for a set of w weights that exemplifies the presence of typicality and causal status effects but the absence of coherence and relational centrality, namely, $w_C = 2$, $w_E = 1$, $w_N = 1$, and $w_H = 0$. For test item T_C, these parameters yield ev_k (10xx) = 1 and ev_k (01xx) = –1 and thus $diff_k = 2$ and $choice_k = .88$; that is, X should be chosen over Y.

The predictions for the remaining test items can be generated in a similar manner. In test items T_D and T_F, the X alternatives (10xx and 11xx) should be preferred over the Ys (xx10 and xx11) because the Xs have the heavily weighed cause. This results in $diff_k = 1$ and $choice_k = .73$ for those test items. Conversely, for test items T_E and T_G the Y alternatives (xx01 and xx00) should be preferred over the Xs (01xx and 00xx) because the Xs are missing the cause feature, resulting in $diff_k = –1$ and $choice_k = .27$. In summary, the overall pattern of responses for test items T_A–T_G associated with the causal status effect is {X, X, X, X, Y, X, Y}.

Typicality + coherence effect only. The predictions for the coherence effect ($w_H > 0$) are generated assuming an effect of typicality ($w_i > 0$) and the absence of both a causal status effect ($w_C = w_E$) and a relational centrality effect ($w_E = w_N$). For test item T_C, both X (10xx) and Y (01xx) are equally incoherent; thus performance should be at chance. A choice probability of .50 is indicated by the example parameter values $w_C = 1$, $w_E = 1$, $w_N = 1$, and $w_H = 1$ in Table 3, which results in $ev_k(10xx) = ev_k(01xx) = -1$ and $diff_k = 0$. In test item T_D, Y (xx10) should be chosen over X (10xx) because of the incoherence exhibited by 10xx. This choice is indicated by the values $ev_k(10xx) = -1$, $ev_k(xx10) = 0$, and $diff_k = -1$. Y (xx01) should be chosen over X (01xx) in test item T_E for the same reason. In test items T_F and T_G, the Xs (11xx and 00xx) should be chosen over the Ys (xx11 and xx00) because of the coherence exhibited by the former. In summary, the pattern of responses for items T_A–T_G associated with a coherence effect is {X, X, =, Y, Y, X, X}.

Typicality + relational centrality effect only. Finally, the predictions for relational centrality effect ($w_E > w_N$) are generated assuming an effect of typicality ($w_i > 0$) and the absence of a causal status effect ($w_C = w_E$) and a coherence effect ($w_H = 0$). For T_D, both X (10xx) and Y (01xx) have an equal number of relationally central features (one), thus performance should be at chance for this item. A choice probability of .50 is indicated by the example parameter values $w_C = 2$, $w_E = 2$, $w_N = 1$, and $w_H = 0$ in Table 2, which results in ev_k (10xx) = ev_k (01xx) = –1 and $diff_k = 0$. In test items T_D and T_E, the fact that the Xs (10xx and 01xx) have one central feature is compensated by the fact they are missing the other; thus subjects should be at chance on both test items. In test item T_F, X (11xx) should be chosen over the Y (xx11) because X possesses two central features. In contrast, in T_G Y (xx00) should be chosen over X (00xx) because X is missing two central features. In summary, the response pattern associated with a relational centrality effect is {X, X, =, =, =, X, Y}.

The preceding analyses indicate how the causal status, coherence, and relational centrality effects each imply a unique pattern of performance on the seven test items in Table 2. Importantly, they also allow for the detection of two or more effects operating simultaneously. They do so because Equations 2 and 3 suggest that a logistic regression model applied to all seven choice problems in Table 2 can yield estimates of parameters w_C, w_E, w_N, and w_H, estimates which in turn will allow us to assess the simultaneous presence of an effect of causal status ($w_C > w_E$), coherence ($w_H > 0$), and relational centrality ($w_E > w_N$). This insight is important because nothing prevents subjects from exhibiting two or more of the effects we have identified, a finding that would lead to a pattern of responding that is more complex than those shown in Table 2. Because it simultaneously evaluates multiple test items, logistic regression also provides a more powerful statistical test. Accordingly, in two experiments, subjects were presented with the test items in Table 2 and logistic regression analyses were performed to test for the presence of these multiple effects on categorization.

Experiment 1

In Experiment 1, university students and 5-6 year old children were taught artificial categories like those in Table 1, and then presented with the test items in Table 2. Participants were taught an explicit causal link between features on two dimensions. Note that, in Experiment 1, no information about the strength of this causal link was provided. Subjects were told, for example, that "because they have big lungs Rogos can stay underwater for a long time", without being told whether this was true for every Rogo that has big lungs (a deterministic causal link) or only some of them (a probabilistic causal link). After reporting the results from our first experiment, we argue that subjects in Experiment 1 were likely to have interpreted the causal link as deterministic. Hence, in Experiment 2 we extend our results by explicitly stating that the causal link was probabilistic.

Method

Participants. The participants were 41 kindergarten and first grade children (M_{AGE} = 5 years, 10 months; Range: 5 years, 5 months – 6 years, 6 months) from a number of private elementary schools in middle-class metropolitan areas of Sydney, Australia. Thirty eight undergraduate Psychology students participated for course credit.

Materials. Four sets of target animal categories labeled Rogos, Waddos, Daxes, and Bliks were created. Waddos, Daxes, and Bliks had the same statistical and causal structures as the Rogos in Table 1 but with different surface features (see Appendix for a list of the features used in each category). Two study instances from each training category were presented, one typical instance and one atypical instance. As shown in Table 1, each instance had four features, two that were causally related and two neutral features. The typical instance (with notation 1111) showed the causal and neutral features that were true of most category members. The atypical instance (with notation 0000) showed the features that were true of some category members. The assignment of features to typical and atypical instances was counterbalanced across participants (i.e., half the participants saw the features as presented in Table 1 while for the remainder the assignment of features to the typical and atypical roles was reversed). The seven test items for each category followed the feature structure shown in the left columns of Table 2. These items always involved a comparison of two instances, each with two features. All individual features were illustrated with 13-cm x 9-cm black-and-white line drawings. In addition, sets of 13-cm x 9-cm, black silhouettes were developed for each of the four categories. These were used to illustrate the body shape of category members and to demonstrate differences in the base rates of typical and atypical instances within a category (see Figure 2 for an example).

Procedure. All participants were tested individually in a quiet room and told that they were going to learn about different kinds of extraterrestrial animals. Adults received four study-test trials focusing on each of the four target categories. Children received two study-test trials for only two of the target categories, with the presentation of the four categories counterbalanced across this age group. The decision to present children with only two categories was made after pilot testing showed that children became fatigued when more than two training/test sequences were presented. Trial order was randomized across participants.

In each study phase trial, participants were told that they were going to learn about a particular kind of extraterrestrial animal (e.g., Rogos). The four features comprising each study instance were then attached using Velcro tabs to a poster board positioned 1m from the participant, with typical and atypical instances placed at the left and right extremes of the board. The features of each instance were described twice by an experimenter. On the first occasion, the experimenter simply labeled the features of each instance (as shown in Table 1). On the second occasion, the experimenter highlighted the relationship between the cause and effect feature by saying, "It's because this animal has [causal feature] that it can [effect feature]. Separate statements about the relevant causal relations were made for typical and atypical instances. For children, this causal relation was enhanced by the placement of a solid arrow on the poster board indicating the direction of causality. The labels of the neutral features were also repeated (e.g., "And don't forget they have long ears and sleep during the day").

Participants were told that "most" members of the category had the features like the typical instance but that "some" had features like the atypical instance. The different base rates of typical and atypical instances for a given category was reinforced by placing five identical silhouettes of category members above the typical instance and two identical silhouettes above the atypical instance (see Figure 2). After presentation of the study items participants were shown pairs of features for each of the four dimensions and asked to identify the feature associated with typical instances (e.g., Can you show me which things *most* Rogos have?). If this question was answered incorrectly all the features and causal links between features for each instance were repeated. Typical instances were always presented before atypical instances. The left-right position of the typical and atypical instances was counterbalanced across trials so each instance type appeared in each position on an equal number of occasions. The order of presentation of causal and neutral pairs was always the same for the typical and atypical instances on a given trial. However, the order of presentation of causal and neutral feature pairs was counterbalanced across trials so that each type of feature pair was presented first on an equal number of occasions. The cause feature for a given instance always preceded the effect feature. The ordering of the two neutral features within each instance was randomized across trials.

The test phase for each trial commenced immediately after study. Study phase instances remained in view throughout testing. Participants were told that they would see pairs of animals and that their task was to decide which was most likely to be a category member (e.g., "Which animal do you think is more likely to be a Rogo?"). Participants were told that guessing was acceptable. The features of the test pairs were presented on a table to the left and right of the participant. Features were described by the experimenter and illustrated using the same sorts of pictures used during study. Participants could indicate their choice of category member verbally or by pointing. The two test items designed to check that participants could correctly apply the base rates of typical and atypical features to determine category membership (items T_A and T_B from Table 2), were always presented first with the ordering of these items randomized across trials. The remaining five test items were then presented in random order.

