Demystifying Theory-based Categorization

Woo-kyoung Ahn

Christian C. Luhmann

Vanderbilt University

To appear in *Building object categories in developmental time*

When cognitive psychologists first started studying concept learning, they relied on artificial stimulus materials that were not meaningful to participants in their studies. The idea was that if we want to study how people acquire novel concepts, we should use completely novel categories, therefore controlling for the influence of people's already possessed concepts. In the last 20 years or so, however, many researchers have argued that this practice misses one of the most important components of concept learning processes. Concepts are not represented in isolation, but are instead linked to and defined in relation to other concepts. In order to understand the concept of "shoes", for instance, we need to understand the concept of legs and walking. Therefore, it is no surprise that people have a strong tendency to apply their existing knowledge when learning a new concept because that new concept must be embedded in a complex network of existing knowledge. This emphasis on the influence of existing knowledge on concept learning has been called a theory-based approach to concept learning. The name theory-based is derived from the idea that our existing knowledge is represented like scientific theories where concepts are causally related to each other and there are explanations underlying what we can directly observe. Unfortunately, the mechanism underlying theory-based categorization had not been explicitly articulated.

Recently, various attempts have been made to formalize background knowledge and the different ways in which this knowledge can influence concept learning (Heit, 1998; Rehder, 1999). The purpose of this chapter is to describe one such attempt, called the causal status hypothesis (Ahn, 1998), and to argue that this mechanism can account for numerous findings in children's conceptual representations that have been grossly described under the blanket of children's theory use. We will first explain the causal status hypothesis, followed by illustrations of how this mechanism can provide parsimonious accounts for many important theory-based effects found in children's concept learning. Finally, we will describe a series of recent studies showing that this mechanism is a primary process in adult categorization, opening up the possibility that it might be developmentally privileged as well.

Causal Status Hypothesis: General Introduction

Concepts are connected to each other in many complex ways that resemble theories. For example, our concept of "boomerang" is connected with other concepts such as "throwing," "air," "speed," and so on, all of which are intricately connected as in a scientific theory. Furthermore, features within a concept exist in a rich structure, rather than as a set of independent features (Carey, 1985; Gelman & Kalish, 1993; Murphy & Medin, 1985; Wellman, 1990). In particular, features in concepts tend to be *causally* related. For instance, Kim and Ahn (in press) found that more than 76% of symptom relations that clinicians drew for mental disorders could be classified as either causal or as implying causality. Ahn, Marsh, Luhmann, and Lee (2002) found that 58% of feature relations that lay-people recognized in sample natural kinds and artifacts were causal. As shown in Figure 1, for instance, the most frequent label undergraduate participants in Ahn, Marsh, et al.'s study (2002) provided for relations among features in the category furniture was "causes" (e.g., having cushions causes furniture to be comfortable). Given that features in a concept are causally related to each other, the causal status hypothesis states that people regard cause features as more important and essential than effect features in their conceptual representations. (See also Gelman & Hirschfeld, 1999; Gelman & Kalish, 1993; Kahneman & Miller, 1986 for similar proposals.)

Insert Figure 1 about here.

Ahn, Kim, Lassaline, and Dennis (2000) report a number of tests of the causal status hypothesis. Participants in their Experiment 1 learned three characteristic features of a novel category (e.g., animals called "roobans" tend to eat fruits, have sticky feet, and build nests in trees). While the Control group received no further information about the target category, the Causal group was told that one feature tends to cause a second feature, which in turn tends to cause a third feature (e.g., eating fruits tends to cause roobans to have sticky feet, and having sticky feet tends to allow roobans to build nests in trees). Finally, all participants rated the membership likelihood of three transfer items, each of which had two features characteristic of the target category and one non-characteristic feature (e.g., an animal that likes to eat worms, has feet that are sticky, and builds nests in trees). For the Control group, likelihood ratings remained constant regardless of which feature the exemplar animal was missing. In the Causal condition, however, when an exemplar was missing the target category's fundamental cause in the causal chain, the mean likelihood of being a target category member was lower than when an object was missing its intermediate cause in the causal chain, which in turn was lower than when an object was missing its terminal effect. That is, the deeper a feature was in a causal chain, the more central it was in a categorization judgment.

In a further test of the causal status hypothesis, Ahn, Kim, et al.'s (2000) Experiment 3 examined whether people prefer to free-sort objects based on cause features rather than effect features. Participants received a triad of objects, consisting of a target (e.g., Jane who is depressed because she has low self-esteem) and two options (e.g., Susan who is depressed because she has been drinking; Barbara who is defensive because she has low self-esteem), and were asked to choose an option that they would like to categorize with the target. As in the above examples, one option (Susan) had the same effect as the target but a differing cause (Matching-Effect), and the other option (Barbara) had a matching cause but a differing effect (Matching-Cause). If the causal status effect occurs, matching on a cause feature will be considered more important than matching on an effect feature and consequently, participants would prefer the Matching-Cause case to the Matching-Effect case. In order to ensure that preference for the Matching-Cause case was not due to any a priori salience of features that were chosen as causes, a control condition was employed in which all features and tasks were identical to those used in the experimental condition, except that the causal relations among features were explicitly denied (e.g., Jane is depressed. Jane has low self-esteem, which is NOT the reason why she is depressed.). Participants in this Non-Causal Condition showed no preference for either of the options. However, when causal relations were specified, 73.6% of responses were to prefer the Matching-Cause.

Why the Causal Status Effect?

Why should a cause feature be more central than its effect? Does the causal status effect stem from a rational basis so fundamental that we should expect it even from young children?

It would not be an exaggeration to assert that the most critical cognitive capacity an adaptive system should be equipped with is an ability to predict future events. The more one can infer future circumstances, the more one can be prepared for possible environments, resulting in a greater likelihood to survive. Indeed, one of the reasons why we possess concepts is so that we can infer or predict non-obvious properties (e. g., “is dangerous") based on category membership information (e.g., wolf). Thus, a "good" category is considered to be the one that allows rich inductive inferences (e.g., Anderson, 1990).

In particular, the more that underlying causes are revealed, the more inductive power the concept seems to gain. For instance, discovering a cause of a symptom such as nausea (e.g., Is it caused by bacteria or pregnancy?) allows doctors to determine the proper course of treatment, and also to make a better prognosis of the condition (e.g., will it lead to a fever, or to a new baby?). In contrast, merely learning the effect of the symptom (e.g., nausea usually causes a person to throw up) does not necessarily help us come up with a treatment plan. Similarly, understanding the motive of a person's nice behavior (e.g., does he want a promotion or is he genuinely nice?) would allow us to predict many more behaviors of the person than would discovering the consequence of the person's nice behavior (e.g., people were impressed).

Content-based versus Structure-based Approach

By definition, the causal status hypothesis states that the determinant of feature centrality is not the specific content of features but rather the causal role that a feature plays. That is, the critical determinant is the position a feature takes in a conceptual structure.

