COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 36 (2012) 1102–1128

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2012.01244.x

# The Development of Causal Categorization

Brett K. Hayes,[a] Bob Rehder[b]

[a]*School of Psychology, University of New South Wales*
[b]*Department of Psychology, New York University*

Received 12 May 2011; received in revised form 16 October 2011; accepted 17 October 2011

**Abstract**

Two experiments examined the impact of causal relations between features on categorization in 5- to 6-year-old children and adults. Participants learned artificial categories containing instances with causally related features and noncausal 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 when links were probabilistic (Experiment 2). Children's classification was based primarily on causal coherence in both cases. There was no effect of relational centrality in either age group. These results suggest that the generative model (Rehder, 2003a) provides a good account of causal categorization in children as well as adults.

*Keywords:* Categorization; Causal reasoning; Cognitive development

## 1. Introduction

It is well established that causal knowledge plays an important role in adult categorization. Adults are more likely to assign an object to a category if it has the same causal features as known category members (Rehder, 2003a; Rehder & Hastie, 2001; Sloman, Love, & Ahn, 1998). They often predict the features of category members on the basis of causal relations (Rehder & Burnett, 2005; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005) and use those inferences to establish category membership (Chaigneau, Barsalou, & Sloman, 2004; Rehder & Kim, 2009).

Recent evidence suggests that young children are also sensitive to the role of causal features in categorization (e.g., Gelman & Kalish, 1993; Gopnik et al., 2004; Hayes, 2006;

Correspondence should be sent to Brett K. Hayes, School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: b.hayes@unsw.edu.au

Hayes, Foster, & Gadd, 2003; Opfer & Bulloch, 2007). In particular, Ahn, Gelman, Amsterlaw, Hohenstein, and Kalish (2000) found that causal knowledge taught to 7- to 9-year-olds influenced judgments of category membership. These children were told, for example, that a fictitious animal called a Taliboo had promicin in their nerves, thick bones, and large eyes, and that the thick bones and large eyes were *caused by* the promicin (see Fig. 1A). Ahn et al. then presented children with two test 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 Fig. 1B). Meunier and Cordier (2009) found a similar effect with 5-year-olds using a related design.

Despite this evidence, at least three critical issues concerning the development of causal categorization remain unresolved. First, there is still a lack of consensus concerning exactly *how* causal knowledge influences categorization judgments, especially in children (Hayes, 2006; Sloman & Lagnado, 2005). This, in turn, has led to an ongoing debate about the cognitive mechanisms that underlie causal categorization (cf., Ahn & Kim, 2001; Rehder, 2010). Finally, the nature and extent of developmental change in these mechanisms has yet to be examined in detail. The current studies address each of these issues.

### 1.1. 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 colleagues have documented a *causal status effect* (Ahn, Kim, Lassaline, & Dennis, 2000; Sloman et al., 1998; also see Rehder, 2003a; Rehder & Kim, 2006, 2010). 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. (2000a) created novel categories consisting of instances with three typical features that were related in a causal chain (e.g., Roobans ''eat fruit'' that causes them to ''have sticky feet'' that in turn causes them to be good at ''building nests''). Adults were then shown novel items with two typical and one atypical

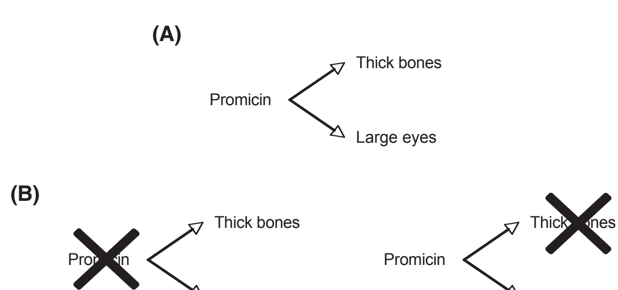



Fig. 1. (A) Schematic representation of one of the causal networks of category features taught to children in Ahn et al. (2000a). (B) One of the test trials in which children chose which of two items was a better category member. Crossed-out features were described to subjects as missing.

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 when compared with 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 Fig. 1B). This may have reflected their belief that the cause feature provided more evidence for category membership than the effect.

Second, several studies have documented a causal *coherence effect* in adult categorization (Ahn, Marsh, Luhmann, & Lee, 2002; Malt & Smith, 1984; Marsh & Ahn, 2006; Murphy & Wisniewski, 1989; Rehder & Hastie, 2001; Rehder, 2003a, 2003b; Rehder & Kim, 2006, 2010). 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, 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 Fig. 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).[1] Children's preferences in these studies could have reflected a 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 Fig. 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 item 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 (2002) 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, children 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 response measure (forced-choice) when compared with 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.

### *1.2. Theories of causal categorization*

The question of precisely how causal knowledge affects children's categorization is important because the answer has implications for whether theoretical accounts proposed to explain causal categorization in adults can be generalized to children. One of the first accounts of how causal relations affect categorization and inference was Sloman et al.'s (1998) *dependency model*. This model assumes that category knowledge can be represented as a network of dependency relations among features, where a causal relation represents a special type of relation (an effect depends on its causes). A key assumption of the model is that features will be weighed more heavily to the extent they are strongly depended on by other features. Features that are at lower levels in the causal chain have more dependency relations (i.e., there will be a larger number of effect features that depend on the presence of these causes) than features higher up the chain. Hence, the model predicts a causal status effect in categorization and that the size of that effect should increase as the strength of the causal relations increase. Notably though, because the model assumes that an object's degree of category membership is a (weighted) sum of its features (with causes weighing more than effects), it does not predict any *interactive* effects of features (e.g., a cause and effect both present or both absent). That is, it does not predict coherence effects.