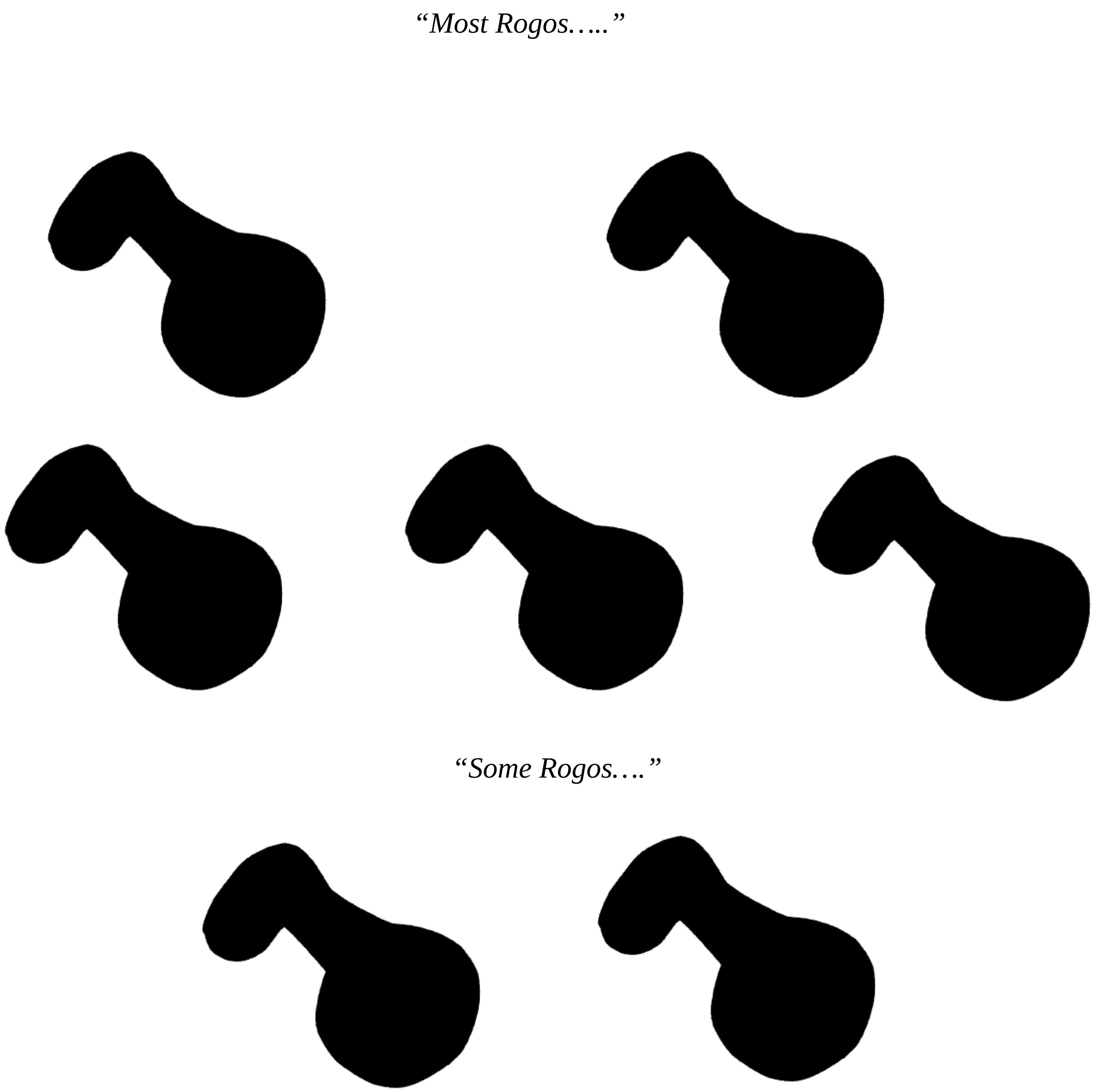

Figure 2. Example of silhouettes used to illustrate category members.

Results

Recall that the purpose of test items T_A and T_B was to confirm that subjects were sensitive to the feature typicality information we provided. Accordingly, subjects whose average response to these two items did not exceed .50 (i.e., did not favor alternative the more typical instance X), were excluded from further analyses. This criterion resulted in the exclusion of data from one child and zero adults. For the remaining subjects, Table 2 presents averaged choice proportions for the seven test items for adults and children.

Recall that if subjects' responses are based on typicality alone, they should be at chance on all test items where the two alternatives had the same number of typical features, namely, T_C–T_G. Neither group of subjects was at chance for all five pairs. This shows that categorization by both adults and children was affected by the presence of interfeature causal links. To understand which effects were driving causal categorization, we now consider performance of each of age group separately.

Adult responses. Examination of the overall pattern of adult responses in Table 2 indicates the presence of a coherence effect and the absence of causal status and relational centrality effects. First, subjects were at chance on item T_C and significantly more likely to choose alternatives Y, Y, X, and X for test items T_D–T_G, respectively, a pattern of responding consistent with the coherence effect (Table 2). Second, contra the predictions of the causal status effect, subjects were not more likely to choose alternative X on test items T_C and T_D and not more likely to choose Y on test item T_G. Third, contra the predictions of the relational centrality effect, subjects were not more likely to choose alternative Y on item T_G and differed from chance in their assignments of T_D and T_E.

These informal arguments are augmented by a quantitative analysis using logistic regression. The logistic regression model represented by Equations 2 and 3 were applied to the choice data of each adult subject.[1] The estimated values of w_C, w_E, w_N, and w_H averaged over subjects are presented in Table 4. Parameter w_H provides information about the presence of a coherence effect. In addition, for each subject we computed measures corresponding to the effect of typicality (*average*(w_C, w_E, w_N, w_N)), the causal status effect ($w_C - w_E$), and the relational centrality effect ($w_E - w_N$). These measures averaged over subjects are included in Table 4.

Table 4 confirms the presence of a significant coherence effect in adult categorization choices, indicated by a value for parameter w_H of .85 that was significantly greater than 0, $t(37) = 6.10$, $p < .0001$. Instances were judged more likely to be category members when they were coherent (cause and effect both present or both absent) and less likely to be category members when they were incoherent (cause present and effect absent or vice versa). The absence of a significant difference between a weight on the cause feature ($w_C = .66$) and the effect feature ($w_E = .70$), $t < 1$, reflects the absence of a causal status effect. Finally, the absence of a significant difference between w_N (.65) and w_E,(.70), $t < 1$, confirms the absence of a relational centrality effect.

These parameter values elucidate the pattern of responding on individual test items. For example, test item T_D pits the causal status and coherence effects against one another: the former predicts choice X whereas the latter predicts Y. In fact, the logistic regression model predicts a choice probability of .33 favoring Y (cf. the observed value of .30) reflecting a coherence effect. This prediction arose because the value of parameter of w_H (that drives the model to prefer Y) is large whereas a difference between w_C and w_E (that drives the model to prefer X) is absent.

Table 4
Average parameter estimates for adults and children. Standard errors are presented in parentheses.

	Experiment 1		Experiment 2	
	Adults	5-6 Year Olds	Adults	5-6 Year Olds
Parameter				
w_C	.66 (.09)	0.33 (.05)	.93 (.10)	.20 (.08)
w_E	.70 (.08)	0.29 (.06)	.63 (.08)	.29 (.04)
w_N	.65 (.07)	0.20 (.04)	.51 (.04)	.21 (.04)
w_H	.85 (.14)	0.22 (.08)	.74 (.12)	.31 (.09)
Effect				
Typicality [*average* (w_C, w_E, w_N, w_H)]				
Causal status [$w_C - w_E$]	–.03 (.10)	.04 (.07)	.30*** (.08)	–.09 (.05)
Coherence [w_H]	.85*** (.14)	.22** (.08)	.74*** (.12)	.31*** (.09)
Relational centrality [$w_E - w_N$]	.05 (.10)	.08 (.07)	.12 (.09)	.08 (.05)

Note. Standard errors are presented in parentheses. Effects are tested against 0. * $p < .05$. ** $p < .01$. *** $p < .001$.

Children's responses. In comparison to the adult data, informal analysis of children's responses to the individual test items was less informative regarding the presence of the three effects. On one hand, children were like the adults in that they showed no preference for choice X on item T_C (evidence against a causal status effect; see Table 2). On the other hand, while their preferences on test items T_D, T_E, T_F, and T_G were in a direction consistent with a coherence effect (and in the same direction as the adults' preferences), these choice probabilities reached significance for test item T_F only. Once again, to untangle the possible presence of multiple effects (and provide more powerful statistical tests), we submitted children's responses to the same logistic regressions applied to the adults, the results of which are presented in Table 4.

First, Table 4 shows the presence of a coherence effect, indicated by a value of parameter w_H (.22) that was significantly greater than zero, $t(39) = 2.71$, $p < .05$. This parameter reflects children's overall pattern of choices. Small (and mostly nonsignificant) preferences for Y on test items T_D and T_E, and for X on T_F and T_G, combine to produce a significant coherence effect. Second, the absence of a difference between the weight on the cause feature ($w_C = .33$) and the effect ($w_E = .29$), $t < 1$, reflects the absence of a causal status effect. These parameters reflect children's agnostic performance on test item T_C: The logistic regression model predicts a choice probability of .52 favoring X for this item (cf. the observed value of .51). Finally, the absence of a significant difference between w_N (.20) and and w_E (.29) indicates the absence of a relational centrality effect, $t(39) = 1.28$, $p > .20$. Note that the significant coherence effect explains children's preference for alternative X in test item T_F. The logistic regression model predicts a choice probability of .60, favoring X for this item (cf. the observed value of .67).