This structure-based approach contrasts with what we have termed the content-based approach (Ahn & Kim, 2000), in which the focus lies on which specific feature is central in which type of concept. One example of this approach is a prominent debate in the developmental literature about whether young children categorize objects based on perceptually salient dimensions, such as overall shape, or, alternatively, based on what is known as “conceptual” dimensions, such as functions, intentions, and non-obvious inside features (e.g., Gelman & Koenig, in press). For instance, Landau, Smith, and Jones (1988) presented two- and three-year-old children with a small, blue, wooden inverted U-shaped object, and told the children that the object was a "dax." When asked to select other objects that were also daxes, children preferred objects with the same shape to those with the same size or material (known henceforth as a shape bias). For other types of categories, such as biological kinds, inside matters are shown to be more central than overall shape (e.g., Gelman, 1988). In general, one camp argues that early in conceptual development children form concepts that are primarily based on perceptual features, whereas the other camp argues that even young children’s concepts include more abstract information about causation, intentions, and other properties that are not directly observable. Thus, there are two sets of features differing in content (perceptual and conceptual ones) that are seen as core to children’s concepts and it is this distinction in content that is the focus of debate.

Unlike this traditional content-based approach, we advocate the structural approach. Part of its appeal to us derives from our training background in Cognitive Psychology. As John Macnamara (1998) put it, cognitive psychology is “more concerned with the apparatus of the mind than with the particular uses to which the apparatus is put – just as an anatomist is concerned about the skeletal and muscular structures of the body and how they function together, rather than about the particular purposes to which they are put, such as walking to work or playing tennis” (p. 4). That is, we are more concerned with how concepts are structured and how mind works over various components of concepts, rather than what concepts actually contain as their content matter. In this chapter, we will show that various developmental findings can in fact be explained in terms of the structural approach (the causal status hypothesis, in particular). Furthermore, we will demonstrate that the structural approach can provide a more parsimonious account for the developmental debate pertaining to the bases of children's conceptual representations.

Account of Developmental Findings in Terms of Causal Status Hypothesis

In essence, the hypothesis we entertain in this chapter is that the causal status effect is a processing bias that takes place in all domains across (almost) all developmental stages. How can this view account for variances, such as developmental differences or domain differences? Although the process might be uniform, the input to the system might differ, resulting in different outcomes.

For instance, two different experts might weigh features differently because of differences in their causal background knowledge rather than differences in their processing bias. Consider Figure 2 showing the two radically different theories that expert clinical psychologists in Kim and Ahn's study (in press) drew about schizotypal personality disorder. The symptom "Excessive social anxiety" is a cause of all other symptoms for the first expert, but the same symptom is an effect of other symptoms in the second expert. While clinicians' theories were idiosyncratic, Kim and Ahn found that within a given clinician's theory the causal status effect held. Therefore, the apparently different outcomes result from differences in the input (i.e., causal background knowledge) operating on the same processing mechanism. If a person does not know any causal relations among features one has no way of weighting features based on their causal status, and would rely on other factors (e.g., perceptual saliency, base rates, etc.).

This may be the case with novices or with young children and domains in which they are lacking extensive causal knowledge (e.g. astrology).

Insert Figure 2 about here.

Natural Kinds Versus Artifacts

Previous studies (e.g., Barton & Komatsu, 1989; Gelman, 1988; Keil, 1989; Rips, 1989) have shown that different features are central for natural kinds and artifacts: in natural kinds internal or molecular features are more conceptually central than functional features, but in artifacts functional features are more conceptually central than internal or molecular features.

In Gelman's study (1988) children first learned a feature novel to them for each type of category (e.g., this rabbit has a spleen inside) and were asked whether this feature is generalizable to another instance of the same category. The second graders in this study responded that features referring to substance and internal structure (e.g., "has a spleen inside", p. 74) were more generalizable for natural kinds, whereas functional features (e.g., "you can loll with it", p. 75) were more generalizable for artifacts.

Keil (1989) demonstrated a similar phenomenon using a transformation task. In this study, children learned, for instance, a story that a raccoon was dyed and painted with a white stripe, or a story that a coffeepot was transformed into a bird feeder by changing its parts. Fourth graders as well as adults in this study judged that changes in perceptual appearance did not matter for natural kinds' identity, but these changes did matter for artifacts' identity. Presumably, a perceptual change in an artifact is directly related to the function it performs, and, therefore, perceptual changes were judged to matter for artifact category membership. However, a new discovery about the origin of an artifact (e.g., a key that was, in fact, made of pennies) did not affect category membership.

It is tempting to take the content-based approach based on these findings and conclude that there is something inherently special about molecular features for natural kinds and functional features for artifacts (e.g., Barton & Komatsu, 1989). In contrast, we argue that the mechanism underlying this phenomenon is the causal status effect. That is, in natural kinds, internal / molecular features tend to cause functional features (e.g., cow DNA determines whether or not cows give milk) but in artifacts, functional features determine its compositional structure (e.g., chairs are used for sitting, and for that reason, they are made of a hard substance). In particular, Dennett (1987) dubbed the latter pattern in artifacts as the 'design stance'; a designer intentionally creates an artifact to fulfill a function and this intended function constrains its form, actual function, and the material it can be made from.

To test this structure-based hypothesis, Ahn (1998) examined the real-life categories used in previous studies (Barton & Komatsu, 1989; Malt & Johnson, 1992). Participants were asked to draw causal relations among features within a category. At the same time, they judged the centrality of features as measured by the degree to which a feature impacts categorization when that feature is missing. It was found that across natural and artifactual kinds, the more features any particular feature caused, the more influential the causal feature was in categorization. In addition, Ahn (1998) directly manipulated the causal status of features using artificial stimuli, and showed that when a compositional feature caused a functional feature, a compositional feature was more influential in categorization of both natural and artifactual kinds, whereas the opposite was true when the causal direction was reversed.

These results suggest that the content of features does not play a major role in categorization once their causal roles are taken into account. Thus, the more fundamental determinant of feature centrality is not content but rather the structural roles played by specific features. This way, the causal status hypothesis offers a more fundamental account than has previously been put forth for why people might treat natural kinds and artifacts differently.

Form versus Function Debate

Within the domain of artifacts, there has been debate on whether children prefer form or function of artifacts as a basis of extending novel names. For instance, children are taught that a novel artifact is named "Figley" and then asked which object they prefer to call "Figley," an object that shares the same perceptually salient feature but does not perform the same function, or an object that looks different but performs the same function. The results from studies using this paradigm are still controversial. We will describe two representative studies from recent literature, and argue why these results are more parsimoniously described by the causal status hypothesis.

The strongest evidence supporting the perceptual camp is described in Smith, Jones, and Landau (1996). Three-year-old children learned the names of novel artifacts, and were asked whether the same name applied to test objects. There were four test objects, constructed by replacing the bases and parts of the target objects with contrast objects. Across two experiments, it was found that children extended the novel names to test objects that share the same perceptual dimensions that were salient in that experiment. That is, when parts were salient, they extended the names to test objects that share the same part whereas when bases were salient, they extended names to test objects that share the same base. In two other experiments, Smith et al. taught participants novel functions of the same novel objects. Adult participants' naming was based on this functional information rather than perceptual salience. However, the responses of 3-year-old children remained based on the perceptually salient dimensions, and were not shifted by functional information.

