How Categories Shape Causality

Michael R. Waldmann (michael.waldmann@bio.uni-goettingen.de)
Department of Psychology, University of Göttingen,
Gosslerstr. 14, 37073 Göttingen, Germany

York Hagmayer (york.hagmayer@bio.uni-goettingen.de)
Department of Psychology, University of Göttingen,
Gosslerstr. 14, 37073 Göttingen, Germany

Abstract

The standard approach guiding research on the relationship between categories and causality views categories as reflecting causal relations in the world. We provide evidence that the opposite direction also holds: Categories that have been acquired in previous learning contexts may influence subsequent causal learning. In three experiments we show that identical causal learning experiences yield different attributions of causal capacity depending on the pre-existing categories that the learning exemplars are assigned to. There is a strong tendency to continue to use old conceptual schemes rather than switch to new ones even when the old categories are not optimal for predicting the new effect. This tendency is particularly strong when there is a plausible semantic link between the categories and the new causal hypothesis under investigation.

Introduction

The Standard View: Causality Shapes Categories

The standard view guiding research on causality presupposes the existence of networks of causes and effects in the world that cognitive systems try to mirror. Regardless of whether causal learning is viewed as the attempt to induce causality on the basis of statistical information or on the basis of mechanism information it is typically assumed that the goal of causal learning is to form adequate representations of the texture of the causal world (see Shanks, Holyoak, & Medin, 1996, for an overview of recent research). This view also underlies research on the relationship between categories and causality. According to the view that categorization is theory-based traditional similarity-based accounts of categorization are deficient because they ignore the fact that many categories are grounded in knowledge about causal structures (Murphy & Medin, 1985). As Murphy and Medin pointed out, natural kind categories (e.g., animals) are not adequately represented as lists of features because this format excludes functional and causal relations that also are part of our category knowledge. Categories should rather be seen as embodying intuitive theories of the target domain. One example of the standard view is Waldmann, Holyoak, and Fratianne's (1995) work on causal categories (also see Waldmann, 1996). In a series of experiments they have shown that category learning is affected by assumptions about the causal structure underlying the categories (e.g., disease categories).

The Neglected Direction: Categories Shape Causality

Even though it is certainly true that in many cases knowledge of causal structures influences the way categories are formed, the opposite may also hold true: The categories that have been acquired in previous learning contexts may have a crucial influence on subsequent causal learning. This direction has typically been neglected in research on the relationship between categories and causality.

The basis of the potential influence of categories on causal induction lies in the fact that the acquisition and use of causal knowledge is based on categorized events. Regardless of whether causal relations are viewed as statistical relations (probabilistic causality view) or as mechanisms (mechanism view), both accounts postulate causal regularities that refer to *types* of events. Causal laws, such as the fact that smoking causes heart disease, can only be noticed on the basis of events that are categorized (e.g., events of smoking and cases of heart disease). Without such categories causal laws neither could be detected nor could causal knowledge be applied to new cases. Thus, causal knowledge not only affects the creation of categories, it also presupposes already existing categories for the description of causes and effects.

Given that the induction of new causal knowledge is based on already existing categories the question arises whether the outcome of causal learning may be influenced by the categories that are used. The potential influence of categories is due to the fact that one of the most important cues to causality is statistical covariation between causes and effects. Many (otherwise conflicting) views agree that causal induction is based on the observation of causes altering the probability of effects (e.g., contingency view; associationist theories)(see Shanks et al., 1996). However, statistical regularities are not invariant across different categorial segmentations of domains. This can easily be shown with a simple example. Let us assume, for example, a world with four different (uncategorized) event tokens, A, B, C, and D, that represent potential causes. It has been observed that A and C are followed by a specific effect but B and D are not. Now the statistical regularities that are observed in this mini-world are crucially dependent on how these four events are categorized. If A and B are exemplars of Category 1, and C and D exemplars of Category 2, no causal regularity would be observed. Within this conceptual

framework the effect has a base rate of 0.5 that is invariant across the two categories. By contrast, categorizing A and C (Category 3), and B and D (Category 4) together would lead to the induction of a deterministic causal law. Events that belong to Category 3 always produce the effect, whereas Category 4 is never associated with the effect. Thus, the causal regularities observed in a domain are dependent on the way the domain is categorized. In fact, as pointed out by Clark and Thornton (1997) in an example with (non-causal) continuous features, there is an infinite number of descriptions of the world with a potentially infinite number of statistical regularities entailed by these descriptions.