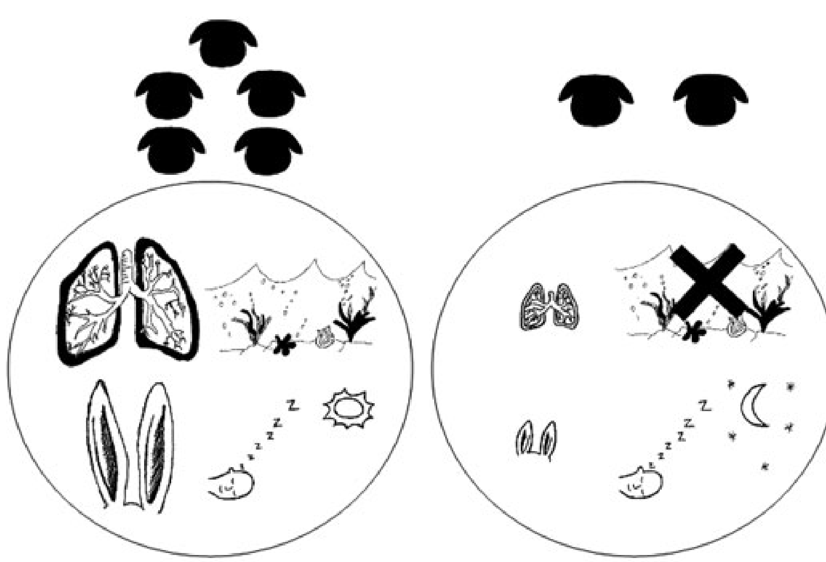
In contrast, the *generative model* (Rehder, 2003a, 2003b; Rehder & Kim, 2006) builds on causal model theory (Sloman, 2005; Waldmann & Holyoak, 1992) by assuming that inter-feature causal relations are represented as probabilistic causal mechanisms (Rehder, 2003a, 2003b; Rehder & Kim, 2006). The central intuition behind the generative model is that objects that are likely to have been produced, or *generated*, by the causal mechanism associated with a particular category should be considered good category members and those unlikely to be generated should be considered poor ones.[2] A central prediction of this account is that the causal coherence of instances should be an important constraint on how they are categorized. Items that maintain the pattern of causes and effects that one expects to be generated by a category's causal mechanism should be likely to be accepted as category members. Incoherent items that violate the causal pattern (have a cause present and effect absent, or vice versa) are likely to be rejected. Notably, the generative model predicts that an additional contribution of causal status will be observed under certain circumstances (e.g., when cause–effect relations are probabilistic, see Rehder & Kim, 2010). This aspect of the generative model is discussed at greater length in Experiment 2.

The current experiments represent the first attempt to test the competing predictions of the dependency and generative predictions in children's categorization. If children's categorization is constrained by causal coherence (and if the causal status effect only emerges for probabilistic causal relations), then the generative model will be supported over the dependency model. However, a finding of a causal status effect that increases with the strength of the causal relations and the absence of causal coherence would constitute support for the dependency model. Finally, neither model predicts an effect of relational centrality and so finding that effect would constitute evidence against both.

### *1.3. Evaluating the effects of causal knowledge on categorization*

The aim of the current study was to test for the presence of causal status, coherence, and relational centrality effects in 5- to 6-year-old children and adults. 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.

Children and adults were taught artificial categories with four features. The typical and atypical features of category members were conveyed through the presentation of two study instances. The ''typical'' instance illustrated the four features possessed by most category members. The ''atypical'' instance had opposite values on each feature dimension, found only in a minority of category members. These different base rates were conveyed by different numbers of silhouettes representing the animal's shape above the typical and atypical items (see Fig. 2 for an example). Two of the four features in each instance 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 (e.g., have long ears and sleep during the day) were causally unrelated (hereafter ''neutral features''). 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



*"Most Rogos……."* *"Some Rogos……"*

Fig. 2. Example of typical and atypical members of the Rogo category.

Table 1
Examples of typical and atypical study instances from the Rogos category (see Appendix A for a list of all study phase categories)

| Typical Study Instance (1111) ''Most ROGOS have …'' | | | | Atypical Study 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 |

cause, effect, and two neutral feature dimensions, respectively. Fig. 2 gives an example of how the causal and noncausal features were illustrated.

After learning a category's typical and atypical features and the causal linkages, participants were tested 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 chose the object that was more likely to be a category member. Table 2 shows the feature values on each of four dimensions for each test pair. An ''x'' indicates that no information about that dimension was provided for that item. For example, ''10xx'' denotes an instance that was described as having the typical cause feature, an atypical effect feature, and was missing features on the third and fourth dimensions. A forced choice categorization test was used rather than ratings of category membership because young children often have difficulty in using such scales reliably (Surber, 1984). The forced-choice test procedure is also a closer match to the testing methods used by Ahn et al. (2000a) and Meunier and Cordier (2009).

The seven test items in Table 2 were designed to allow for the identification of multiple effects of causal knowledge on categorization. First, note that only two 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 whether subjects attended to information about feature typicality (if so, 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. 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 classified on the basis of causal status, coherence, relational centrality, or some combination of all three.

These effects have not previously been assessed in the context of forced-choice judgments. Hence, to present our predictions rigorously, we introduce the following definitions. Let $w_C$, $w_E$, and $w_N$ represent 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 $i$ is present and decreased

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

| | | | | Predicted Values for $choice_k$ (X, Y) [Preferred Alternative] | | | | Experiment 1 (Preference for X) | | Experiment 2 (Preference for X) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Typicality Only | Typicality + Causal Status Effect Only | Typicality + Coherence Effect Only | Typicality + Relational Centrality Effect Only | | | | | |
| Test Item | Choice X | Choice Y | $diff_k$ (X, Y) | $w_C = 1$ $w_E = 1$ $w_N = 1$ $w_H = 0$ | $w_C = 2$ $w_E = 1$ $w_N = 1$ $w_H = 0$ | $w_C = 1$ $w_E = 1$ $w_N = 1$ $w_H = 1$ | $w_C = 2$ $w_E = 2$ $w_N = 1$ $w_H = 0$ | Adults | 5- to 6-Year Olds | Adults (''Tends to'') | Adults (''Sometimes'') | 5- to 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** | .96** | .80** |
| $T_B$ | xx11 | xx00 | $4w_N$ | .98 [X] | .98 [X] | .98 [X] | .98 [X] | .99** | .75** | .94** | .94** | .71** |
| $T_C$ | 10xx | 01xx | $2(w_C - w_E)$ | .50 [=] | .88 [X] | .50 [=] | .50 [=] | .52 | .51 | .71** | .63** | .48 |
| $T_D$ | 10xx | xx10 | $w_C - w_E - w_H$ | .50 [=] | .73 [X] | .27 [Y] | .50 [=] | .30** | .48 | .34** | .38* | .28** |
| $T_E$ | 01xx | xx01 | $-w_C + w_E - w_H$ | .50 [=] | .27 [Y] | .27 [Y] | .50 [=] | .32** | .43 | .37** | .34** | .42 |
| $T_F$ | 11xx | xx11 | $w_C + w_E - 2w_N + w_H$ | .50 [=] | .73 [X] | .73 [X] | .88 [X] | .70** | .68** | .78** | .67** | .57 |
| $T_G$ | 00xx | xx00 | $-w_C - w_E + 2w_N + w_H$ | .50 [=] | .27 [Y] | .73 [X] | .12 [Y] | .62** | .55 | .57 | .54 | .54 |