Kemler Nelson, et al. (2000), however, argued that in order for functional information to be used by young children, it has to be compelling and nonarbitrary. That is, the function of an artifact should be based on principles of causality that are familiar to young children. For instance, consider an exemplar of "Amlas" in Smith et al.'s (1996) study. Its object-part biasing function was described to be, "A toy dog sits in." This function, however, might not have been used as a basis for naming due to its arbitrariness (e.g. a toy dog can sit almost anywhere). To test their hypothesis, Kemler Nelson et al. used either plausible or implausible functions. When a function was plausible, it provided a compelling explanation for why the objects had the structures they did. Drawing an analogy to real-life artifacts, "used for sitting" would be a plausible function for chairs because it explains why chairs have seats placed at a certain height, and backs attached to seats, etc., whereas "used to prop a door" would be a possible but implausible function for chairs because it does not explain many structural components of chairs. Kemler Nelson et al. taught 4-year- olds labels of novel artifacts and examined whether they extended these labels to functional but perceptually similar objects or dysfunctional but perceptually less similar objects. They found preference for functional objects in the plausible function condition but not in the implausible function condition. Note, however, that this study still provides no theoretical reason as to why causal relations between functions and structure have to be compelling.

The causal status hypothesis provides a natural account in these described situations. Unless children can understand that the object was created to perform a certain function, that is, the parts and shape of an object are constrained by the intended function, the functional feature would not have a high causal status. Consequently, functions would be less likely to be conceptually central. Once again, what is critical here is not whether any feature is "functional" per se, but rather the features' causal status.

Even in adult categorization literature, the form-function debate remains controversial. For instance, while Barton and Komatsu (1989) found that physical appearance matters much less than functional features in artifact categorization, Malt and Johnson (1992) found that the physical appearance of objects is more important than functional features in artifact categories. At the surface level, these two studies seem contradictory to each other. However, in the framework of the causal status theory, they are not necessarily conflicting because the particular physical features chosen as the stimulus materials in each study could have varied with respect to causal status. Malt and Johnson used a set of complex physical features, some of which seem causally connected to other features, and some not. For example, a taxi's physical feature was "has a meter for fare, two seats, and is painted yellow" and its function was "to provide private land travel for 1-4 people at a time when their own cars are unavailable and they are willing to pay a variable amount of money depending upon their specific destination(s)." Of the features that were classified as physical ones in Malt and Johnson, "has a meter for fare" seems to determine the taxi's function whereas "painted yellow" does not. Based on this observation, the causal status hypothesis can explain the apparent discrepancy in the following way. The physical features used in Barton and Komatsu were not causal, whereas at least some of the physical features used in Malt and Johnson appear to be more causal. In order to test this hypothesis, Ahn (1998, Experiment 2) examined the physical features used in Malt and Johnson. Participants received all the physical features used in Malt and Johnson and assessed conceptual centrality of each individual feature ("would X still be X if it were in all ways like X except that it does not have property Y?") and causal centrality of that feature by drawing causal relations among the features within a target category. The results showed that not all physical features were equally conceptually central, and more importantly, that the conceptual centrality of physical features correlated with their causal centrality.

Again, these studies demonstrate that the form-function debate should be considered from a new perspective. Dichotomizing features based on their content (i.e., forms and functions) does not aid in providing a clear picture of how people assess feature importance. Looking at the causal status of features, apart from their content, allows for a more coherent interpretation of the studies described thus far.

Appreciation for Intentionality

Recently, several developmental studies demonstrated that even young children appreciate intentionality over appearance for naming external representations such as drawings. Bloom and Markson (1998) asked 3- and 4-year-old children to draw, for instance, a balloon and a lollipop. As one might expect, these drawings looked almost identical. Later, when the children were asked to name these drawings, the children named the pictures on the basis of what they had intended to depict. Similarly, Gelman and Ebeling (1998) showed that when the same picture was produced either intentionally or accidentally (e.g., drawings had been intentionally created to be a bear versus somebody spilled paint), intentional representation led to higher rates of naming responses (e.g., "bear") than did accidental representation. Clearly, the intention behind a drawing is a causal factor of the drawings' appearance. Thus, appreciation for the drawer's intention when naming drawings can be construed as an example of the causal status effect.

Is this effect limited to the naming of artwork or visual representations? Even 3-year- old children are familiar with paintings, and therefore have had first-hand experiences of intentions behind drawings. Gelman and Bloom (2000) examined whether children would also be influenced by creator's intent when naming mundane artifacts such as tools and clothing. Participants in this study (3-year-olds, 5-year-olds, and adults) were presented with real artifacts (e.g., newspaper folded into the shape of a hat). They were told that the object was either intentionally created (e.g., "Jane went and got a newspaper. Then she carefully bent it and folded it till it was just right. Then she was done. This is what it looked like." p. 94) or accidentally created (e.g., "Jane was holding a newspaper. Then she dropped it by accident, and it fell under a car. She ran to get it and picked it up. This is what it looked like." p. 94). When told the object was intentionally created, participants in all age groups were more likely to name the object as the transformed object (e.g., hat in the above example). However, when told the object was accidentally created, they were more likely to name it by its material composition (e.g., newspaper). Again, the cause of an object's existence (i.e., intent) derived naming of the object.

Apparently conflicting results to the afore-mentioned studies were obtained in Matan and Carey (2001). Participants were asked to judge whether an artifact that was originally made for one purpose (e.g., making tea) and was actually being used for another purpose (e.g., watering flowers) should be named as the object originally intended (e.g., a teapot) or as the object actually used (e.g., a watering can). Consistent with the previous studies, 6-year-olds and adults named the object based on its original intended function. However, unlike the three studies described earlier in this section that showed even 3-year-olds are sensitive to the original intent (e.g., Gelman & Bloom, 2000), 4-year-olds in Matan and Carey's study virtually never named objects based on their intended function.

There are two inter-related differences that might explain the discrepancy. First, in studies showing the importance of creator's intent, the intent of the creator was that the object be a certain kind (e.g., intent to draw a lollipop). In contrast, in Matan and Carey's (2001) study, the intent was about the function rather than the kind. As reviewed in the previous section, function alone does not consistently determine artifact naming in children's categorization. It is when the function's causal status is plausible and explicit that the function predominantly constrains naming. However, participants in Matan and Carey's (2001) study were presented only with ambiguous unidentifiable objects (e.g., a spout sticking out from behind a wall, which could be interpreted as belonging to either a teapot or a watering can), which made it difficult for them to clearly grasp the causal status of the intended function (i.e., that the form and materials of the object are caused by the intended function). (See also Chaigneau & Barsalou, 2002.) Older children and adults might have an abstract framework theory about artifacts, such as design stance (Dennett, 1987), which they could rely on in the absence of concrete instantiations. But younger children might need more specific instantiations of such relations in a given object in order to let intended function override actual function in naming artifacts.

Although learning about abstract causal structures underlying the design stance may come later in development (approximately 6 years old according to Matan & Carey, 2001), recent studies present evidence supporting the idea that even infants prefer to categorize events based on an agent's goal rather than mere perceptual information when the goal information is concrete. For instance, Gergley, Nádasdy, Csibra, and Bíró (1995) presented 12-month-old infants with a scene, in which a small disc and a large disc were separated by a block wall. The small disc moved toward the large disc and then jumped over the block, and moved to the large disc. In the test phase, the block was removed. In one condition, the small disc repeated the same jumping action, and in the other condition, the small disc moved directly toward the large disc without jumping. Infants looked longer at the less optimal event where the small disc repeated the same jumping action even though that event should be more perceptually familiar to them. Presumably, the original scene gives viewers an impression that the goal of the small disc is to be by the large disc, so it jumps over the block. It does not need to jump when the block is removed so the infants look at the jumping disc longer because its behavior is no longer consistent with the imagined goal. Therefore, the results suggest that even infants are sensitive to the goal of an agent.