Comparison of adults and children. The parameter estimates produced by our logistic regression analyses also provide an opportunity to compare the magnitude of the effects in the two age groups. First, adults exhibited a larger typicality effect (average of w_C, w_E, w_N, and w_N) as compared to children (.66 vs. .27), $t(76) = 6.93$, $p < .0001$. Second, although both groups exhibited sensitivity to coherence, that effect was significantly larger in adults ($w_H = .85$) than children (.22), $t(76) = 3.95$, $p < .001$. Finally, neither the causal status effect nor the relational centrality effect differed between groups, both *t*s < 1.

Discussion

The purpose of Experiment 1 was to establish how causal knowledge affects categorization decisions in 5–6 year old children and adults. A central finding was that both adults and children exhibited coherence effects. Instances were more likely to be accepted as category members if they confirmed the category's causal link (cause and effect both present or both absent) and less likely to be accepted if they violated that link (cause present and effect absent or vice versa). The presence of a coherence effect replicates previous demonstrations of that effect in adults (Marsh & Ahn, 2006; Rehder

& Hastie, 2001, Rehder 2003a; b, Rehder & Kim 2006; in press) and extends those findings to a category with only one causal link and use of a forced-choice procedure. Importantly, Experiment 1 is the first to establish that 5–6 year old children also exhibit a coherence effect in causal-based categorization.

Another key finding was the lack of a causal status effect in both adults and children. The lack of a causal status effect may seem surprising given the many previous reports that both age groups give more weight to cause than effect features (e.g., Ahn, 1998; Ahn et al., 2000a; Rehder, 2003b, Rehder & Kim, 2006; Sloman et al. 1998). Recall though that certain kinds of empirical results seen as evidence for causal status may actually reflect relational centrality or coherence. When the effects of centrality and coherence are controlled, causal status effects often disappear or are relatively weak (e.g., Rehder, 2003b, Rehder & Kim, 2006). Moreover, more recent work has shown that the causal status effect occurs only under particular conditions, conditions that may not have obtained in Experiment 1 (Rehder & Kim, in press). Hence, in Experiment 2 we will induce conditions more favorable to the causal status effect in adults and then determine whether it also occurs in children.

A final result is that the relational centrality effect was absent in both adults and 5-6 year-olds. We defer further discussion of this finding until the results of Experiment 2 are reported.

Experiment 2

One key result from Experiment 1 was the lack of a causal status effect in children's categorization. On one hand, this result supports one of our alternative interpretations of the findings of Ahn et al. (2000a) and Meunier and Cordier (2009), namely that the causal status effect in children observed in each study may actually be due to coherence. However, the fact that adults also did not exhibit a causal status effect raises the possibility that its absence in Experiment 1 may have been due to other factors. Experiment 1 differs from previous studies in many ways (having only one causal link per category, different materials, etc.) and perhaps one of these other factors was responsible for the lack of a causal status effect. Clearly, before concluding that children do not exhibit a causal status effect, it is desirable to establish experimental conditions under which it occurs in adults.

Why didn't adults exhibit a causal status effect in Experiment 1? Recently, Rehder and Kim (in press) have demonstrated that the causal status effect is influenced by a number of variables. One important factor is the perceived strength of the causal relationship. A causal status effect tends to occur for probabilistic but not deterministic causal links. For example, Ahn et al. (2000b) found a large causal status effect when a causal relationship was described in probabilistic terms by use of the phrase "tends to" (e.g., "Sticky feet *tends to* allow roobans to build nests on trees"). In contrast, Rehder and Kim (2006) found only a weak causal status effect when they omitted any information about the strength of the causal link. A subsequent experiment that tested Rehder and Kim's materials by asking subjects to judge how often the cause produced the effect, found the modal response was 100%. This suggests that causal links in that study were interpreted as effectively deterministic. Finally, Rehder and Kim (in press) directly manipulated the strength of the causal links and confirmed the presence of a causal status effect for probabilistic links and its absence for deterministic ones.

A causal status effect may have failed to obtain in adults (and perhaps children) in Experiment 1 because subjects interpreted the causal link as deterministic. For example, when told that Rogos can stay underwater for a long time because of their lung size, participants may have interpreted this to mean that *every* Rogo manifests this feature relation. On this account, the effect feature is at least as prevalent as the cause feature. If both effect and cause are treated as equally diagnostic of category membership, this would eliminate an effect of causal status.

The aim of Experiment 2, therefore, was to establish conditions under which a causal status effect obtains in adults, in order to determine whether it then also occurs in children. To this end, the description of the causal link was changed to imply a probabilistic causal relation. For adults, this was accomplished by following Ahn et al. (2000b) and using the phrase "tends to" in the description of the causal link. For example, for Rogos the link was described as "because they have big lungs, Rogos tend to be able to stay underwater for a long time."

Because they may not understand the "tends to" phrase, "sometimes" was used to express probabilistic relations for 5-6 year-olds. The term "sometimes" has been used in previous work where probabilistic causal relations have been presented to preschoolers (e.g., Shultz & Somerville, 2006). Moreover, there is evidence that by around 6 years of age, a majority of children grasp the meaning of adverbs that indicate probabilistic rather deterministic relations (Hoffner, Cantor & Badzinski, 1990).

Hence, we expected that 5-6 year-olds would understand the probabilistic language used to describe the causal relations in this experiment.

In addition to providing a test of the causal status effect in children, Experiment 2 provided additional tests of the presence of the coherence and relational centrality effects in both age groups.

Method

Participants. The participants were 51 kindergarten and first grade children (M_{AGE} = 5 years, 9 months; Range: 5 years, 6 months – 6 years, 6 months) from a private elementary school in a middle-class metropolitan area of Sydney, Australia. Forty five undergraduate Psychology students participated for course credit.

Materials and Procedure. These were identical to Experiment 1 except for the way that cause and effect features were presented. For adults the relation between these features was always instantiated by saying "Because they have [C] Rogos *tend to* [E]". For children the same relations were instantiated by saying "Because they have [C] Rogos *sometimes* [E]". [2]

Results

Following Experiment 1, we excluded subjects (five children, zero adults) whose average responses to test items T_A and T_B did not exceed .50 (i.e., showed no sensitivity to feature typicality). Table 2 gives averaged choice proportions for the remaining participants. Both age groups showed above chance reasoning to one or more of test items T_C-T_G. This indicates that categorization in both age groups was affected by causal features links even when these were explicitly presented as probabilistic. We now consider the details of these causal effects for each age group.

Adult responses. Adult responses in Experiment 2 were qualitatively similar to those in Experiment 1, with one major exception: Whereas adults were at chance on test item T_C in the first experiment, they showed a significant preference for alternative X (.71) in Experiment 2. Because T_C pits alternatives (10xx and 01xx) that are equivalent except for the relative importance of the cause and effect, this result provides unambiguous support for the presence of a causal status effect.

The results of a logistic regression analysis of the adult data are given in Table 4. As before, in addition to parameter w_H (which provides information about the coherence effect), Table 4 reports measures corresponding to the typicality effect (*average*(w_C, w_E, w_N, w_N)), the causal status effect (w_C – w_E), and the relational centrality effect (w_E – w_N). A significant difference between a weight on the cause and effect features ($w_C - w_E = .30$) confirms the presence of a causal status effect, $t(44) = 3.94$, $p < .001$. These parameter values reflect adults' preference for alternative X on test item T_C. These results indicate that Experiment 2 succeeded in inducing a causal status effect in adults by describing the link between the cause and effect as probabilistic.

Table 4 also shows that Experiment 2 replicated the large coherence effect in adults, as indicated by a weight on parameter w_H (0.74) that was significantly greater than zero, $t(44) = 6.16$, $p < .0001$. Finally, Table 4 also indicates the absence of a relational centrality effect, $t(44) = 1.39$, $p = .17$.

Children's responses. The results of the logistic regression analysis of children's responses are presented in Table 4. Whereas adult responses in Experiment 2 differed from those in Experiment 1 in exhibiting a causal status effect, children's logistic regression parameters in Experiment 2 were largely the same as in the first experiment. First, Table 4 indicates the presence of a coherence effect as indicated by a value for parameter w_H (.31) that was significantly greater than zero, $t(45) = 3.66$, $p < .001$. Second, the absence of a positive difference between a weight on the cause feature and its effect (.20 vs. .29) reflects the absence of a causal status effect; indeed, that difference was nonsignificantly negative, $t(45) = –1.66$, $p = .10$. Finally, the difference of .08 between w_E and w_N (the relational centrality effect) did not reach significance, $t(45) = 1.59$, $p = .12$.