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

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 consequence, indicates whether $i$ is coherent (+1), incoherent (−1), or neither (0). 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 ($-w_H$). 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 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 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.

Given the definition for an object's degree of category membership in Equation 1, we can predict which of two instances should be preferred when they are paired together in a forced-choice judgment (see Appendix B for the derivation of these probabilities). Table 2 gives example values for the *w* parameters and the predicted pattern of categorization corresponding to four cases: (a) an effect of feature typicality only; (b) typicality + causal status; (c) typicality + coherence; and (d) typicality + relational centrality. The qualitative predictions based on each of the four alternative models are summarized below (see Appendix B for more detailed discussion of quantitative predictions).

### *1.3.1. Typicality effect only*

As noted above, the predicted pattern of categorization based on typicality alone is relatively straightforward. When one member of the test pairs has more typical features than the alternative (as in $T_A$ and $T_B$), then this item should be preferred in test choices. In the remaining pairs ($T_C$–$T_G$), the test alternatives had the same number of typical features so choice should be at chance (indicated by a ''='' in Table 2).

### *1.3.2. Typicality + causal status effect only*

Now consider the case where participants use both typicality and causal status (and assuming the absence of both coherence, $w_H = 0$, and relational centrality effects, $w_E = w_N$). Subjects should still choose alternative X on test items $T_A$ and $T_B$ because it has the more typical features. More interesting is item $T_C$, which involves a choice between 10xx (cause present but effect absent) and 01xx (cause absent but effect present). These instances have the same number of typical features. However, 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 $T_C$. 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. 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. 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}.

*1.3.3. Typicality + coherence effect only*

The predictions for the coherence effect are generated assuming an effect of typicality ($w_i > 0$) and the absence of both causal status ($w_C = w_E$) 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.[3] In test item $T_D$, Y (xx10) should be chosen over X (10xx) because of the incoherence exhibited by 10xx. 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}.

*1.3.4. 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 causal status and coherence. 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. 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 these 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 show how causal status, coherence, and relational centrality effects each imply a unique pattern of performance on the seven test items in Table 2. They also suggest how one may test for the presence of two or more of these effects operating simultaneously. As we show in Appendix B, a logistic regression model applied to all seven choice problems in Table 2 can yield beta weights that represent the importance to classification of the cause feature C ($w_C$), the effect feature E ($w_E$), and the neutral features $N_1$, and $N_2$ ($w_N$), as well as a beta weight representing the importance of the item being coherent ($w_H$). As described earlier, these weights can be interpreted as reflecting the presence of a causal status effect ($w_C > w_E$), a relational centrality effect ($w_E > w_N$), and a coherence effect ($w_H > 0$). In addition, because it simultaneously evaluates multiple test items, logistic regression 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.

## 2. Experiment 1

In Experiment 1, university students and 5- to 6-year old children were taught artificial categories like those in Table 1. Note that in Experiment 1, no information about the strength of the causal link between exemplar features was provided. For example, it was not stated whether the causal link with staying underwater was true for every Rogo that had 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.

### *2.1. Method*

#### *2.1.1. Participants*

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

#### *2.1.2. Materials and procedure*

Four sets of animal categories (labeled Rogos, Waddos, Daxes, and Bliks) were created for the experiment. All categories had the same statistical and causal structure as the Rogos in Table 1 but had different surface features (see Appendix A).

All participants were tested individually in a quiet room and told they were going to learn about different kinds of extraterrestrial animals. Participants were trained and tested with either four categories (adults) or two categories (children).[4] For each category, there was a study phase that was followed immediately by a test phase. In each study phase, participants were shown one typical instance (which had features that were true of ''most'' category members) and one atypical instance (with features that were true of ''some'' category members). The different base rates were reinforced by placing five identical silhouettes of representing the animal's shape above the typical item and two identical silhouettes above the atypical item (as shown in Fig. 2).

The four features of each study instance were described twice by an experimenter. On the first occasion, the experimenter labeled the features of each item and illustrated them by placing 13-cm × 9-cm black-and-white line drawings on a poster board positioned 1 m from the participant (see Fig. 2 for an example). On the second occasion, the experimenter highlighted the relationship between the cause and effect feature in each instance by saying, ''It's because this animal has [causal feature] that it can [effect feature].'' For children, causal relations were further emphasized 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'').

Presentation of the cause feature always preceded the relevant effect. Order of presentation of the two neutral features within each instance was randomized. Typical study instances were always presented before atypical instances. The left-right position of typical and atypical instances on the poster board was counterbalanced across categories. The assignment of features to typical and atypical roles in each category was counterbalanced across participants (i.e., half the participants saw the instances in Appendix A; for the remainder feature, assignment to typical and atypical roles was reversed).

After presentation of the study instances for a given category, the silhouettes were removed and participants were asked to identify the set of features associated with typical items (e.g., Can you show me which things *most* Rogos have?). If this question was answered incorrectly, the description of the study instances (together with base rate information) was repeated.