Woodward (1998) presents similar findings from even younger infants (as young as 6 months old). In this study infants were habituated to an event in which a hand and arm moved to grasp one of two toys. In the test events, the position of the two toys was reversed and infants saw the hand and arm reaching for the same toy (i.e., along a different path than in the habituation sessions) or a different toy (i.e., along the previously used path). Infants looked longer at the same path event; that is, when the goal was changed. Again, an agent's goal is the cause of the agent's behavior, and these two studies are consistent with the causal status hypothesis: participants categorize events based on the causal factor – goal – rather than the effect – behavior.

The causal relations between an agent's goal and the agent's behavior appear to be acquired very early on (compared to more abstract causal relations, say, design stance) because infants themselves carry out such causal actions. They themselves intend to perform a certain behavior and that intention leads to a certain action. Searle (1983) describes how an ascription of intentional cause can be acquired without multiple observations:

> For example, suppose I am thirsty and I take a drink of water. If someone asks me why I took a drink of water, I know the answer without any further observation: I was thirsty. Furthermore, in this sort of case it seems that I know the truth of the counterfactual without any further observations or any appeal to general laws… (Searle, 1983, p. 118)

To summarize, we propose that intentionality behind creation of an object is more conceptually central than the object's appearance because intentionality determines what the object looks like. If this causal status of intentionality is not obvious to a categorizer, intentionality may not be as conceptually central. In situations where the causal role of intentionality is most prominent (e.g., intention or goal of an actor), even young infants categorize events based on that dimension.

Evidence for the Primacy of the Causal Status Effect

Primacy in Development

Thus far, we illustrated how a large number of findings in children's categorization can be explained in terms of the causal status hypothesis. Yet, none of these tests with children directly tested the causal status hypothesis by holding the content constant and manipulating the causal status of features only. Ahn, Gelman, Amsterlaw, Hohenstein, and Kalish (2000) provide direct evidence of the causal status effect in 7- to 9-year-old children. In this study, adults and children learned descriptions of novel animals, in which one feature caused two other features. When asked to determine which transfer item was more likely to be an example of the animal they had learned, both adults and children preferred an animal with a cause feature and an effect feature than an animal with two effect features. Thus, children in this age group do indeed show the causal status bias. The paradigm used in this study would be difficult to use with younger children because participants had to learn novel features of an animal along with the causal relations among them, followed by a choice task between two transfer items. At this point, it remains to be seen whether direct evidence of the causal status effect can be obtained from children under seven.

Primacy in Categorization Processes

As discussed above, the causal status hypothesis posits that children possess abstract beliefs about the implications of causal structure. This suggestion has met with implicit resistance. As addressed in Keil, Smith, Simons, and Levin (1998), one reason for this is that theory-based reasoning has traditionally been thought of as a slow, reflective process (e.g., Sloman, 1996; Smith & Sloman, 1994). Perhaps because of this, theory-use has also been assumed to be difficult and thus more readily available to adults. This assumption is contained in the concrete to abstract developmental shift (e.g., Werner & Kaplan, 1963) and in the claim that perceptually-based categorization arises before categorization of conceptual features (e.g., Smith & Kemler Nelson, 1984). Keil et al. (1998) argue that this assumption is unwarranted. Below we present more direct evidence that theory-based categorization, and the causal status hypothesis in particular, is not a slow, deliberative process. (See Luhmann, Ahn, & Palmeri, 2002 for more detailed descriptions of this study.)

Experiment 1. Our stimuli consisted of four fictional animals (see Figure 3). Each animal was described as possessing three features (e.g., A, B, and C). The features were described as having a causal chain structure such that feature A causes feature B, and feature B causes feature C. In order to ensure that the three features' saliency did not vary in the absence of causal information, we removed the explicit causal information from the animal descriptions and asked a separate set of subjects to rate the likelihood of category membership of items missing a single feature (see Figure 4). The results of this pre-test showed no significant differences between the ratings of items missing the first feature, items missing the second feature, and items missing the third feature (all p's>.4), confirming that the features were equated for a priori strength.

Insert Figures 3 and 4 about here.

To allow for speeded responses, subjects in the main experiment were required to learn and memorize the four animals, their features, and the causal relations between the features. First, subjects were given the opportunity to study the description of each animal at the beginning of the experiment. While studying each description subjects were instructed to “write about how you think each feature causes the next,” in an attempt to force subjects to think causally about the features (instead of as a simple ordered list). To help subjects further learn the items, they were then presented with six trial blocks, during which they were prompted with the name of one of the animals and were required to select (using a mouse-click) in the appropriate causal order the features of that animal from an array containing the features of all four animals. In the first two blocks responses were unspeeded, while in the last four blocks responses had 5-second deadlines so that the novel causal background knowledge would be sufficiently internalized. In addition, on half of the blocks subjects were asked for the causal relations in the forward order (e.g., A, B, C), while in the other half in the backward order (e.g., C, B, A).

Once subjects completed these six blocks they proceeded to the experimental transfer task. Subjects were presented with items missing a single feature and were asked to rate the likelihood that the item belonged to its target category on an 8-point scale (with 1 labeled as “Definitely Unlikely” and 8 labeled as “Definitely Likely”). There were four blocks of trials in the transfer task. In two of the blocks subjects were instructed to answer as quickly as possible. In the other two blocks they were told to take as much time as needed. The reaction times in the speeded blocks (M = 1560ms) were indeed significantly faster than the times in the unspeeded blocks (M = 3202ms), $p<.05$, Tukey’s HSD.

For the unspeeded trials we expected to find results similar to those of Experiment 1 of Ahn, Kim, et al. (2000). That is, items missing the terminal effect feature should be rated as more likely category members than those missing the initial cause feature. The critical question was whether this causal status effect would disappear during the speeded trials.

The results for subjects' categorization responses are summarized in Figure 5. A 2 (speed condition: speeded vs. unspeeded) X 3 (item type: missing first feature vs. missing second feature vs. missing third feature) repeated measures ANOVA was performed on the data. We observed a significant main effect of item type, $F(2, 56)=22.69$, $p<.0001$. Planned comparisons showed that in both the speeded and unspeeded conditions items missing the third feature were rated significantly higher than those missing the first or second features (p's<.05, Tukey's HSD). The difference between items missing the first feature and those missing the second feature was not significant (p's>.05, Tukey's HSD), possibly because the second feature also served as a cause of another feature, making the difference between the first and the second feature less pronounced (see also Kim and Ahn, 2002a). In addition, we observed no main effect of speed, $F<1$, and the speed X item type interaction was also not significant, $F(2,56)=1.03$, $p>.05$. Overall, these results demonstrate that the causal status effect occurs even when time for lengthy reflection is not allowed.