Comparison of adults and children. Following Experiment 1, we compared the magnitude of the effects in the two groups of subjects. First, the average weight on the four features (w_C, w_E, w_N, and w_N) indicated a larger typicality effect in adults (.64) than children (.22), $t(89) = 7.47$, $p < .0001$. Second, although both groups exhibited sensitivity to coherence, adults exhibited a larger coherence effect (w_H = .74) than children (.31), $t(89) = 2.89$, $p < .01$. Third, adults exhibited a larger causal status effect (.30) than children (–.09), $t(89) = 2.89$, $p < .0001$. Finally, the size of the relational centrality effect did not differ across groups, $t < 1$.

Discussion

The purpose of Experiment 2 was to establish conditions that induced a causal status effect in adults in order to determine whether that effect would also arise in 5–6 year old children. Describing the category's causal relation as probabilistic in Experiment 2 was sufficient to induce a large causal status effect in adults. These results are consistent with previous results showing that a causal status effect occurs for probabilistic causal relationships but not deterministic ones (Rehder & Kim, in press). In contrast, 5-6 year olds showed no sign of a causal status effect (as in Experiment 1), despite the use of probabilistic causal link. Below we discuss the theoretical significance of this persistent lack of a causal status effect in children in this age group.

In Experiment 2 both adults and children exhibited significant coherence effects. These results replicate previous findings of coherence effects in adults with probabilistic causal links (Rehder & Kim, in press) and extend them to use of a forced-choice procedure and a category with a single causal relation. More importantly, the results of Experiment 2 extend those of Experiment 1 by showing that a coherence effect obtains in children even when the causal links are described as probabilistic. We consider our findings regarding the coherence and relational centrality effects at greater length in the General Discussion.

Evaluating Theoretical Accounts of Causal-Based Classification

The preceding experiments establish the presence of new effects in children's causal-based categorization. But another goal of this article is to evaluate whether models that have been proposed as theoretical explanations of causal-based categorization in adults also apply to children. In this section we evaluate the generative and dependency models as accounts of children's causal categorization by presenting their quantitative fits to the data from Experiments 1 and 2.

Note that both models specify a rule that assigns to an object a measure of its degree of membership in a category on the basis of that category's network of interfeature causal relations. However, neither model denies the existence of other effects on classification. It is well known, for example, that feature weighting in categorization is also influenced by a feature's perceptual salience (e.g., Kruschke & Johansen, 1999; Lamberts, 1995) and *cue validity* (the extent to which it is diagnostic of that category versus another, Rosch & Mervis, 1975). Rather, the claim is that causal relations will have the predicted effects when these factors are controlled.

To reflect the fact that Experiments 1 and 2 were identical except for the information provided about causal strength (conveyed as probabilistic in Experiment 2 but not 1), for both the adults and children the models were simultaneously fit to their data from both experiments. All parameters were held fixed over the two experiments except for those that represent the strength of the causal link.

The Generative Model

The generative model builds on *causal-model theory* (Sloman, 2005; Waldmann & Holyoak, 1992) by assuming that interfeature causal relations are represented as probabilistic causal mechanisms (Rehder, 2003a, b; Rehder & Kim, 2006). The central intuition behind the generative model is that objects that are likely to have been produced, or *generated*, by a category's causal model should be considered good category members and those unlikely to be generated should be considered poor ones.

An important advantage of the generative model is that it provides an intuitively straightforward account of the coherence effect. The generative model predicts that a population of category members generated by a causal network should exhibit the expected pattern of correlations between causally related features and thus a likely category member is one that maintains those correlations. Of course, two features will be correlated when they are directly linked by a causal relation. In addition, the generative model can also predict a causal status effects depending the particular parameters associated with the category's causal model, as demonstrated below.

Quantitative predictions for the generative model can be derived assuming a particular representation of causal relations first introduced by Cheng (1997) and later applied to a variety of category-based tasks (Rehder & Hastie, 2001; Rehder, 2003a, b; Rehder, 2009; Rehder & Kim, 2006, in press; Rehder & Burnett, 2005). In the current experiments, assume that the causal mechanism relating features C and E operates (i.e., C produces E) with probability m_{CE} when C is present, and that any other potential background causes of E collectively operate with probability b_E. Given other reasonable assumptions (e.g., the independence of causal mechanisms, see Cheng & Novick, 2005), then C and the background causes form a "fuzzy-or" network that together produce E in members of category k with probability,

$$p_k(E \mid C) = m_{CE} + b_E - m_{CE}b_E \tag{8}$$

When C is absent, the causal mechanism has no effect on E and thus the probability of E is simply

$$p_k(E \mid \overline{C}) = b_E \tag{9}$$

In addition, the probability of the cause C is a free parameter c_C and the probability of the two neutral features N_1 and N_2 is given by a free parameter c_N.

From these definitions, the probability that the category generates any combination of the features C, E, N_1, and N_2 can be specified. Table 5 presents the likelihood equations for the eight test objects presented in this study. For example, the probability of 11xx is the probability the cause C is present (c_C) multiplied by the probability that the effect is produced when C is present ($m_{CE} + b_E - m_{CE}b_E$). Similarly, the probability of 00xx is the probability the cause C is absent ($1 - c_C$) multiplied by the probability that the effect is absent when C is absent ($1 - b_E$). Finally, the probabilities of the test instances with values on the neutral dimensions are straightforward functions of parameter c_N (e.g., the probability of xx11 is c_N^2, the probability of xx00 is $(1 - c_N)^2$, etc.).

Table 5
Generative model likelihood equations for the eight test instances.

Object i	$p_k(i)$
11xx	$c_C(m_{CE} + b_E - m_{CE}b_E)$
10xx	$c_C(1 - m_{CE})(1 - b_E)$
01xx	$(1 - c_C)b_E$
00xx	$(1 - c_C)(1 - b_E)$
xx11	c_N^2
xx10	$c_N(1 - c_N)$
xx01	$(1 - c_N)c_N$
xx00	$(1 - c_N)^2$

Forced-choice performance on a test instance that includes object i and j is calculated by computing the difference between $p_k(i)$ and $p_k(j)$. Because these probabilities are ≥ 0 and ≤ 1, their difference is restricted to the range [–1, 1], and so an additional scaling constant *K* was applied before passing the difference through a logistic. That is,

$$choice_k^{GenerativeModel}(i,j) = \frac{1}{1+e^{-K(p_k(i)-p_k(j))}} \quad (10)$$

As mentioned, because Experiments 1 and 2 were identical except for the causal strength information, the generative model was simultaneously fit to the results of both experiments with parameter m_{CE} allowed to vary across experiments and parameters b_E, c_N, and *K* held fixed. To incorporate our assumption that the link was interpreted as deterministic in Experiment 1 but as probabilistic in Experiment 2, m_{CE} was set to 1 for the first experiment but was allowed to vary freely in the second (thus, because m_{CE} is constrained to the range [0–1], its value in Experiment 2 must be less than or equal to its value in Experiment 1). Separate sets of parameter values that minimized squared error were obtained for the adults and children.

The Dependency Model

The dependency model (Sloman et al., 1998) characterizes the theoretical knowledge that classifiers have about categories in terms of a network of *dependency relations* among category features, where a causal relation is one type of dependency relation (an effect depends on its causes). The key intuition behind the dependency model is that features will be weighed more heavily to the extent they are strongly depended on by other features. This includes the features that they directly cause as well as those they cause indirectly through other features.

Specifically, according to the dependency model, feature *i*'s weight or *centrality* is a nonnegative real number computed from the iterative equation,

$$c_{i,t+1} = \sum d_{ij} c_{j,t} \quad (4)$$

where $c_{i,0}$ is *i*'s initial centrality *t*, $c_{i,t}$ is its centrality at iteration *t*, and d_{ij} is the strength of the causal link between *i* and its dependent *j*. The feature weights are the centralities that obtain when Equation 4 converges (when the c_is stop changing).

One important advantage of the dependency model is that it can predict a causal status effect. For example, in a category such as the one used present study in which one feature (*C*) causes another (*E*) (and C and E are involved in no other causal relations), then when $c_{E,0}$ is initialized to 1 and the causal link between C and E (d_{CE}) has a strength of 2, the centralities of C and E will converge after a single iteration to 2 and 1, respectively. Because C's centrality is greater than E's, C will have more influence on categorization decisions, that is, a causal status effect will obtain. The size of the causal status effect will increase as d_{CE} increases and to the extent C (or E) has additional effects. The dependency model has been used to successfully model the causal status effect found in adults with both natural (Kim & Ahn, 2002; Sloman et al. 1998) and experimental (Ahn et al. 2000b) categories.

The dependency model's predictions for the current experiments were derived by assuming that the evidence that an object i provides for membership in category k is an additive function of its features, each weighted by its centrality. Assume that the centralities for features C, E, and the two neutral features are c_C, c_E, and c_N, respectively. Thus,

$$ev_k^{DependencyModel}(i) = c_C f_{i,C} + c_E f_{i,E} + c_N f_{i,N_1} + c_N f_{i,N_2} \tag{5}$$

where $f_{i,j}$ is defined as above, namely, equal to +1 when the typical feature on dimension j is present, –1 when it is absent, and 0 otherwise. From Equation 4, the centralities in Equation 5 can be rewritten as functions of d_{CE} and the initial centralities of E, N_1, and N_2, yielding,

$$ev_k^{DependencyModel}(i) = d_{CE} c_{E,0} f_{i,C} + c_{E,0} f_{i,E} + c_{N,0} f_{i,N_1} + c_{N,0} f_{i,N_2} \tag{6}$$

Note that an important property of this definition of ev_k is that it fails to incorporate an interactive effect of cause and effect features (e.g., whether they are both present or absent). As a consequence, the dependency model is unable to account for coherence effects, a fact we demonstrate below.