At this point it could be argued that the dependence of causal knowledge on pre-existing categories is a philosophical rather than a psychological problem as long as it has not been shown that there is evidence for the possibility of different categorizations of the same domains. Following the work of Rosch on natural categories many psychologists have assumed that natural categories are relatively stable since they are reflecting the correlational structure in the world (see Rosch, 1978). However, recently it became clear that this assumption is too strong. For example, Medin, Lynch, Coley, and Atran (1997) have shown that the way natural objects (e.g., trees) are categorized is dependent on pragmatic factors such as the profession of the categorizer (also see Barsalou, 1983). Schyns, Goldstone, and Thibaut (1998) have demonstrated that not even the object features used in categorizations are invariant. Their work demonstrates that the way the world is perceived may be influenced by the categories that are being used.

Another reason for the potential bi-directional interaction of categories and causality is the dynamic character of knowledge acquisition. Causal knowledge, in everyday life as well as in science, is typically not acquired at one point in time after which it remains stable but is rather the result of a long process in which it undergoes dynamic changes such as continuous modifications or even paradigm shifts (see Carey, 1991; Horwich, 1993). Categories acquired in specific contexts may not always be optimal for the new learning task at hand. For example, a learner who is equipped with Categories 1 and 2 in our example may be better off abandoning the old conceptual scheme altogether and instead forming Categories 3 and 4 that allow her to optimize predictability. On the other hand, switching to a novel conceptual scheme or keeping two different schemes in parallel incurs a cost that is computationally demanding. Therefore, there is a possible trade-off between sticking to old conceptual schemes that may not be currently optimal and switching to a new paradigm.

The hypothesized impact of pre-existing categories on causal learning constitutes a new type of transfer effect. Unlike in research on analogical transfer (see Holyoak & Thagard, 1995), no specific relational knowledge is transferred. The transfer effect is rather based on the indirect influence pre-existing categories may have on the statistical regularities observed in a domain. For example, traditionally, psychiatric diseases were classified on the basis of a taxonomy of symptoms, whereas today many researcher are more interested in neuropsychological analyses for which the old categories may not be optimal anymore. The original categories have never been created with the new research questions in mind. Nevertheless, it is possible that the continued use of the old categories in the new context may seriously bias the outcome of the causal investigations.

The following three experiments demonstrate how causal induction is affected by the way a novel domain is categorized. It will be shown that participants tend to use category knowledge acquired in a different context in a subsequent causal learning task.

Experiment 1

The goal of this experiment was to demonstrate how the way exemplars in a domain are categorized affects causal learning. The experiment consisted of three phases: In Phase 1, the *category learning* phase, participants were told that scientists had discovered new types of viruses that vary in the dimensions brightness, size, shape, and number of molecules on the surface. Cytophysiological investigations had revealed two types of viruses which can be distinguished on the basis of their appearance, *allovedic* and *hemovedic* viruses. After these instructions participants received index cards with pictures of viruses one after another, and had to judge whether the respective exemplar represented a hemovedic or an allovedic virus. After each judgment feedback was given. Learning proceeded until participants met a learning criterion, 10 correct classifications in a row. The exemplars varied continuously in the four features. The two relevant features were size and brightness. The diameter of the viruses varied between 30 and 48mm (Levels 1 to 4 in Table 1), and brightness was manipulated by using four equally spaced levels of grayness (20% to 80%)(Levels 1 to 4 in Table 1). The two irrelevant features also came in four levels which allowed us to create 256 different items. Our goal was to discourage exemplar learning. Table 1 shows examples of the 16 crucial types of viruses than can be created by factorially combining four values of size and brightness.

Two conditions were compared: In the size condition participants learned, for example, that the bigger viruses were allovedic, and the smaller ones hemovedic, in the orthogonal brightness condition they learned, for example, that the darker exemplars were allovedic and the lighter ones hemovedic. In Phase 1 128 different exemplars were presented to the participants.

While Phase 1 differed between conditions, the subsequent Phases 2 and 3 were *identical* across conditions. In Phase 2, the *causal learning* phase, participants were told that physicians were interested in exploring the relationship between the newly discovered viruses and diseases in animals. In particular, they wanted to find out whether the viruses cause splenomegaly, that is a swelling of the spleen. Therefore they studied animals that were infected with the new viruses. It was pointed out that any outcome of this study was possible including the possibility that there was no causal relationship between the viruses and splenomegaly. In Phase 2 participants saw a new set of 32 viruses one after another representing single instances of the viruses. On the backside of each card information was given on whether the respective virus had caused splenomegaly or not. In all

conditions the same items with identical associations with the effect were presented to participants.