Insert Figure 5 about here.

<u>Experiment 2.</u> In Experiment 1, subjects were simply asked to respond as quickly as possible to the "speeded" items. Given this freedom, some subjects responded very quickly but others responded significantly more slowly. Although the speed manipulation

we used in Experiment 1 is naturalistic in that subjects carried out what they thought to be a rapid decision making process, forcing subjects to respond within a specific deadline ensures uniform time pressure across all subjects and items. Therefore, in Experiment 2, we imposed stricter control over subjects' response times by enforcing deadlines on subjects' responses.

One methodological complication with establishing appropriate response deadlines is that it is difficult to determine beforehand whether a particular deadline is short enough to challenge the categorization system but not so short as to make accurate responses impossible. That is, if the speeded condition does not show the causal status effect, it may be because theory-based reasoning does not take place during rapid categorization, or because the deadline is too short to produce any reasonable responses.

For this reason, we also tested whether similarity information could be used under similar deadlines. By testing both kinds of knowledge, the casual status effects can be compared to similarity-based categorization at each deadline. In this way it can be inferred whether any breakdown of the causal status effect is due to the inability to complete the processes necessary for theory-based categorization or if reasonable responses at that deadline are impossible for both kinds of categorization.

There were two conditions in Experiment 2, one representing a theory-based situation and the other representing a similarity-based situation. The condition representing a theory-based situation was the Causal condition, where subjects were given the same stimuli and accompanying causal information as used in Experiment 1. The condition representing a similarity-based situation was the Base-Rate condition, where subjects were provided with information about relative base rates of each feature within a category. More specifically, each category was described as having three features (e.g., A, B, and C) such that 100% of category members possessed feature A, 80% of possessed feature B, and 60% possessed feature C. It was thought that these base rates (or a measure also known as category validity, Rosch & Mervis, 1975) serve as a similarity-based determinant for feature weighting because similarity is frequently calculated based on how many attributes an item has in common with other members of the category (e.g., Tversky, 1977). Paralleling the results of Experiment 1, items in the Base-Rate condition missing the third (60%) feature should be rated as better category members than those missing the first (100%) feature. Using this condition as a point of comparison, and with the addition of strict response deadlines, we hoped to provide a more rigorous test of the causal status effect under speeded conditions.

The learning phase for the Causal condition was identical to that used in Experiment 1. Subjects in the Base-Rate condition did not have to generate explanations but instead categorized exemplars into one of the four animal categories. For this task, each exemplar always possessed the first feature of its category, possessed the second feature on 80% of the trials, and the third feature 60% of the time (thus mirroring the stated base rates). Feedback was given after each trial during the learning phase.

Blocks of 30 such trials alternated with blocks of a “selection task” like that used in the Causal condition. The directions for the selection task instructed those subjects in the Base-Rate condition to select features in an order (forward or backwards) dictated by their base rate percentages rather than their position in the causal chain.

The transfer phase for both conditions was nearly identical to that used in Experiment 1 except for a modified speed manipulation. Instead of an instruction to respond quickly, Experiment 2 employed a signal-to-respond technique (Lamberts, 1998). Thus, every trial presented the feature triad (Figure 4) for a certain set amount of time (see below). Subjects responded to the item once the presentation was completed and the item disappeared from the screen. If a response was made more than 300ms after the disappearance of the triad, subjects were told to respond more rapidly. There were four blocks of trials. Each block used one of four durations (5000ms, 2250ms, 1500ms, and 750ms) for the presentation of the triads.

The results from the categorization task can be seen in Figures 6 and 7. A 2 (knowledge condition: Causal vs. Base-Rate) X 4 (speed condition: 5000ms vs. 2250ms vs. 1500ms vs. 750ms) X 2 (item type: missing first feature vs. missing third) ANOVA was performed with repeated measures on the latter two factors. Neither the main effect of, nor any interaction with the knowledge condition (feature frequency vs. causal) was found to be significant. The main effect of item type was found to be significant, $F(1, 15)=45.6$, $p<.0001$. Planned comparisons were conducted to determine whether a significant effect of item type was present at each of the response deadlines for each of the knowledge conditions. For simplicity, we include only those comparisons between items missing the first feature and those missing the third, the difference CSH predicts to be the largest. For both knowledge conditions and at all response deadlines the items missing the third feature were rated as significantly better category members than those items missing the first feature (all $p<.05$).

Insert Figures 6 and 7 about here.

<u>Experiment 3</u>. Experiment 3 used faster response deadlines (1500, 750ms, 500ms, and 300ms). The other aspect of the method was identical to that of Experiment 2.

The results from the categorization task can be seen in Figures 8 and 9. A 2 (knowledge condition: Causal vs. Base-Rate) X 4 (speed condition: 1500ms vs. 750ms vs. 500ms vs. 300ms) X 2 (item type: missing first feature vs. missing third) ANOVA was performed with repeated measures on the latter two factors. We observed a significant main effect of item type, $F(1, 58)=15.14$, $p<.0005$, that did not interact with knowledge condition, $F<1$, demonstrating that both background conditions had the predicted effect on categorization behaviors. Planned comparisons were carried out to determine at what response deadlines the background information had an effect on categorization (items missing the second feature were again excluded). For the Base-Rate condition, items missing the first (100%) feature significantly differed from items missing the third (60%) feature in the 1500ms condition, $t(29)=3.43$, $p<.005$, and the 750ms condition, $t(29)=2.41$, $p<.05$, but not in the 500ms, $t(29)=.3$, $p>.05$, or 300ms, $t(29)=1.59$, $p>.05$, conditions. In the Causal condition, items missing the first (initial cause) feature differed from those missing the third (terminal effect) feature in the 1500ms condition, $t(29)=2.22$, $p<.05$, the 750ms condition, $t(29)=2.86$, $p<.01$, and the 500ms condition, $t(29)=2.06$, $p<.05$, but not the 300ms condition, $t(29)=.81$, $p>.05$.

Insert Figures 8 and 9 about here.

<u>Discussion</u>. These results support the idea that theory-use is as fast as, if not faster than, comparable similarity-use. In particular, subjects in Experiment 3 were able to categorize according to their theory even when allowed only 500ms to view the exemplar and make a response. Furthermore, our results indicate that the base-rate information,

which has been considered a key determinant of similarity (Rosch and Mervis, 1975), did not result in differential responses under this deadline. The results taken together provide strong evidence that the causal status effect cannot be slower than the frequency effect.

Conclusion

One dominant line of theory is that children's initial concepts are concrete and perceptually based, and only later do they acquire the more conceptually based categories that adults have (e.g., Inhelder & Piaget, 1964). Presumably because of this tradition, most debates on children's conceptual representations are framed in terms of the use of perceptual versus non-perceptual features. In this chapter, we offer a new perspective on these issues. We propose that recent findings favoring either the perceptual camp or the non-perceptual camp can be both interpreted in terms of causal structures in concepts, and provided several examples in this chapter. We presented several arguments for the advantages of this structure-based approach. The causal status hypothesis provides a more parsimonious account for various phenomena without having to resort to ill-defined concepts, such as domains. Furthermore, the causal status hypothesis can provide precise predictions about novel domains once we know the causal structure of concepts, whereas the content-based approach is merely descriptive and fails to provide predictions about novel domains. In most existing studies, however, the content effect was confounded with the causal status effect. When they are pitted against each other, evidence favors the causal status effect.