An expression representing whether a subject chooses object i or j as the better member of category k involves computing the difference between $ev_k^{DependencyModel}(i)$ and $ev_k^{DependencyModel}(j)$ and then passing that difference through a logistic. That is,

$$choice_k^{DependencyModel}(i,j) = \frac{1}{1 + e^{-\left(ev_k^{DependencyModel}(i) - ev_k^{DependencyModel}(j)\right)}} \tag{7}$$

Just as for the generative model, the dependency model was simultaneously fit to the results of both experiments, where parameter d_{CE} was allowed to vary across experiments but parameters $c_{E,0}$ and

$c_{N,0}$ were held fixed.[3] Because there is no value of d_{CE} that a priori corresponds to a deterministic link, it was allowed to vary freely in Experiment 1 but was constrained be larger than the value in Experiment 2. Parameter values that minimized squared error were obtained separately for the adult and child participant groups.

Model Fitting Results

The best fitting parameters for the two models are presented in Table 6. Figure 3A presents the generative and dependency model fits to adults' responses to the 14 test items (7 each from Experiments 1 and 2). Figure 3B does the same for the children.[4] We discuss the results from the two subject groups separately.

Table 6
Parameter values resulting from fits of the generative and dependency model to the data of Experiments 1 and 2.

Parameter	Adults	5–6 Year Olds
Generative model		
c_C	.84	.78
m_{CE} [for Expt. 2]	.90	1
b_E	0	.14
c_N	.76	.74
K	3.40	1.35
R^2	.95	.91
RMSE	.054	.037
Dependency model		
d_{CE} [for Expt. 1]	1.22	1.10
d_{CE} [for Expt. 2]	1.22	.73
c_E	.48	.22
c_N	.43	.16
R^2	.68	.72
RMSE	.131	.063

Model fits for the adults. As is apparent from Figure 3A, the generative model provides a generally good account of the adult data from both Experiments 1 and 2. Most prominently, it accounts for the effect of coherence observed in both experiments by yielding choice probabilities strongly favoring alternative Y in test items T_D and T_E and X for T_F and T_G. It does so because the large values of the m_{CE} parameter in both experiments leads the model to favor items that manifest the expected correlation between features *C* and *E* (i.e., 11xx and 00xx) and disfavor items that break those correlations (i.e., 10xx and 01xx).

The generative model was also able to reproduce the different pattern of feature weights that obtained in the two experiments. Specifically, the fit yielded a choice probability of .50 and .59 for test item T_C in Experiments 1 and 2, respectively, reflecting the presence of a causal status effect in the

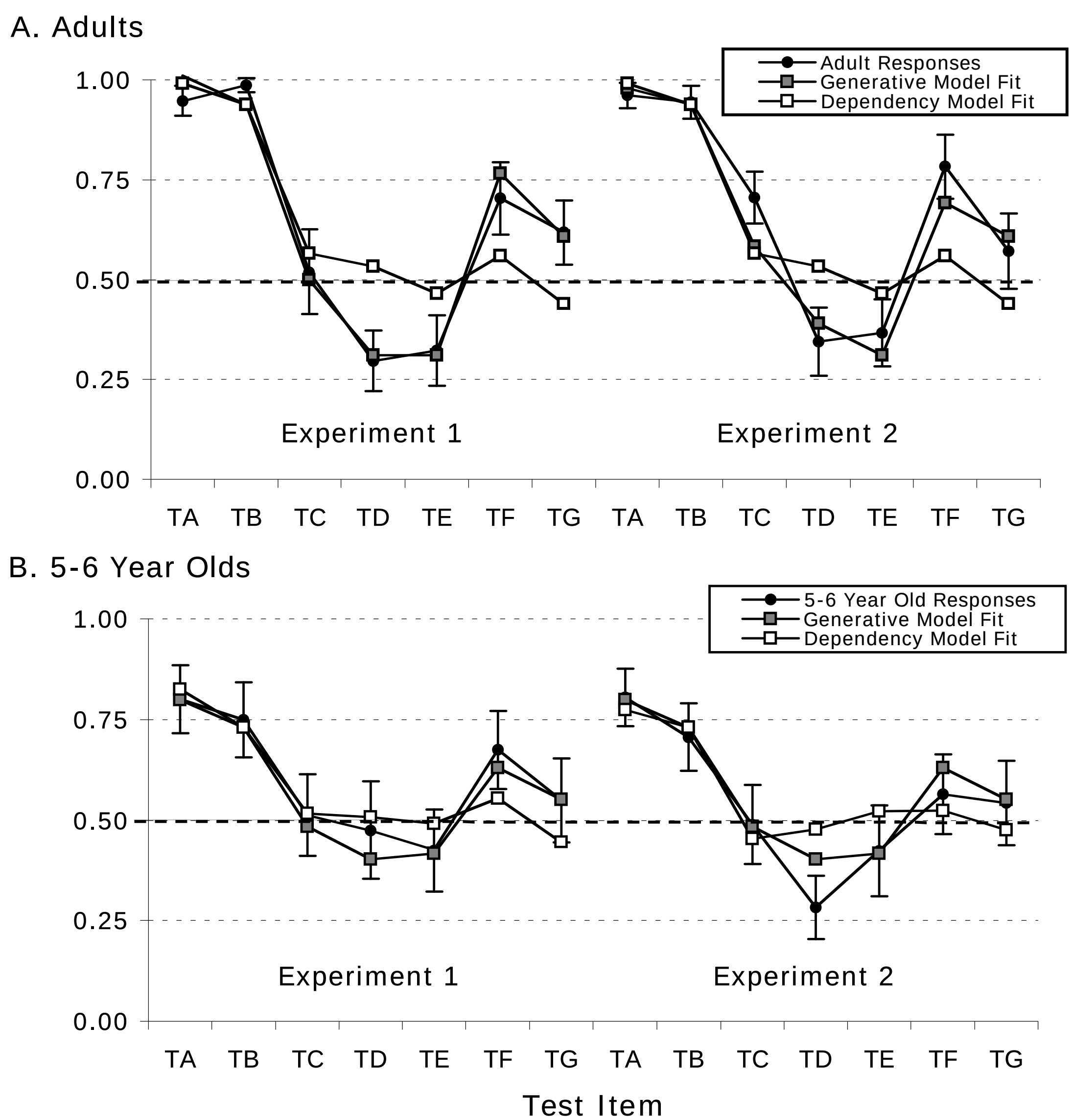

Figure 3. Fits of the generative model to the data of (A) adults and (B) 5-6 year olds from Experiments 1 and 2. Error bars are 95% confidence intervals.

second experiment but not the first (although note the fitted value of .59 undershoots Experiment 2's observed value of .71). It is able to predict this qualitative pattern because the value for the m_{CE} parameter in the second experiment (.90) is lower than in the first (1.0). The lower value of m_{CE} yields a causal status effect because the cause feature C is less likely to "generate" the effect feature E. Thus, E is less prevalent in category members, which in turn means it is given less weight in categorization decisions.[5] Note that even with this lower value of m_{CE}, the causally linked features C and E should still be highly correlated and thus the model continues to yield a coherence effect on items T_D–T_G for Experiment 2.

In contrast, Figure 3A shows that the dependency model yielded a much poorer fit to the adult data. One reason for this poorer fit is that the dependency model fails to predict an effect of coherence, and so is unable to reproduce the response pattern Y, Y, X, and X for test items T_D–T_G. The model is also unable to reproduce the relative size of the causal status effect in the two experiments. As mentioned, the dependency model predicts a larger causal status effect for stronger causal links. However, the observed choice probabilities showed the opposite pattern: there was a larger causal status effect for weaker causal links (Experiment 2) relative to stronger ones (in Experiment 1). When fit to these data, the model is thus forced to produce a modest causal status effect for *both* experiments. This behavior is reflected in a value for the d_{CE} parameter that is the same in the two experiments (1.22) and that yields the same modest difference in the importance of the cause feature C ($c_C = d_{CE}c_E = 1.22 \times .48 = .59$) and the effect feature E ($c_E = .48$).

The better fit of the generative model apparent in Figure 3A is also supported by quantitative measures of degrees of fit. The generative model explains 95% of the variance in the adults' responses as compared to 70% for the dependency model. The better fit also holds according to a measure, *RMSE*, that corrects for the different number of parameters in the two models (.054 vs. .128).[6]

Model fits for the 5-6 year olds. Comparison of Figures 3A and 3B reveals that the better fit of the generative model is less prominent for the children versus the adults. Nevertheless, because the children's responses were a less extreme version of the pattern exhibited by the Experiment 1 adults, the generative model but not the dependency model is able to reproduce the empirical choice probabilities that reflect a coherence effect, namely, Y, Y, X, and X for test items T_D–T_G (although note that the fitted value of .40

for T_D in Experiment 2 overshoots the observed value of .28). This better fit is reflected in both total variance explained (R^2s of .91 and .72 for the generative model and the dependency model, respectively) and *RMSE*s (.037 vs. .063).