The test phase for each category commenced immediately after presentation of the study instances, which remained in view during test. On each test trial, a pair of items, each containing two features, was presented. Participants were asked to choose which item was most likely to be a member of the given category (e.g., ''Which animal do you think is more likely to be a Rogo?''). The features of each instance were described verbally by the experimenter and illustrated using line drawings (see Fig. 2 for an example). Categorization choices could be indicated verbally or by pointing to the appropriate item. Participants were told that guessing was acceptable. The two test items designed to check understanding of the base rates of typical and atypical features (items $T_A$ and $T_B$ from Table 2) were always presented first. The remaining five test items from Table 2 were then presented in random order.

## 2.2. Results

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

If test responses were 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 age group was at chance for all five pairs. This shows that the interfeature causal links influence categorization in both age groups. To understand exactly which effects were driving causal categorization, we now consider the performance of each of age group in detail.

### 2.2.1. Adult responses

Examination of the pattern of adult responses in Table 2 indicates the presence of a coherence effect and the absence of causal status and relational centrality effects:

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). In contrast, this pattern is inconsistent with both causal status (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 $T_G$) and relational centrality (they 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 presented in Appendix B was applied to the choice data of each adult subject.[5] The estimated values of $w_C$, $w_E$, $w_N$, and $w_H$ averaged over subjects are presented in Table 3. 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*($w_C$, $w_E$, $w_N$, $w_N$)), causal status ($w_C - w_E$), and relational centrality ($w_E - w_N$).

Table 3 confirms that objects with more typical features were more likely to be judged category members (average of all feature weights = .66), $t(37) = 13.50$, $p < .0001$. More important, Table 3 also 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 >0, $t(37) = 6.10$, $p < .0001$. Instances were judged more likely to be category members when they were coherent and less likely to be category members when they were incoherent. Notably, 33 of the 38 adults had a $w_H$ parameter that was >0, indicating that the coherence effect was not due to a minority of adults showing a large influence of coherence. The absence of a significant difference between a weight on the cause feature ($w_C$ = .66) and

Table 3
Average parameter estimates for adults and children

| | Experiment 1 | | Experiment 2 | | |
|---|---|---|---|---|---|
| | Adults | 5- to 6-Year Olds | Adults ''Tends to'' | Adults ''Sometimes'' | 5- to 6-Year Olds |
| Parameter | | | | | |
| $w_C$ | 0.66 (0.09) | 0.33 (0.05) | 0.93 (0.10) | 0.81 (0.12) | 0.20 (0.08) |
| $w_E$ | 0.70 (0.08) | 0.29 (0.06) | 0.63 (0.08) | 0.51 (0.10) | 0.29 (0.04) |
| $w_N$ | 0.65 (0.07) | 0.20 (0.04) | 0.51 (0.04) | 0.48 (0.03) | 0.21 (0.04) |
| $w_H$ | 0.85 (0.14) | 0.22 (0.08) | 0.74 (0.12) | 0.59 (0.14) | 0.31 (0.09) |
| Effect | | | | | |
| Typicality | 0.66*** (0.05) | 0.26*** (0.03) | 0.64*** (0.05) | 0.57*** (0.05) | 0.22*** (0.03) |
| Causal status | −0.03 (0.10) | 0.04 (0.07) | 0.30*** (0.08) | 0.31* (0.15) | –0.09 (0.05) |
| Coherence | 0.85*** (0.14) | 0.22** (0.08) | 0.74*** (0.12) | 0.59*** (0.14) | 0.31*** (0.09) |
| Relational centrality | 0.05 (0.10) | 0.08 (0.07) | 0.12 (0.09) | 0.03 (0.10) | 0.08 (0.05) |

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

the effect feature ($w_E$ = .70), $t < 1$, reflects the absence of a causal status effect. Finally, the absence of a significant difference between $w_E$ (.70) and $w_N$ (.65), $t < 1$ 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 positive difference between $w_C$ and $w_E$ (that drives the model to prefer X) is absent.

### *2.2.2. Children's responses*

Informal analysis of children's responses to the test items $T_C$–$T_G$ was less straightforward. On one hand, children were like 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, although 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 adults' preferences), these choice probabilities reached significance for 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 3.

Table 3 confirms that children, like the adults, judged that objects with more typical features were more likely to be category members (typicality effect = .26), $t(39) = 7.72$, $p < .0001$. In addition, it shows the presence of a coherence effect, indicated by a value for parameter $w_H$ (.22) that was significantly >0, $t(39) = 2.71$, $p < .05$. 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. Twenty-nine of 40 children had a $w_H$ parameter >0, indicating that the coherence effect was not due to a minority of children showing a large influence of coherence. Moreover, the absence of a difference between the weight on the cause ($w_C$ = .33) and the effect features ($w_E$ = .29), $t < 1$, reflects the absence of a causal status effect. These values for $w_C$ and $w_E$ reflect children's agnostic performance on test item $T_C$. Finally, the absence of a significant difference between $w_E$ (.29) and $w_N$ (.20) indicates the absence of a relational centrality effect, $t(39) = 1.28$, $p > .20$.

### *2.2.3. 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 $w_N$) when compared with 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 age groups, both *t*s < 1.

To address the question of whether the group means presented in Table 2 reflect the responses of most subjects within an age group or of subsets with different response profiles, a cluster analysis of the responses of each age group was performed. Three clusters of adults were identified. One included a large majority (33 of 38) who responded in a manner reflected by the group means. The other two subgroups represented the idiosyncratic response patterns of two and three individuals, respectively. For children, five clusters were identified, including two representing a substantial number of children (19 and 13). The primary difference was that children in the first subgroup exhibited smaller typicality and coherence effects than those in the first, a result we interpret as more confident responding by the second subgroup. The other three subgroups of three, three, and two individuals reflected idiosyncratic responding.

### 2.3. Discussion

Experiment 1 examined how causal knowledge affects categorization in 5- to 6-year old children and adults. A major finding was that although children were somewhat less consistent in their response patterns than adults, both age groups 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). A robust effect of coherence was reflected in both the mean pattern of responding shown by each age group and in the individual response profiles of both children and adults. Notably, although there was somewhat more diversity in individual patterns of test responding among children, a majority of participants in both age groups showed evidence of an effect of causal coherence. The presence of a coherence effect replicates previous demonstrations in adults and extends those findings to use of a forced-choice procedure. More important, Experiment 1 represents the first evidence that 5- to 6-year old children also exhibit a coherence effect in causal categorization.