Unfortunately, no existing study compared the content effect against the causal status effect among infants. Therefore, it is difficult to determine how developmentally primary the causal status effect is. One possible reason to argue why the causal status effect might not be present in young infants is the idea that theory-based reasoning is deliberate and takes a long time: after all, those tasks that adults need time to complete (e.g., long division) are not usually readily available to children. Our recent study, however, found that the causal status effect occurs as rapidly as the similarity-based effect in adult categorization. Thus, the results rule out the possibility that the causal status effect is too analytic to be used by very young children. Instead, it further opens up a door to the possibility that children would use this process, however rudimentarily, early in life.

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.

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

Ahn, W., & Kim, N. S. (2000). The role of causal status of features in categorization: An overview. In D. L. Medin (Ed.), *Psychology of Learning and Motivation* (Vol. 40, pp. 23-65). New York: Academic Press.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 1-55.

Ahn, Marsh, Luhmann, and Lee (2002). Effect of theory-based feature correlations on typicality judgments. *Memory and Cognition, 30,* 107-118.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research, 18*, 433-447.

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science, 9*, 200-204.

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

Chaigneau, S. E., & Barsalou, L. W. (2002). Testing the roles of design history and affordances in the HIPE theory of function. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 30). W. D. Gray & C. Schunn (Eds.) Lawrence Erlbaum Associates, NJ.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: The MIT Press.

Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review, 6*, 338-346.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.

Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition, 76*, 91-103.

Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146-196). New York: Cambridge University Press.

Gelman, S. A., & Diesendruck, G. (1999). What's in a concept? Context, variability, and psychological essentialism. In I. E. Sigel (Ed.), *Development of mental representation: Theories and applications* (pp. 87-111). Mahwah, NJ: Lawrence Erlbaum Associates.

Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition, 66*, B35-B47.

Gelman, S. A., & Hirschfeld, L. A. (1999). How biological is essentialism? In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 403-446). Cambridge, MA: The MIT Press.

Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development* (Vol. 2). New York: Springer Verlag.

Gelman, S. A., & Koenig, M. A. (in press). Theory-based categorization in early childhood. In D. Rakison & L. Oakes (Eds.), *Category and concept development*. Oxford University Press.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165-193.

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

Inhelder, B., & Piaget, J. *The early growth of logic in the child*. New York: W. W. Norton & Company.

Johnson, K. E., Mervis, C. B., & Boster, J. S. (1992). Developmental changes within the structure of the mammal domain. *Developmental Psychology, 28*, 74-83.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development, 62*, 499-516.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136-153.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: The MIT Press.

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

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

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: implications for hybrid models of the structure of knowledge. *Cognition, 65*, 103-135.

Kemler Nelson, D. G., Frankenfield, A., Morris,C., & Blair, C. (2000). Young children's use of functional information to categorize artifacts: three factors that matter. *Cognition*, 77, 133-168.

Kim, N. S., & Ahn, W. (in press). Clinical Psychologists' Theory-Based Representations of Mental Disorders Affect their Diagnostic Reasoning and Memory. *Journal of Experimental Psychology: General*.

Kim, N. S., & Ahn, W. (2002). The influence of naive causal theories on lay concepts of mental illness. *American Journal of Psychology*, 115, 33-65.

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 695-711.

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

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1153-1169.

Luhmann, C. C., Ahn, W., Palmeri, T. J. (in press). Theories and similarity: Categorization under speeded conditions. Paper presented at the 24th Annual Conference of the Cognitive Science Society, George Mason University.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of memory and language, 31*, 195-217.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development, 8*, 291-318.

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

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Rakison, D. H., & Poulin Dubois, D. (2001). Developmental origin of the animate inanimate distinction. *Psychological Bulletin, 127*, 209-228.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York: Cambridge University Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Searle, J. R. (1983). *Intentionality*. New York: Cambridge University Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22*, 377-386.

Smith, J. D., & Kemler-Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General, 113*, 137-159.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*, 143-171.

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

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: The MIT Press.

Werner, H., & Kaplan, B. (1963). Symbol formation: An organismic-developmental approach to language and the expression of thought. New York: Wiley.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.