Both models account for the children's lack of a causal status effect, although for different reasons. The dependency model does so with values of d_{CE} in both experiments that do not greatly exceed 1. As a result, features *C* and *E* are weighed about the same, thus explaining children's agnostic performance on test item T_C. The generative model explains the lack of a causal status effect with deterministic causal links for both experiments. When m_{CE} = 1.0, the effect feature *E* is "generated" at least as often as *C* and thus will be weighed as least as heavily as *C* in categorization decisions. Importantly, that the fitted value for m_{CE} in Experiment 2 was 1.0 for children but .90 for adults suggests why only adults exhibited a causal status effect in that experiment. These m_{CE} values are consistent with the view that adults but not children perceived the causal link in Experiment 2 as probabilistic.

We make two other observations about the parameters yielded by the children's model fits. First, those parameters also reflect the small (and nonsignificant) negative causal status effect exhibited by the children in Experiment 2. The dependency model accounts for this with a value of .73 for parameter d_{CE} that results in the weight for feature *E* (c_E = .22) being slightly larger than *C* ($c_C = d_{CE}c_E$ = .73 × .22 = .16). The generative model accounts for it with a value of .14 for parameter b_E that results in feature *E* being a bit more probable than *C*.[7] Second, both models capture the less extreme responding in the children as compared to the adults. The dependency model does so with smaller overall feature weights (e.g., values for c_N of .16 vs. .43). The generative model does so with a smaller value of *K* for the children (1.35 vs. 3.40 for the adults).

Summary. Although it mispredicts a few specific items (e.g., T_C for adults in Experiment 1, T_D for children in Experiment 2), the generative model provides a better account of the overall pattern of responses in these experiments as compared to the dependency model. It also provides a better quantitative account, suggesting that, just like adults, 5-6 year old children decide whether an object is a likely category member by judging whether it was likely generated by the category's causal model. Nevertheless, a few cautionary notes are in order. One is the relatively low ratio between the number of parameters used by the generative model (5) and the number of data points being fit (14). Another is that the two models were compared on only one causal network and that network was extremely simple (one causal link). Clearly, future research that presents a larger number of test items and tests more complex networks of causal relations is desirable. Nevertheless, the broader point we would like to make is that these analyses provide initial evidence that a model of causal-based classification shown to account for many aspects of adult performance may also apply to 5-6 year old children. They also demonstrate how the sorts of model fitting and comparisons carried out in adult studies can also be applied to children making forced-choice judgments.

General Discussion

This study tested for the presence of causal coherence, causal status and relational centrality effects in adult and child categorization. Unlike previous studies (e.g., Ahn et al., 2000a; Meunier & Cordier, 2009), our design and logistic regression analysis allowed us to evaluate the independent contribution of each of these effects to categorization. In adults, the effect of causal relations on categorization depended on the perceived strength of the causal relation. Consistent with previous work (Rehder & Kim, 2006, in press), when causal relations could be interpreted as deterministic, categorization was dominated by causal coherence. Instances where cause and effect features were both present, or both absent, were more likely to be accepted as category members, while incoherent instances (cause present, effect absent or vice versa) were less likely to be accepted. In contrast, when causal relations were explicitly described as probabilistic, adults showed an additional effect of causal status in which a cause was given more weight in categorization judgments than its effect.

These results replicate and extend previous work showing causal status and coherence effects in adult categorization (Ahn, et al., 2000; Rehder & Hastie, 2001; Rehder 2003a, b) and induction (e.g., Rehder & Hastie, 2004). They also reinforce previous findings that both effects can be found concurrently for a given set of causally related category members (Rehder & Kim, in press).

These findings are theoretically significant because only certain models can explain concurrent effects of causal coherence and causal status. First, Rehder's (2003a, b; Rehder & Kim, 2006) generative model assumes that people expect that the respective roles of cause and effect features that exist in familiar category members should remain stable in novel instances. Hence, people will be more likely to accept coherent instances that preserve the specific pattern of causal relations observed in other category members, and reject incoherent instances that violate these patterns. This pattern was observed in Experiments 1 and 2. Second, the generative model predicts that the magnitude of the causal status effect should be larger for probabilistic versus deterministic relations. It predicts this because when a cause feature only sometimes generates an effect feature, it will be more prevalent in category members than the effect (assuming that the effect has no alternative causes). This in turn means that the cause will be more diagnostic of category membership (see Rehder & Kim, in press, for a formal derivation). Consistent with this prediction, a causal status effect was observed in Experiment 2 but not Experiment 1. As a consequence of these predictions, the generative model yielded generally good quantitative fits to both experiments. These findings support the claim that adult classifiers judge whether an object was generated by a category's causal model and decide category membership on the basis.

In contrast, the dependency model (Sloman, et al., 1998) is consistent with neither of these findings. First, that model is unable to predict an effect of coherence. Second, while it is able to predict a causal status effect, it predicts that that effect should be stronger for stronger causal links, because a stronger link means that the effect depends more heavily on the cause. In contrast, we found the opposite pattern. A causal status effect was only observed when the causal links were explicitly described as probabilistic (in Experiment 2). As a result, model fitting revealed that the dependency model yielded poor quantitative fits and was unable to reproduce important patterns in adults' choices. These results are consistent with previous demonstrations showing that the presence of interfeature causal relations does not necessarily make "depended on" features more important to category membership (Rehder & Kim, 2006, in press).

A final notable result from the adult data was absence of a relational centrality effect. This result contrasts with previous adult studies that found that neutral features were weighed less heavily than the terminal effect feature in a three-element causal chain (Kim & Ahn, 2002a; b). Causally related features are also weighed more heavily than neutral ones when they are involved in more complex networks involving four features (Rehder & Kim, 2005, reported in Rehder, 2010). It may be that a strong effect of relational centrality only holds when features are involved in larger causal networks (i.e., those involving more than one causal link).

The interpretation of children's categorization patterns was less straightforward, with many of their test phase category assignments indistinguishable from chance. Nevertheless, logistic regression analyses revealed that five and six year olds showed a reliable effect of category coherence for both deterministic and probabilistic causal relations. In contrast with the adults, there was no evidence of an independent contribution of causal status to children's category judgments. Like the adults, children exhibited no effect of relational centrality.

These results provide further support for the view that young children use causal knowledge about feature relations to guide categorization of novel instances (cf. Ahn et al., 2000a; Hayes et al., 2003). However, our explanation for the mechanism that drives causal categorization in children differs from that advanced in earlier work. Previous work has shown that children think that an object missing an effect is a better category member than one missing a cause (Ahn et al., 2001; Hayes et al., 2003; Meunier and Cordier, 2009; Opfer & Bulloch, 2007). Ahn et al. and Meunier and Cordier interpreted this as evidence for the early emergence of causal status as a crucial factor in children's categorization. The current work suggests, however, that these previous findings can be equally well explained by assuming the early development of sensitivity to causal *coherence*. Our work shows that young children are sensitive to whether patterns of causes and effects are consistent with the causal model for a given category. Instances that contained a cause and its expected effect, and those lacking both the cause and effect, were treated as better category members than would be predicted by a simple effect of typicality alone. Conversely, children often rejected instances that violated the causal model through the presence of a causal feature without the effect, or vice versa.

When the effects of causal coherence were controlled, no evidence was found of an independent contribution of causal status to children's categorization. Because the generative model is the only one that predicts an effect of causal coherence, it offers the best account of not only adult's categorization in causally rich domains, but children's as well. This conclusion is reinforced by our model comparisons that showed that the generative model consistently provided a better fit to children's categorization responses than the dependency model.

It should be acknowledged that whereas we taught our subjects a single link between two features, Ahn et al. and Meunier and Cordier taught theirs a common cause structure in which one feature caused two others (Figure 1A). It may be that a cause with at least two effects is necessary to induce a causal status effect in children. Arguing against this possibility, however, are findings indicating that, at least for adults, a feature's importance does not generally increase with its number of dependents (Rehder & Kim, 2006). It is also important to note that Ahn et al. tested children that were somewhat older (7-9 years old) than those in the present study (5-6 years old) and of course it is possible that children at this older age in fact exhibit a causal status effect, just like adults do. We now consider possible reasons why adults but not children exhibited a causal status effect in our Experiment 2 and what implications those potential explanations have for the development of causal-based categorization.

The Generative Model, Essentialism and Determinism

The current studies reveal both continuities and discontinuities in the development of causal categorization. The data and modeling suggest that an ability to understand causal generative mechanisms emerges at a relatively early age and continues to influence categorization during adulthood. Indeed, both categorization data and model fitting suggested that the effect of coherence increases with age.

It is interesting to compare the finding of the early emergence of a coherence effect with previous work on the early emergence (and persistence) of essentialist thinking in categorization (see Gelman, 2003 for a review). Essentialism entails an assumption that that the observable features of category members are caused by a less obvious core feature, common to category members. Children as young as three years of age have been shown to hold essentialist assumptions about many biological and social categories (Gelman, 2003; Taylor, Rhodes & Gelman, 2009). The current findings are consistent with this work in that children, as young as five years of age, expected stable patterns of cause and effect within members of the same category. Notably, however, whereas essentialism focuses on belief in a shared *underlying cause* for category membership, the current results highlight children's sensitivity to cause and effect *relations*. To be accepted as a category member an instance had to have an underlying causal feature *together with* a plausible effect that could be generated by the cause.