Another key finding was the lack of a causal status effect in both adults and children. This null result 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 coherence or relational centrality. When the effects of coherence and centrality are controlled, the effect of causal status often disappears or is relatively weak (e.g., Rehder, 2003b; Rehder & Kim, 2006). Moreover, 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, 2010). Hence, in Experiment 2, we induce conditions more favorable to the causal status effect in adults and then test whether it also occurs in children.

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

## 3. Experiment 2

A key result from Experiment 1 was the lack of a causal status effect in children's categorization. On one hand, this result supports our alternative interpretation of the findings of Ahn et al. (2000a) and Meunier and Cordier (2009); namely, that the causal status effect in children observed in those studies may actually be due to coherence. However, the lack of a causal status effect in adults raises the possibility that its absence in Experiment 1 may have been due to other factors. Experiment 1 differs from previous studies of causal status in many ways (having only one causal link per category, different materials, etc.) and one of these factors may have been responsible for the null finding. 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 did not adults exhibit a causal status effect in Experiment 1? Recently, Rehder and Kim (2010) have demonstrated that causal status is influenced by a number of variables. One 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''). Rehder and Kim (2010) 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. Moreover, when all feature relations were probabilistic but varied in strength, a stronger causal status effect was found in the context of weaker causal relations (e.g., when the propensity of the cause to produce the effect was 0.60 rather than 0.90).

We may have failed to find a causal status effect 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 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. This explains why classifiers in Experiment 1 were at chance on test item $T_C$: They had no basis for preferring the alternative with the cause feature (X: 10xx) over the one that has the effect (Y: 01xx). But if classifiers believed instead that only *some* Rogos with large lungs ended up being able to stay underwater for a long time, then the cause would be viewed as more prevalent than the effect and thus more diagnostic of category membership. Classifiers might then prefer alternative X in test item $T_C$ (see Note 3 for elaboration). That is, a probabilistic causal relation would result in a causal status effect.

Experiment 2, therefore, aimed 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- to 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- to 6-year-olds would understand the probabilistic language used to describe the causal relations in this experiment. In order to check that the use of the different terms to describe probabilistic relations did not, in itself, lead to different patterns of responding in children and adults, we tested an additional group of adults using the ''sometimes'' wording.

In addition to providing a stronger test of the causal status effect in children, Experiment 2 again examined the independent contribution of coherence and relational centrality to causal categorization in each age group.

## 3.1. Method

### 3.1.1. Participants

The participants were 51 kindergarten and first-grade children ($M_{\mathrm{AGE}}$ = 5 years 9 months; Range: 5 years 6 months to 6 years 6 months) from a private elementary school in a middle-class metropolitan area of Sydney, Australia. Forty-five Psychology undergraduates participated for course credit. They were tested using the ''tends to'' wording to describe probabilistic relations. A further 32 undergraduate and graduate students participated for a small payment. They were tested using the ''sometimes'' wording.

### 3.1.2. Materials and procedure

These were identical to Experiment 1 except for the way that cause and effect features were presented. For one group of adults, the relation between these features was always instantiated by saying, ''Because they have [cause feature] Rogos tend to [effect feature].'' For children and the other group of adults, the same relations were instantiated by saying, ''Because they have [cause feature] Rogos sometimes [effect feature].''

## 3.2. Results

Following Experiment 1, we excluded participants (5 children, 0 adults) whose average responses to test items TA and TB 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 shows that categorization in both age groups was affected by causal links between features even when these links were presented as probabilistic.

### 3.2.1. 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 in Experiment 2. This was true regardless of whether probabilistic causal relations were described using the ''tends to'' or ''sometimes'' wording. 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 3. As before, in addition to parameter $w_H$ (which provides information about the coherence effect), Table 3 shows measures corresponding to typicality (*average*($w_C$, $w_E$, $w_N$, $w_N$)), causal status ($w_C - w_E$), and relational centrality ($w_E - w_N$). 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 when the phrase ''tends to'' was used to describe probabilistic causal relations, $t(44) = 3.94$, $p < .001$. A similar result was found for the ''sometimes'' wording, $w_C - w_E = .31$, $t(31) = 2.08$, $p < .05$. Hence, Experiment 2 succeeded in inducing a causal status effect in adults by describing the link between the cause and effect as probabilistic.

Table 3 also shows that Experiment 2 replicated the large coherence effect in adults (with both ''tends to'' and ''sometimes'' wordings), as indicated by weights on parameter $w_H$ (0.74 and 0.59) that were both significantly >0, $t(44) = 6.16$, $p < .0001$ and $t(31) = 4.14$, $p < .001$. A majority of adults (38 of 45 who received the ''tends to'' wording, and 26 of 32 who received the ''sometimes'' wording) had a $w_H$ parameter >0. This shows that, like Experiment 1, the coherence effect was not merely due to a minority of subjects showing a large influence of coherence. The relational centrality effect, in contrast, did not reach significance in either condition (''tends to'': $t(44) = 1.39$, $p = .17$; ''sometimes'': $t < 1$).

### 3.2.2. Children's responses

Table 3 shows the results of the logistic regression analysis of children's responses. Whereas adult responses in Experiment 2 differed from 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 3 shows a coherence effect as indicated by a value for parameter $w_H$ (.31) that was significantly >0, $t(45) = 3.66$, $p < .001$. This reflects the fact that a majority of children (36 of 46) had a $w_H$ parameter that was above zero. Second, the absence of a positive difference between the 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 $w_N$ did not reach significance, $t(45) = 1.59$, $p = .12$, indicating an absence of a relational centrality effect.

### 3.2.3. Comparison of adults and children

We again compared the magnitude of the effects in the two age groups. First, the average weight on the four features ($w_C$, $w_E$, $w_N$, and $w_N$) indicated a larger typicality effect in adults than children (.61 vs. .22), $t(122) = 7.81$, $p < .0001$. Second, although both groups exhibited sensitivity to coherence, adults exhibited a larger coherence effect than children (.68 vs. .31), $t(122) = 2.66$, $p < .01$. Third, adults exhibited a larger causal status effect (.31) than children (.31 vs. −.09), $t(122) = 3.71$, $p < .001$. Finally, the size of the relational centrality effect did not differ across groups, $t < 1$.