Table 1: Structure of learning items in Experiments 1 and 3 (see text for explanations).

1	1 (A)	2 (B) ~ effect	3 (B) ~ effect	4 (A)
2	5 (B) effect	6 (A)	7 (A)	8 (B) ~ effect
3	9 (B) effect	10 (A)	11 (A)	12 (B) ~ effect
4	13 (A)	14 (B) effect	15 (B) effect	16 (A)
Size	1	2	3	4
	Brightness			

Table 1 displays the structure of the items with respect to the two relevant dimensions size and brightness. In the brightness condition the left half of the table (Levels 1 and 2 of the brightness dimension) may represent allovedic viruses and the right half hemovedic viruses (Levels 3 and 4). By contrast, in the size condition the upper half represented one category (Levels 1 and 2 of the size dimension), and the lower half the other category (Levels 3 and 4). Half of the items, indicated by an A, were shown in the category learning phase (Phase 1), the other half (indicated by a B) in the causal learning phase (Phase 2). (Again, our goal was to discourage exemplar encoding by presenting items that differed in their appearance.) Table 1 also shows which of the B-items caused the effect splenomegaly (*effect*), and which did not (*~effect*). The assignment of exemplars to the learning phases (A, B), and the association of dimensional attributes (dark, light, big, small) with the effects was counterbalanced.

In Phase 3, the *test* phase, we switched back to exemplars corresponding to the A-items from the category learning phase. These items had not been presented in the causal learning phase. Participants received eight exemplars. Their task was to express their assessment of the likelihood that the respective virus causes splenomegaly by using a rating scale that ranged from 0 ("never") to 100 ("always"). After these ratings participants also gave a general assessment of the likelihood that the two virus types, allovedic and hemovedic viruses, caused the effect. These ratings allowed us to check whether participants encoded the causal relation on the category level.

Despite the fact that participants in the two conditions received identical learning inputs in the causal learning phase (Phase 2), and were confronted with identical test items (Phase 3) we expected that the different categories learned in Phase 1 would influence the causal judgments. Item 1 (A) in Table 1 may exemplify our predictions: We expected that participants in the size condition would group this item along with the other items in the category of relatively small viruses (Items 1 to 8). Since only one out of four items (Items 2, 3, 5, 8) from the group presented in the causal learning phase produced the effect, it seems reasonable to infer that Item 1 would have a relatively low likelihood of causing splenomegaly. By contrast, participants in the brightness condition were expected to classify this item along with the other relatively light viruses. Within this group three out of four viruses (Items 2, 5, 9, 14) caused splenomegaly which should lead participants in this condition to give relatively high ratings. A similar prediction can be derived for Item 6, whereas for Items 11 and 16 the inverse pattern is predicted (i.e., high ratings in the size condition, and low ratings in the brightness condition). The remaining four test items (4, 7, 10, 13) should not yield any differences across the category conditions because they were associated with the same number of effects regardless of which category structure was used.

Alternatively, participants could ignore the category-level information from Phase 1 in the causal learning phase, and compare the test exemplar with the causal pattern of adjacent exemplars (exemplar-based learning), or they could create new categories that are more optimal for the causal task at hand. Both strategies should not lead to any differences in the ratings of the items across the two category conditions.

Results and Discussion

The results are based on 48 participants, 24 in the size condition and 24 in the brightness condition. (Two participants were excluded because they did not meet the learning criterion in Phase 1.) The most interesting analyses involved the test items whose ratings should differ across conditions (e.g., Items 1, 6, 11, 16 in Table 1). To make these items comparable, the ratings of Items 11 and 16 were assigned the inverse rating (e.g., 80 was recoded as 20). For the statistical analyses the average of these four items was used.

The results clearly confirm the predictions. The mean ratings of the four critical items clearly differed depending on which categories had been learned in the prior category learning phase, $F(1,46)=11.8$, $p<.001$, $MSE=350.9$. The mean ratings for these critical items were 43.5 versus 62.1 in the two contrasting category learning conditions. By contrast, no reliable difference was obtained for the test items which were not expected to differ across conditions ($M=62.7$ vs. $M=65.8$). A 2 (category conditions) × 2 (critical vs. non-critical items) analysis of variance yielded a

significant interaction, $F(1,46)=5.62$, $p<.05$, $MSE=253.8$. These results were mirrored in the final category-effect ratings. The mean ratings for the category with strong associations with the effect was 70.0, the contrast category with weak associations yielded a mean value of 24.6. All but four participants who rated both categories equally gave ratings consistent with this difference. An analysis of variance with categories as a within-subject factor yielded a clearly significant effect, $F(1,47)=153.5$, $p<.001$, $MSE=299.2$. This effect was independent of whether the categories were based on the size or the brightness of the items. Thus, participants not only registered the association of the individual exemplars with the effect but they also encoded the causal relations on the category level.