An important developmental change found in the current study was that the strength of causal relations between features had a marked effect on adult categorization but little effect on the way that 5-6 year olds categorized. For adults, the presence of probabilistic relations in Experiment 2 was associated with both causal status and coherence effects while causal coherence remained the primary basis for children's categorization. One explanation for this developmental difference is that children did not understand the probabilistic language used during stimulus presentation in Experiment 2. However, this seems unlikely since previous work has shown that, by the time they reach 1st grade, a majority of children understand the different implications of sentences like "We will *probably* go ice skating this afternoon" as opposed to "We will *definitely* go ice skating this afternoon" (Hoffner, et al., 1990). Moreover, a follow-up analysis found few differences between the patterns of classification shown by younger and older children in either experiment.[8]

An alternative explanation is that although children have some grasp of probabilistic language, they are nevertheless biased in favor of interpreting *causal* relations as deterministic rather than probabilistic. A number of lines of evidence suggest that, from an early age, children are "causal determinists" (see Gelman & Kalish, 1993, and Shultz & Somerville, 2006 for reviews). Infants and pre-schoolers expect observed events to have causes (which may be unobservable), and are surprised by events that have no apparent cause (e.g., Gelman, Coley, & Gottfried, 1994). Young children also generally expect that causal relations will be deterministic rather than probabilistic. Shultz and Somerville (2006) found that four-year-olds were generally reluctant to endorse a stochastic relationship between a causal variable and an effect, even when this pattern was consistent with observed patterns of covariation. By contrast, adults have a better (but by no means perfect) appreciation of the complex probabilistic nature of many causal relations (Sloman, 2005).

The current data are consistent with the view that young children generally treat direct cause and effect relations between object features as deterministic. More importantly, though, our results suggest a refinement of the child-as-causal-determinist hypothesis. Both experiments showed that children's classification of novel instances was primarily determined by an expectation of causal *coherence*. So not only do children show a bias to interpret causal relations as deterministic, they also expect causes and effects to conform to a specific causal mechanism. Hence, any object that had a causal feature like "big lungs" was expected to have observable features that could plausibly arise from that cause (like staying underwater for long periods).

It should be noted that the current experiments only explored one factor that promotes sensitivity to causal status. Previous work with adults has shown that a number of other factors besides causal strength influence whether or not causal status contributes to categorization judgments (see Rehder, 2010 for a review). First, a causal status effect between two features *C* and *E* tends to be stronger when there is an unobserved "essential" feature that is common to all category members and that causes *C* (Rehder, 2003b; Rehder & Kim, 2007; in press). This occurs because (a) the essential feature makes *C* more prevalent in category members and this in turn amplifies the difference in prevalence between *C* and *E* and (b) classifiers reason directly from *C* to the essential feature. Nevertheless, evidence that even young children already exhibit "essentialist" intuitions (Gelman, 2003; Taylor et al., 2009) makes this an unlikely explanation for the difference between adults and children in Experiment 2. Second, a causal status effect tends to be weaker to the extent that a classifier assumes alternative causes of *E*. Feature *C* will be more prevalent than *E* when it only probabilistically generates *E*, but this effect can be ameliorated (or reversed) if *E* has alternative causes that are numerous or strong. Again, however, we find this an unlikely explanation of the results from Experiment 2 because it requires assuming that children's causal model for the category was more complicated than adults' (i.e., that children were more likely to assume the presence of alternative causes of *E*).

In summary then, we suggest that the current results imply a developmental pattern such that in the earliest stages of causal categorization young children treat causal relations as deterministic, and as a result their category decisions are based primarily on the coherence between the causal features of category members. A better understanding of probabilistic causal relations has emerged by adulthood, leading to additional effects of causal status, as predicted by the generative model. An important goal for future work is to conduct a more fine-grained analysis of the trajectory of causal categorization over the elementary school years, with a view to identifying just when children begin to factor in the effects of causal status and centrality.

The Generative Model and Bayesian Approaches

The generative model shares many features with Bayesian approaches to causal reasoning (e.g., Gopnik et al., 2004; Sloman, 2005; Tenenbaum & Griffiths, 2003). According to the generative model, causal relations between the features of category exemplars are represented by a network of probabilistic causal links. Classification of a novel object is then based on an estimation of the likelihood that an object's features could have been generated by the causal model. This process could be seen as analogous to calculating Bayesian posterior probabilities after new observations. In fact, Rehder (2003a, b) acknowledges that the causal models assumed by the generative model can be construed as Bayesian networks (cf. Pearl, 2000). This analogue is particularly interesting because of recent suggestions that even very young children follow Bayesian principles in their learning of causal relations from data (Gopnik et al., 2004; Sobel & Buchanan, 2009), in their inductive inferences (Kemp & Tenenbaum, 2009) and extension of word meaning (Xu & Tenenbaum, 2007). Our results could be seen as further evidence for the early development of a Bayesian computational architecture that allows children to capitalize on existing knowledge when making judgments about novel stimuli. Note, however, that the current work goes well beyond the suggestion that young children operate as efficient Bayesians. Bayesian networks, by themselves, convey no information regarding the details of the causal relationships that link variables in a network. In contrast, the generative model makes specific assumptions about the functional form of causal relationships; namely, that people view features as being linked by probabilistic causal mechanisms and expect that when a cause feature is present, it implies the operation of a causal mechanism that will, with some probability, bring about the presence of an effect feature.

Conclusions

The current studies challenge existing accounts of the way that causal knowledge affects children's categorization. We found that 5-6 year olds' category membership judgments were strongly influenced by the coherence between cause and effect features. Contrary to existing accounts (e.g., Ahn et al., 2000; Meunier and Cordier, 2009) children showed little evidence of an independent effect of causal status. These results suggest that the generative model that has been successful in explaining a range of phenomena in adult categorization and induction (Rehder, 2010), also provides a good account of causal categorization in young children.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69*, 135-178. doi: 10.1016/S0010-0277(98)00063-8

Ahn, W., Gelman, S. A., Amsterlaw, A., Hohenstein, J., & Kalish, C. W. (2000a). Causal status effect in children's categorization. *Cognition, 76*, B35-B43. doi: 10.1006/pmed.2000.0716

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000b). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361-416. doi: 10.1006/cogp.2000.0741

Ahn, W., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., Coley, J. D. & Shafto, P. (2001). Why essences are essential in the psychology of concepts. *Cognition, 82*, 59-69. doi: 10.1016/S0010-0277(01)00145-7

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

Carmichael, C., & Hayes, B. K. (2001). Prior knowledge and exemplar encoding in children's concept acquisition. *Child Development, 72,* 1071-1090. doi: 10.1111/1467-8624.00335

Chaigneau, S. E., Barsalou, L. W., & Sloman, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology. General, 133*, 601-625. doi: 10.1037/0096-3445.133.4.601

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367-405. doi: 10.1037/0033-295X.104.2.367

Cheng, P. W. & Novick, L. R. (2005). Constraints and nonconstraints in causal learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review, 112i, 694-707. doi:* 10.1037/0033-295X.112.3.694

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

Gelman, S. A. & Kalish, C.W., (1993). Categories and causality. In R. Pasnak, & M. L. Howe (Ed.) *Emerging themes in cognitive development* (pp. 3–32). New York, NY: Springer

Gelman, S. A., Coley, J. D. & Gottfried, G. M. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. A. Hirschfeld & S. A. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture* (pp. 341-365). New York, NY, US: Cambridge University Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes Nets. *Psychological Review, 111*, 3-32. doi: 10.1037/0033-295X.111.1.3

Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory & Cognition, 35*, 1785-1800.

Hayes, B. K. (2006). Knowledge, development and category learning, In B. Ross (Ed). *The Psychology of Learning and Motivation, 46*, 37-78. doi: 10.1016/S0079-7421(06)46002-3

Hayes, B. K., Foster, K., & Gadd, N. (2003). Prior knowledge and subtyping effects in children's category learning. *Cognition, 88*, 177-199.

Hayes, B. K., & Thompson, S. P. (2007). Causal relations and feature similarity in children's inductive reasoning. *Journal of Experimental Psychology: General, 107,* 470-484. doi: 10.1037/0096-3445.136.3.470

Hoffner, C., Cantor, J., & Badzinski, D. M. (1990). Children's understanding of adverbs denoting degree of likelihood. *Journal of Child Language, 17*, 217-231. doi: 10.1017/S0305000900013192

Kemp, C. & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116*, 20-58. doi: 10.1037/a0014282

Kim, N. S., Ahn, W. (2002). Clinical psychologists' theory-based representations of metal disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131*, 451-476.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25,* 1083-1119.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General, 124*, 161-180.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 754-770. doi: 10.1037/0278-7393.22.3.754

Marsh, J. K. & Ahn, W. (2006). Order effects in contingency learning: The role of task complexity. *Memory & Cognition, 34*, 568-576.