As in Experiment 1, we conducted cluster analyses of the responses of the adults and children to determine whether the group means reflect the responses of most subjects. For this analysis, the two groups of adults were combined. Five adult clusters were identified. One included the large majority of subjects (62 of 77) whose pattern of test responses generally reflected the group means (with item choices consistent with effects of both coherence and causal status). Another cluster of seven individuals exhibited an especially large causal status effect. The other three subgroups (consisting of five, two, and one individuals) were small and reflected idiosyncratic response patterns. There were two clusters of children; the first included the majority of children (35 of 46) who responded in accordance with the group means (i.e., showing a coherence effect). The second cluster of 11 children exhibited an especially large coherence effect and a negative causal status effect in which the effect was weighed more than the cause. As was the case in Experiment 1, most subjects in the current experiment responded in a manner consistent with their respective group means.

### *3.3. 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- to 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. This finding is consistent with previous results showing that a causal status effect occurs for probabilistic causal relationships but not deterministic ones (Rehder & Kim, 2010). In contrast, 5- to 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 finding.

In Experiment 2, both adults and children exhibited significant coherence effects. This replicates previous findings of coherence effects in adults with probabilistic causal links (Rehder & Kim, 2010). More important, 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 now consider at greater length our findings regarding causal status, coherence, and relational centrality effects.

## 4. General discussion

These studies examined the mechanisms that drive causal categorization in children and adults. The first step was to test for the presence of causal coherence, causal status, and relational centrality effects in each age group. Like previous studies (e.g., Ahn et al., 2000a; Meunier & Cordier, 2009), we used a forced-choice classification procedure suitable for young children. Unlike those studies however, our design and logistic regression analysis allowed us to evaluate the independent contribution of each of these effects.

In adults, the coherence between cause and effect features always influenced categorization. An additional effect of causal status was found only when causal relations were presented probabilistic. These results replicate and extend previous work showing both causal status and coherence effects in adult categorization (Ahn et al., 2002; Marsh & Ahn, 2006; Rehder & Hastie, 2001; Rehder, 2003a, 2003b; Rehder & Kim, 2006, 2010) and induction (e.g., Rehder & Hastie, 2004).

The most important novel finding was that categorization by 5- and 6-year-olds was also strongly influenced by causal coherence. Like adults, children judged that instances in which cause and effect features were both present, or both absent, were likely to be category members, whereas items in which the cause was present and the effect absent (or vice versa) were not. Notably, children showed no evidence of an independent causal status effect regardless of whether causal relations were described as deterministic or probabilistic.

These findings represent an important advance on previous work examining causal categorization in children. Although past studies have shown that young children use causal knowledge in categorization (cf., Ahn et al., 2000a; Hayes et al., 2003) and induction (Hayes & Thompson, 2007), they have not shown exactly *how* that knowledge is used and have not tested explicit models of causal categorization. The current studies extend this work by assessing the contribution to children's categorization of multiple kinds of causal effects (coherence, causal status, and centrality). This in turn allowed us to compare the predictions of the generative and dependency models of causal categorization. Our findings are the first to show a clear effect of causal coherence in children and to suggest that a sensitivity to causal status may emerge later in development. In these respects, children's causal categorization conforms more closely to the predictions of the generative model than the dependency model (see below for elaboration of this point).

These results suggest that a reinterpretation of previous developmental findings is in order. Previous work showed that children think that an object missing an effect is more likely to be a category member than one missing a cause (Ahn et al., 2000a; Hayes et al., 2003; Meunier & Cordier, 2009; Opfer & Bulloch, 2007). Ahn et al. and Meunier and Cordier interpreted this as evidence for the early emergence of causal status in children's categorization. The current work suggests, however, that these previous findings are better explained by assuming the early development of sensitivity to causal coherence.

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 (Fig. 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). In addition, it is important to note that Ahn et al. tested children who 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.

A final notable result was the absence of a relational centrality effect in either adults or children. 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, 2002). Causally related features are also weighed more heavily than neutral ones when they are involved in more complex networks involving at least four features (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). To provide a stronger test of relational centrality effects in children, as well as to extend the current results regarding the development of coherence and causal status, future work needs to examine these effects in more complex causal structures (e.g., a common cause scenario like that shown in Fig. 1).

### *4.1. Evaluating theoretical accounts of causal classification*

The current findings have important implications for models of causal categorization. According to the generative model, people believe that category members are generated by the category's causal mechanisms. As a consequence, novel items will be seen as likely category members to the extent they embody the statistical properties of those category members. This view explains our two key empirical findings. First, because causally related features should be correlated, items are more likely to be category members to the extent that cause and effect features are either both present or both absent. Thus, the generative model accounts for the coherence effects exhibited by adults and children in Experiments 1 and 2. Second, the relative prevalence of a cause and effect feature depends on the strength of their causal relation: An effect will be as prevalent as its cause when the causal relation is deterministic and less prevalent when it is probabilistic (assuming the absence of alternative causes). Thus, the generative model explains the presence of a causal status effect in adults when causal relations were described as probabilistic (Experiment 2) and its absence when they were described as deterministic (Experiment 1).

In contrast, the dependency model (Sloman et al., 1998) characterizes the knowledge that classifiers have about category features as a network of dependency relations and assumes that features that are at lower levels in a causal chain will carry more weight in categorization. Because a stronger link means that the effect depends more heavily on the cause, the dependency model predicts that the causal status effect should be stronger for stronger causal links. We found the opposite pattern. A causal status effect was only observed in adults when the causal links were described as probabilistic (in Experiment 2). Moreover, the dependency model fails to predict an independent effect of coherence.

### *4.2. The generative model, essentialism, and determinism*

The current studies reveal both continuities as well as a possible discontinuity in the development of causal categorization. Our results show that an ability to understand causal generative mechanisms emerges at a relatively early age and continues to influence categorization during adulthood. In contrast, causal status only seems to constrain categorization in older individuals, and then only under certain circumstances.