In summary, despite the fact that participants in all conditions received identical learning inputs in the causal learning phase, and had to make predictions about identical exemplars in the test phase, the attribution of causal capacity clearly differed depending on the categories the exemplars were assigned to in Phase 1. The same virus exemplars were either seen as causally effective or ineffective with respect to splenomegaly.

Although it is certainly true that viruses may generally be viewed as categories that are potentially responsible for symptoms such as splenomegaly, it is by no means certain that viruses that have been classified on the basis of their *appearance* represent classes that are optimal with respect to all kinds of effects that later may be studied. Potential causal links between the viruses and specific symptoms never were an issue when the rationale for the classifications in Phase 1 was introduced. In fact, other classifications (e.g., segmenting the exemplars in Table 1 using a diagonal boundary) are better able to capture the observed causal regularities. Nevertheless, Experiment 1 shows that participants rather continued to use categories acquired in a different learning context than create new categories for the induction task at hand.

Experiment 2

Experiment 1 used relatively simple category structures that were based on one-dimensional rules. By contrast, the rules underlying the causal regularities in Phase 2 were quite complex relative to the categorization rules. In Experiment 2 we investigated a task with more realistic, complex category structures, and with a comparably complex causal structure. In this experiment we used linearly separable, family resemblance categories that were based on four binary features. None of these features was individually sufficient for achieving correct classifications. However, correct classifications could be learned by an additive integration of the four features.

As learning exemplars we used items similar to the ones in Experiment 1. In the present experiment the exemplars varied on *four binary* dimensions, however: brightness (20% vs. 60%), size (30mm vs. 42mm), number of corners (5 vs. 7), and number of molecules on the surface (2 vs. 4). Table 2 displays the structures of the learning items with the feature value 1 representing high values and the value 0 low values.

Again the cover stories from Experiment 1 were used so that participants' task was to categorize the items into allovedic and hemovedic viruses in Phase 1 of the experiment. Two categorization conditions were compared that manipulated the location of the category boundaries (see Table 2). In *Condition A* hemovedic viruses had at least two high values on the four dimensions (Items 1 to 11), whereas allovedic viruses (Items 12 to 16) only had one or no high value. (The category labels were counterbalanced.) By contrast, in *Condition B* hemovedic viruses (Items 1 to 5) had at least three high values, whereas allovedic viruses (Items 6 to 16) only had two high values or less. Again we used a learning criterion in Phase 1. Learning proceeded until participants managed to correctly classify one block of 16 items (maximum of 8 blocks). The items were presented in random orders within blocks.

Table 2: Structure of learning items in Experiment 2

Items	Features				Effect	Categorization	
						A	B
1	1	1	1	1	E	Hemovedic viruses	Hemovedic viruses
2	1	1	1	0	E		
3	1	1	0	1	E		
4	1	0	1	1	E		
5	0	1	1	1	E		
6	1	1	0	0	E		Allovedic viruses
7	0	0	1	1	E		
8	0	1	1	0			
9	1	0	0	1	-		
10	1	0	1	0	~E		
11	0	1	0	1	~E		
12	0	0	0	1	~E	Allovedic viruses	
13	0	0	1	0	~E		
14	0	1	0	0	~E		
15	1	0	0	0	~E		
16	0	0	0	0	~E		

Whereas Phase 1 differed across the two conditions, the subsequent causal learning phase and the test phase were *identical* for all participants. Again participants received index cards that depicted exemplars of the viruses with information on the backside on whether the respective virus causes splenomegaly (E) or not ($\sim E$). To avoid an unequal association of individual features with the effect Items 8 and 9 were not presented in this phase. In the particular counterbalancing condition shown in Table 1 the upper half of the items (1-7) caused the effect, whereas the lower half (10-16) did not cause it.

In Phase 3, the test phase, participants received ten exemplars (1, 3, 6 to 11, 14, 16) and had to assess their likelihood of producing splenomegaly using the rating scale from

Experiment 1. The most important results involved the six items (6 to 11) lying between the category boundaries of the two conditions. In Condition A these items should be viewed as being members of the hemovedic virus type. Since within this group seven out of nine viruses caused splenomegaly, high ratings are to be expected. By contrast, the very same items should yield low ratings in Condition B. In this condition the six items belong to the allovedic viruses which cause the effect in only two out of nine cases.