Meunier, B., & Cordier, F. (2009). The biological categorizations made by 4 and 5-year olds: The role of feature type versus their causal status. *Cognitive Development, 24*, 34-48. doi: 10.1016/j.cogdev.2008.05.003

Opfer, J. E., & Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition, 105*, 206-217. doi 10.1016/ j.cognition.2006.08.006

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Piaget, J., & Inhelder, B. (1973). *Memory and intelligence*. New York, NY: Basic Books.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27*, 709-748. doi: 10.1016/S0364-0213(03)00068-5

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1141-1159. doi: 10.1037/0278-7393.29.6.1141

Rehder, B. (2006). When causality and similarity compete in category-based property induction. *Memory & Cognition, 34*, 3-16.

Rehder, B. (2009). Causal-based property generalization. *Cognitive Science*, *33*, 301-344.

Rehder, B. (2010). Causal-based categorization: A review. In B. Ross (Ed.) *Psychology of Learning and Motivation, 52*, 40-116.

Rehder, B. & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264-314. doi: 10.1016/j.cogpsych.2004.09.002

Rehder, B. & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, *130*, 323-360. doi: 10.1037/0096-3445.130.3.323

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

Rehder, B. & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 659-683. doi: 10.1037/0278-7393.32.4.659

Rehder, B. & Kim, S. (2009). Classification as diagnostic reasoning*. Memory & Cognition, 37,* 715-729. doi: 10.3758/MC.37.6.715

Rehder, B. & Kim, S. (in press). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605. doi: 10.1016/0010-0285(75)90024-9

Shultz, L. E. & Somerville, J. (2006). God Does Not Play Dice: Causal determinism and preschoolers' causal inferences. *Child Development, 77*, 427 – 442

Sloman, S. (2005). *Causal models*. New York, NY: Oxford University Press.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science, 29*, 5-39. doi: 10.1207/s15516709cog2901_2

Sloman, S. Love, B. C. & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228. doi: 10.1016/S0364-0213(99)80039-1

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. doi: 10.1037/0096-3445.133.2.166

Sobel, D. M. & Buchanan, D. W. (2009). Bridging the gap: Causality-at-a-distance in children's categorization and inferences about internal properties. *Cognitive Development, 24*, 274-283.

Springer, K. (2001). Perceptual boundedness and perceptual support in conceptual development. *Psychological Review, 108*, 691-708. doi: 10.1037/0033-295X.108.4.691

Taylor, M G., Rhodes, M., & Gelman, S. A. (2009). Boys Will Be Boys; Cows Will Be Cows: Children's essentialist reasoning about gender categories and animal species. *Child Development, 80*, 461-481. doi: 10.1111/j.1467-8624.2009.01272.x

Tenenbaum, J.B., & Griffiths, T.L. (2003). Theory-based causal inference. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* (15; pp. 35–42). Cambridge, MA: The MIT Press.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 216-227. doi: 10.1037/0278-7393.31.2.216

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian Inference, *Psychological Review, 114*, 245–272.

Appendix

Examples of additional causal and non causal features used in experimental categories

	Typical Instance (1111)				*Atypical Instance (0000)*			
Category label	Dimension *C*	Dimension *E*	Dimension N_1	Dimension N_2	Dimension *C*	Dimension *E*	Dimension N_1	Dimension N_2
Dax	Has a small brain	Forgets things a lot	Has long fur	Catches fish	Has a large brain	Always remembers things	Has short fur	Catches birds
Waddo	Small eyes	Not good at seeing in the dark	Curly tail	Eats leaves	Large eyes	Can see in the dark	Straight tail	Eats fruits
Blik	Large protective shell	Can stay in sun for a long time	Blunt teeth	Sleeps under rocks	No shell	Can't stay in the sun for long	Sharp teeth	Sleeps on top of rocks

Author Note

Brett K. Hayes, School of Psychology, University of New South Wales, Sydney, Australia. Bob Rehder, Department of Psychology, New York University. Correspondence concerning this article should be sent to Brett K. Hayes, School of Psychology, University of New South Wales, Sydney, Australia, 2052 (E-mail: B.Hayes@unsw.edu.au).

The authors would like to thank the staff and students at St Mark's Primary School, Drummoyne, Kincoppal School, Rose Bay and Mt. Sinai College, Maroubra, for their enthusiastic participation in these studies. We also thank Brooke Hahn, Lisa Shapiro, Ben Walker and Helen Paton for their assistance with data collection. We thank Marjorie Rhodes for comments on this research.

Footnotes

1 Because of the small sample size associated with each test item for each subject (4 for the adults, 2 for the children), average choice proportions for each test item were adjusted according to the following Bayesian procedure. Observing a subject choose the X alternative *x* times for an item is treated as the probability of *x* successes in a number of trials which can result in success or failure. Define π to be the true probability of choosing X. To estimate $p(\pi|x)$, the probability distribution of π given *x* successes, Bayes' theorem was applied,

$$p(\pi \mid x) = p(x \mid \pi) p(\pi)$$

where $p(\pi)$ is the prior distribution of π and $p(x|\pi)$ is the binomial distribution. A non-informative prior distribution of π was assumed (Box & Tiao, 1973), that is, a prior which makes the weakest assumptions: A beta distribution with parameters α = 0.5 and β = 0.5. α may be interpreted as representing the number of prior successes and β as the number of prior failures. The resulting posterior is a beta distribution with parameter $\alpha = x + 0.5$ and $\beta = (N - x) + 0.5$ where *N* is the total number of trials. The adjusted choice proportions were taken as the mean of the posterior. Because the mean of a beta distribution is given by $E(\pi) = \alpha / (\alpha + \beta)$, the adjusted choice proportions *choice'* were calculated as follows,

$$choice' = \frac{x + .5}{\left[(x + .5) + ((N - x) + .5)\right]} = \frac{x + .5}{N + 1}$$

where $N = 4$ for the adults and 2 for the children. As a result of this adjustment, adults' choice proportions of 0, .25, .50, .75, and 1 were converted to .10, .30, .50, .70, and .90, respectively, and children's choice proportions of 0, .50, and 1 were converted to .167, .50, and .833. Note that performing the logistic regressions without this adjustment produced qualitatively identical results.

2 In order to check that adults saw the meaning of "sometimes" as similar to "tends to", we also ran a group of eight adults (graduate students) through the full train/test procedure using the "sometimes" wording that was used with children. The results were qualitatively similar to those of the 45 adults reported in Experiment 2. In particular, for these eight subjects a logistic regression analysis yielded parameter values of $w_C = 1.33$, $w_E = .61$, $w_N = 54$, and $w_H = 1.05$. That is, they exhibited a causal status

effect, a coherence effect, and no relational centrality effect. Hence, "sometimes" appears just as effective as "tends to" in conveying a probabilistic causal relation.

3 Unlike the generative model, the dependency model was not given a scaling parameter *K* because such a parameter would be absorbed by the feature centralities.

4 As was the case for the logistic regressions, the dependency and generative models were fit to adjusted choice proportions (Footnote 1). Thus, for purposes of creating Figure 3, to make each model prediction $\hat{p}$ comparable to the original (untransformed) empirical choice proportion, those predictions were converted back into the original metric by applying the expression $\left(\hat{p}(N+1)-.5\right)/N$, where *N* is the number of trials per subject (4 for the adults, 2 for the children).

5 To illustrate, according to the generative model the probability of the effect feature *E* in category *k*, $p_k(E)$, is given by

$$p_k(E) = p_k(E \mid C)p_k(C) + p_k(E \mid \overline{C})p_k(\overline{C})$$

Applying Equations 8 and 9, this reduces to,

$$p_k(E) = c_C m_{CE} + b_E - c_C m_{CE} b_E$$

For the adults, $b_E = 0$ and so,

$$p_k(E) = c_C m_{CE}$$

The fitted value for m_{CE} of .90 for the adults in Experiment 2 yields $p_k(E) = .76$. That is, *E* is less prevalent than *C* (.84), reproducing that experiment's causal status effect (and the preference for alternative X in test item T_C). In contrast, in Experiment 1 $m_{CE} = 1$ and so *C* and *E* are equally prevalent, reproducing the absence of a causal status effect in that experiment.

6 *RMSE* = SQRT (*SSE*/(*n* – *p*)) where *SSE* = sum of squared error for a participant, *n* = number of data points fit (14), and *p* = a model's number of parameters (*p* = 5 for the generative model and 4 for the dependency model).

7 Substituting the generative model's parameters for the children ($c_C = .80$, $m_{CE} = 1.0$, and $b_E = .14$) into the expression for $p_k(E)$ in Footnote 5 yields .83, that is, *E* is (slightly) more prevalent in category members than *C* (.80).

8 In a follow-up analysis, children in each study were split into two groups according to median age. Mean preference for X on the seven test pairs was recalculated for each age group. In both studies younger (< 68 months) and older children showed the same patterns of category preference as in Table 2 (with the same test pairs reliably accepted or rejected and the same null effects on other pairs). The only exception was that the older group in Experiment 2 showed an additional preference for assigning T_G to X (M = .68, p =.02). This is evidence of a further strengthening of the coherence effect in older children.