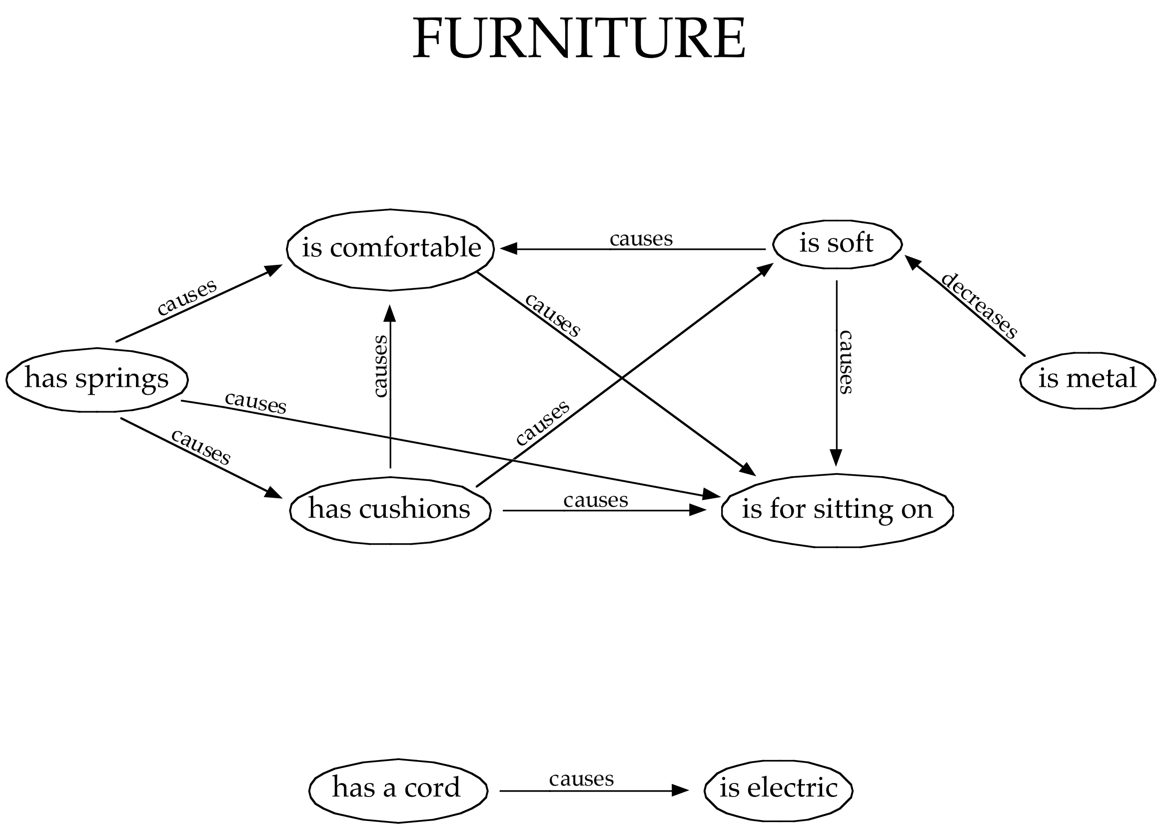
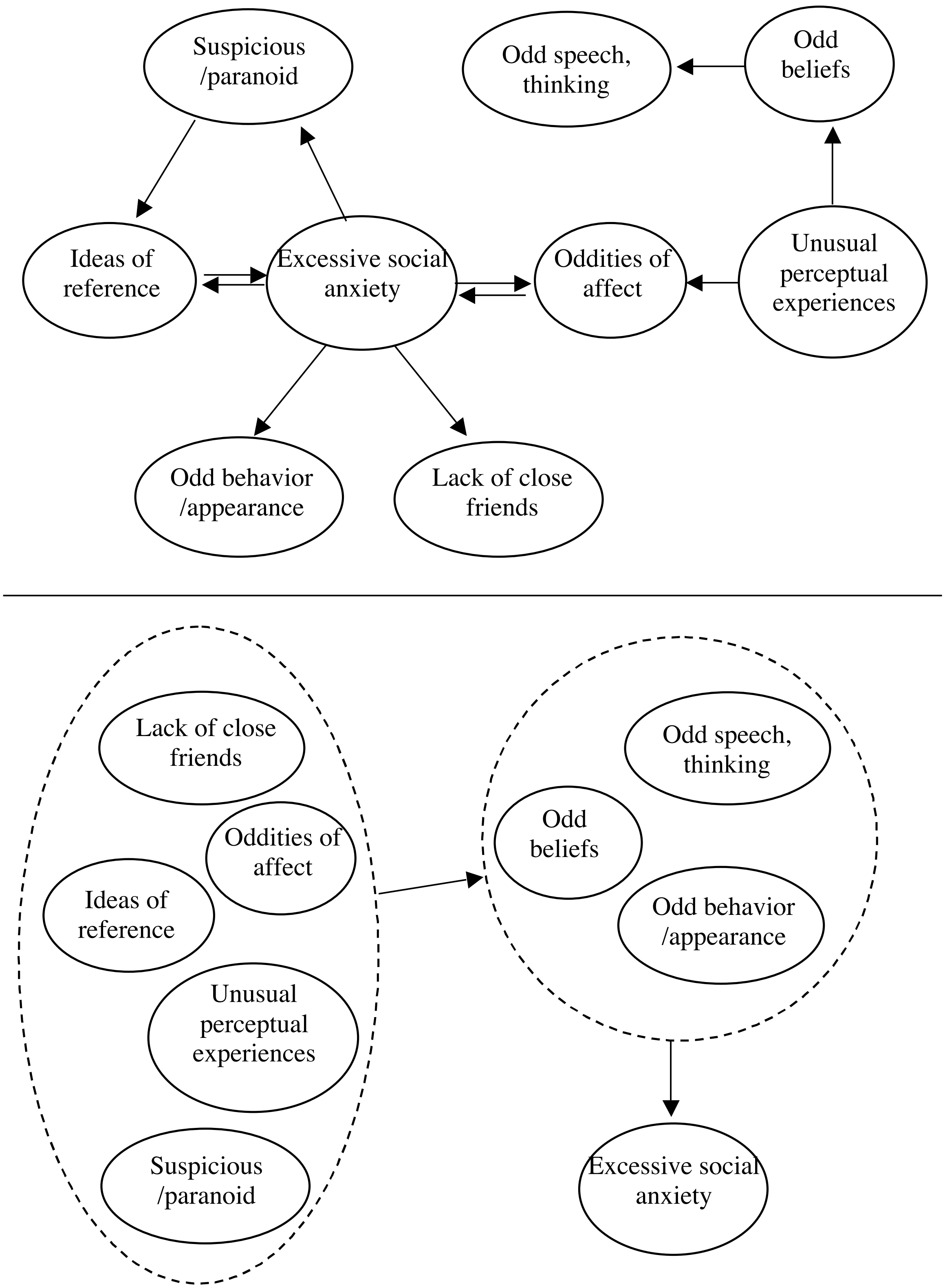


Figure 1. An averaged lay-theory of furniture found in Ahn et al. (2002)

Figure 2. Sample data showing disagreement in theories for schizotypal personality disorder in Kim and Ahn (2002b).



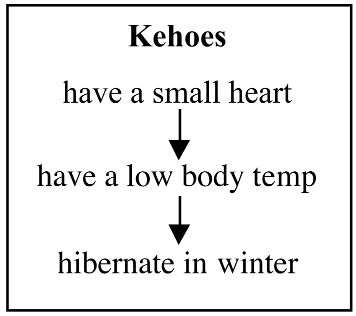


Figure 3. A sample animal with causal links

Kehoe?