Previous work has highlighted the importance of children's understanding of the role of ''mechanism'' in interpreting causal events. When shown a novel event, young children will actively search for a cause and will often evaluate rival causes based on the presence of a plausible mechanism linking the cause to the effect (Bullock, Gelman, & Baillargeon, 1982; Schlottmann, 1999). Bullock et al. (1982), for example, presented children with a novel ''effect'' (a Jack in the box popping up) and asked them to choose between two potential causes (e.g., a light being turned on or a ball rolling toward the target) that had equivalent levels of covariation with the effect. Children of 4 years and older chose the cause that had the most plausible physical mechanism linking the cause and the effect (e.g., the light was chosen when the ball did not make physical contact with the Jack in the Box). The current findings extend such work by highlighting another important aspect of children's beliefs about causal mechanism. When children learn about a specific mechanism that links cause and effect features within a category (as occurred in the study phase of each experiment), they expect this mechanism to operate in a coherent manner (i.e., that the cause and effect will covary) in other category members.

It is also 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 the observable features of category members are caused by a less obvious core feature, common to all category members. Children as young as 3 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 the notion of precocious essentialism in that 5-year-olds expected members of the same category to exhibit features that reflect nonobvious causal factors. But whereas essentialism focuses on a belief in a shared *underlying cause* for category membership, the current results highlight children's sensitivity to shared *underlying causal mechanisms*. We suggest that classifiers interpret the presence of a cause feature along with its effect as evidence for the presence of one of the category's causal mechanisms, which in turn is taken as evidence for the category itself. On this account, adults and children in the present study exhibited coherence effects because, for example, the presence of ''big lungs'' along with ''staying underwater for long periods'' reflected the operation of one of Rogos' normal causal processes. That causal process in turn made them more confident that the animal was a Rogo.

An interesting developmental finding was that the strength of causal relations between features had a marked effect on adult categorization (leading to an additional causal status effect) but little effect on the way that 5- to 6-year olds categorized. It is difficult to say conclusively whether this difference arose because of age changes in understanding of the probabilistic language used in Experiment 2 or because of a more general bias among children to interpret causal links as deterministic rather than probabilistic (cf., Shultz & Somerville, 2006). Previous work (e.g., Hoffner et al., 1990) has shown that 5- to 6-year-olds understand the different implications of adverbs like ''*probably*'' and ''*definitely*,'' suggesting that they are capable of grasping verbally described probabilistic relations. However, in the absence of an independent test of children's understanding of the ''sometimes'' wording, we need to be cautious about interpreting the lack of a causal status effect among children.

One way of making progress on this issue may be to use a manipulation of probabilistic and deterministic relations that does not depend on linguistic cues. Griffiths, Sobel, Tenenbaum, and Gopnik (2011) found that when children as young as 4 years old have an opportunity to observe patterns of covariation between events over successive trials, they can distinguish between probabilistic and deterministic causal relations. This method could be adapted so that during study, children observe a series of category exemplars where cause and effect features show either perfect covariation or probabilistic covariation across trials. Once a reliable technique for teaching children probabilistic causal links is identified, it is possible that they will also exhibit a causal status effect.

## 5. Conclusions

This work shows that the coherence between cause and effect features is a major constraint on categorization in *both* children and adults. When the effects of causal coherence and centrality were controlled, an effect of causal status was only found for adult categorization involving probabilistic causal links. These results challenge previous interpretations of causal categorization in young children (e.g., Ahn et al., 2000a; Meunier & Cordier, 2009) that have assumed that causal status effects are driven by dependency relations. The findings show 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.

## Acknowledgments

This work was supported by Australian Research Council Discovery grant DP1093197 to the first author. 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.

## Notes

1. Note that the common cause structure shown at the top of Fig. 1 may also be interpreted as implying a correlation between the effect features (i.e., if a Taliboo had thick bones, you would expect it to also have large eyes). Under this interpretation the test item on the right-hand side of the figure would still be more likely to be accepted as a category member because it violates two expected correlations, whereas the left-hand test item violates three.
2. In this context, a ''mechanism'' is some feature or set of features within a category that is believed to have causal force or causal necessity.

3. Note that our definition of *coherence* may be viewed as subsuming two conceptually distinct types of coherence that could be used to distinguish between the alternatives in test item $T_C$. The X alternative (10xx) is incoherent because the cause is present and the effect absent. This *violation of sufficiency* might be viewed by people as being maximized when the causal link is deterministic (the cause always produces the effect). In contrast, the Y alternative (01xx) is incoherent because the cause is absent and the effect present. This *violation of necessity* might be viewed by people as being maximized when the effect is believed to have no alternative causes. Thus, a classifier might prefer alternative X in test item $T_C$ when he or she believes that the causal link is not deterministic and there are no alternative causes and prefer Y when he or she believes the link is deterministic and there are alternative causes. Our definition of coherence embodies both sufficiency and necessity both for simplicity and because it provides a measure of the importance of interactions among features that is orthogonal to the importance of features considered independently. Future studies using different sorts of test items could readily distinguish between these two types of (in)coherence.
4. The decision to present children with only two categories was made after pilot testing showed that they became fatigued when more than two study/test sequences were presented.
5. 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 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 that can result in success or failure. Define $\pi$ to be the true probability of choosing X. To estimate $p(\pi|x)$, the probability distribution of $\pi$ given $x$ successes, Bayes' theorem was applied,

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

where $p(\pi)$ is the prior distribution of $\pi$ and $p(x|\pi)$ is the binomial distribution. A noninformative prior distribution of $\pi$ was assumed (Box & Tiao, 1973), that is, a prior that makes the weakest assumptions: A beta distribution with parameters $\alpha = .5$ and $\beta = .5$. $\alpha$ may be interpreted as representing the number of prior successes and $\beta$ as the number of prior failures. The resulting posterior is a beta distribution with parameter $\alpha = x + .5$ and $\beta = (N - x) + .5$, 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:

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

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.

## 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. (2001). The causal status effect in categorization: An overview. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 23–65). San Diego, CA: Academic Press.

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., Marsh, J. K., Luhmann, C. C., & Lee, K. (2002). Effect of theory based correlations on typicality judgments. *Memory & Cognition*, *30*, 107–118.

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

Bullock, M., Gelman, R., & Baillargeon, R., (1982). The development of causal reasoning. In W. J. Friedman (Ed.) *The developmental psychology of time* (pp. 209–254). New York: Academic Press.

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.

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 (Eds.), *Emerging themes in cognitive development* (pp. 3–32). New York: Springer

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.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*, 1407–1455.

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.

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.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, *23*, 250–269.

Marsh, J., & Ahn, W. (2006). The role of causal status versus inter-feature links in feature weighting. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 561–566). Mahwah, NJ: Erlbaum.

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

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberchien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23–45). Chichester, UK: Ellis Horwood.

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

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

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. doi:10.1016/S0010-0277(03)00167-7.

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

Rehder, B., & Kim, S. W. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1171–1206.

Schlottmann, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology*, *35*, 303–317.

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: Oxford University Press.

Sloman, S., & 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.

Surber, C. F. (1984). Issues in using quantitative rating scales in developmental research. *Psychological Bulletin*, *95*, 226–246. doi: 10.1037//0033-2909.95.2.226.

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.

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.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*, 222–236. doi: 10.107/0096-3445.121.2.222.

## Appendix A

Examples of additional causal and noncausal features used in experimental categories

| | Typical Study Item (1111) | | | | Atypical Study Item (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 |

*Note*. Dimension C = cause feature, Dimension E = effect feature, Dimensions $N_1$ and $N_2$ = noncausal features.

## Appendix B: Derivation of regression analysis predictions

Using the definitions of parameter weights specified in the text and following Eq. 1, the probability of choosing X over Y can be described as a function of the difference between $ev_k$ (X) and $ev_k$ (Y),

$$
\begin{aligned}
diff_k(X, Y) &= ev_k(X) - ev_k(Y) \\
&= \left(w_C f_{X,C} + w_E f_{X,E} + w_N f_{X,N_1} + w_N f_{X,N_2} + w_H h_X\right) \\
&\quad - \left(w_C f_{Y,C} + w_E f_{Y,E} + w_N f_{Y,N_1} + w_N f_{Y,N_2} + w_H h_Y\right) \\
&= w_C\left(f_{X,C} - f_{Y,C}\right) + w_E\left(f_{X,E} - f_{Y,E}\right) + w_N\left(f_{X,N_1} - f_{Y,N_1}\right) \\
&\quad + w_N\left(f_{X,N_2} - f_{Y,N_2}\right) + w_H(h_X - h_Y) \\
&= w_C m_{XY,C} + w_E m_{XY,E} + w_N m_{XY,N_1} + w_N m_{XY,N_2} + w_N m_{XY,h}
\end{aligned}
\tag{B1}
$$

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 do not have it. Table 2 presents how $diff_k$ (X, Y) simplifies for each of the seven test items. For example, because $ev_k$ (10xx) = $w_C - w_E - w_H$, and $ev_k$ (01xx) = $-w_C + w_E - w_H$, then $diff_k$ (10xx, 01xx) = $2w_C - 2w_E$ for test pair $T_C$.

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

$$choice_k(X, Y) = \frac{1}{(1 + e^{-diff_k(X,Y)})}. \tag{B2}$$

Equation B2 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.

*Typicality effect only*. Table 2 presents a set of weights that exemplifies the parameter constraints for an effect of typicality with no causal status, coherence, or relational centrality, namely, $w_C = 1$, $w_E = 1$, $w_N = 1$, and $w_H = 0$. For example, for test item $T_A$, $ev_k$ (11xx) = 2 and $ev_k$ (00xx) = −2, which results in $diff_k = 4$. Applying Equation B2, this value of $diff_k$ yields a choice probability $choice_k = .98$; that is, X should be strongly preferred over Y. In addition, there should be no preference between objects with the same number of typical features. For example, for test item $T_C$, $ev_k$ (10xx) = $ev_k$ (01xx) = 0, $diff_k = 0$, and thus $choice_k = .50$. Note that these predictions and those presented below hold not just for the example parameter values shown in Table 2 but for any parameters that satisfy the constraints.

*Typicality + causal status effect only*. Table 2 presents the values of $diff_k$ 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 are generated in a similar manner. For $T_D$ and $T_F$, $diff_k = 1$ and $choice_k = .73$, so X should be chosen over Y for both items. Conversely, for test items $T_E$ and $T_G$, the Y alternatives should be preferred because the Xs are missing the cause feature, resulting in $diff_k = -1$ and $choice_k = .27$.

*Typicality + coherence effect only*. Table 2 shows the *w* weights that reflect a presence of typicality and coherence effects but the absence of causal status and relational centrality, namely $w_C = 1$, $w_E = 1$, $w_N = 1$, and $w_H = 1$. For test item $T_C$, both X (10xx) and Y (01xx) a choice probability of .50 is indicated by $ev_k$(10xx) = $ev_k$(01xx) = −1 and $diff_k = 0$. In test item $T_D$, the choice of Y over X is indicated by the values $ev_k$(10xx) = −1, $ev_k$(xx10) = 0, and $diff_k = -1$. Y should be chosen over X in test item $T_E$ for the same reason.

*Typicality + relational centrality effect only*. Table 2 shows example parameter values for an effect of typicality and relational centrality in the absence of causal status and coherence effects; namely $w_C = 2$, $w_E = 2$, $w_N = 1$, and $w_H = 0$. For $T_D$ and $T_E$, a choice probability of 0.5 is indicated by $ev_k$ (10xx) = $ev_k$ (01xx) = −1 and $diff_k = 0$. For $T_F$, $ev_k$ (11xx) = 4 and $ev_k$ (xx11) = 2, and thus $diff_k = 2$ and $choice_k = .88$; that is, X should be chosen over Y. For $T_G$, $ev_k$ (00xx) = −4 and $ev_k$ (xx00) = 2, and thus $diff_k = -2$ and $choice_k = .12$; that is, Y should be chosen over X.