does not have a low body temp

has a small heart   hibernates in winter

Figure 4: A sample item from Experiment 1

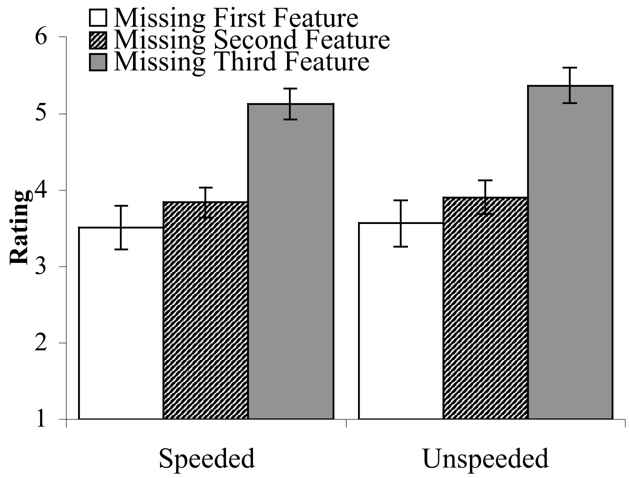


Figure 5. Results from Experiment 1

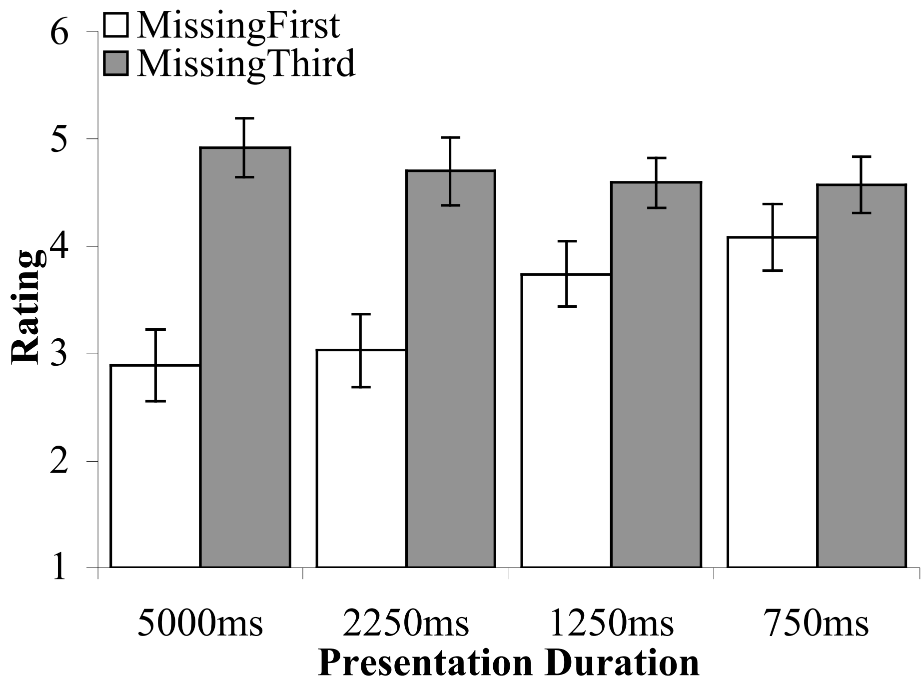


Figure 6. Results from the Causal Condition, Experiment 2

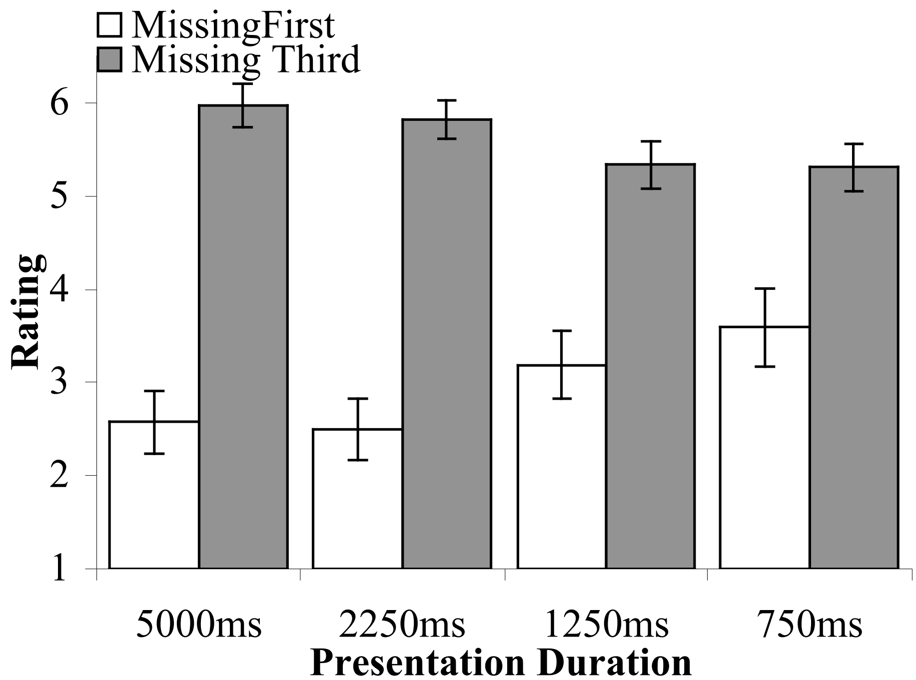


Figure 7. Results from the Base Rate Condition, Experiment 2

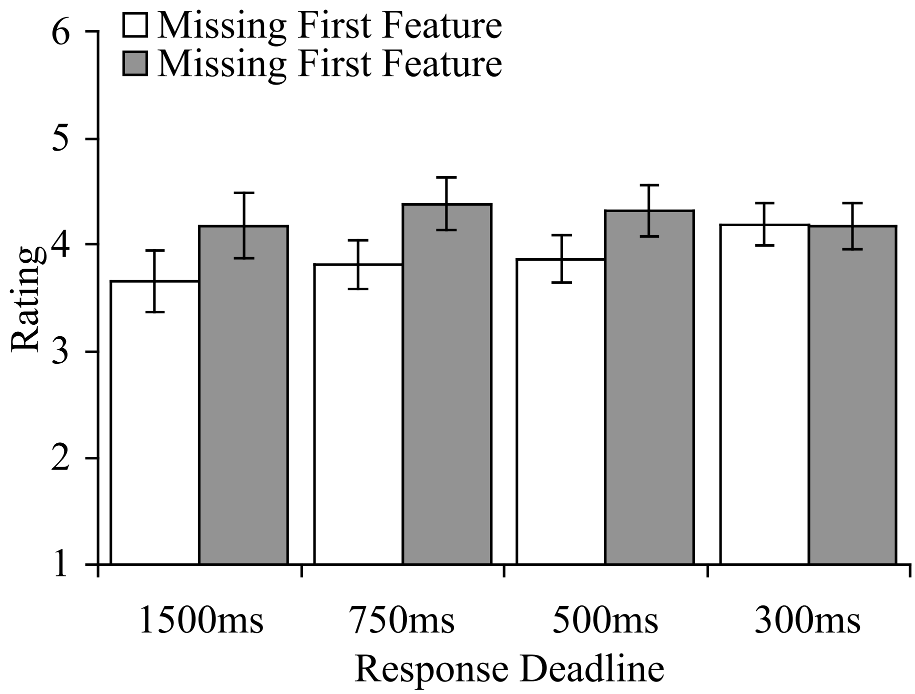


Figure 8. Results from the Causal Condition, Experiment 3

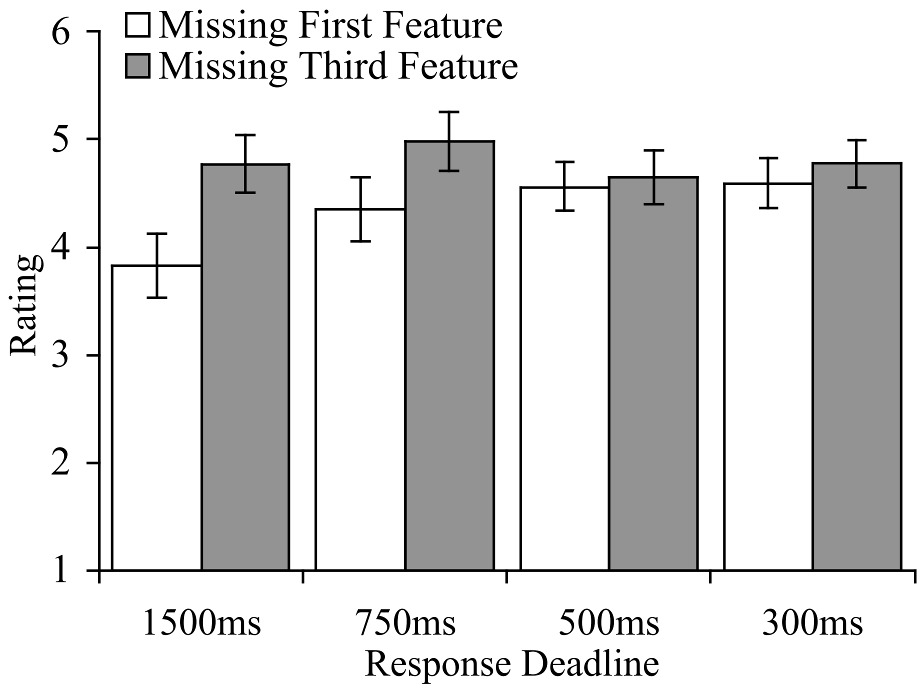


Figure 9. Results from the Base Rate Condition, Experiment 3