Results and Discussion

The analyses are based on 32 participants (16 in Condition A and 16 in Condition B). All participants met the learning criterion. The most important analysis involved the test items between the two category boundaries (Items 6-11). The mean ratings for these six items clearly differed across the two category conditions A and B, $F(1,30)=14.7$, $p<.01$, $MSE=224.3$. The two contrasting mean values (averaged over the six items) were 65.9 versus 45.6. No significant differences were obtained for the (non-critical) test items that were not expected to differ across conditions. The average ratings of these items (with the items expected to yield low ratings being recoded to the corresponding high values) were 82.7 and 82.6. A 2 (category conditions) × 2 (critical vs. non-critical items) analysis of variance revealed a significant interaction, $F(1,30)=9.19$, $p<.01$, $MSE=178.3$.

Interestingly, the influence of the categories was strongest for the exemplars participants had seen in the causal learning phase (Phase 2). The mean ratings for these items (6, 7, 10, 11) were 70.8 versus 43.4 which was, of course, highly reliable, $F(1,30)=25.1$, $p<.01$, $MSE=238.3$. Thus, even though all participants had, for example, seen Item 6 as causing splenomegaly, they nevertheless gave this exemplar a lower rating in the test phase when it was categorized as an allovedic virus in Condition B than when it belonged to the hemovedic category in Condition A. (In this experiment the appearance of the items did not vary across phases.) By contrast, the two non-presented items 8 and 9 did not significantly differ across category conditions ($M=56.3$ vs. $M=50.0$). This rather surprising result seems to indicate that category level information is only used when it is actively encoded along with the item in the causal learning phase. Since these two items were not presented during this phase the relation between these items and the effects were apparently not encoded on the category level.

Again the final ratings showed that participants generally encoded the relationship between categories and the effect. They rated the causal efficacy of the two categories clearly different regardless of the location of the category boundary, $F(1,31)=80.5$,$p<.01$, $MSE=509.5$ ($M=74.8$ vs. $M=24.2$). All but five participants gave ratings consistent with this trend.

In summary, Experiment 2 confirms the results of Experiment 1 with family resemblance category structures. Despite the fact that all participants received identical cause-effect information in the causal learning phase, the ratings of the causal efficacy of the exemplars seen in this phase were moderated by the categories to which they belonged.

Experiment 3

The two previous experiments have shown that participants tended to use category knowledge that they had acquired in a previous learning context when learning about a new causal relation. Even though there was no reason to believe that the appearance-based categories learned in Phase 1 would provide useful classifications for the induction of the cause-effect relations in Phase 2, participants rather continued to use these categories than switch to a new conceptual scheme. Experiment 3 aimed at exploring the boundary conditions for this effect. In Experiments 1 and 2 category labels were used in Phase 1 (types of viruses) that seem to be plausible candidates for being causes of the target effect in Phase 2 (splenomegaly). Thus, despite the fact that the classification of the virus types was based on a rationale which was conceptually independent of the causal hypothesis in Phase 2 it may still be plausible to assume that viruses are generally useful categories for predicting health-related symptoms. It is possible that in a situation in which the semantic relatedness between categories and target effect is reduced fewer participants would continue to use the old categories.

To test the relevance of the semantic relation between categories and effect we designed cover stories that attempted to exclude all possible associations between categories and effects. Thus, our goal was to present participants with a learning situation in which there was no a priori reason to transfer the category knowledge from Phase 1 to the causal learning situation in Phase 2.

We used the same learning exemplars and the same learning procedure as in Experiment 1 but changed the cover stories. In Experiment 3 we introduced the items displayed in Table 1 as belonging to two types of objects, *Alpha-Objects* and *Beta-Objects*, that were distinguished on the basis of their appearance. In Phase 1 participants learned to classify the items into these two classes. In Phase 2, the causal learning phase, it was mentioned that these objects may be the causes of a novel *Effect*. No further semantic characterization of the kind of effect was given. In Phase 3 participants gave ratings of the likelihood that the test items caused this unknown effect. After these ratings we also required participants to assess the causal efficacy of the two contrasting categories, Alpha- and Beta-Objects.

Results and Discussion

The results are based on 48 participants (24 in the size condition and 24 in the brightness condition). Three further participants were excluded because they did not meet the learning criterion. Again the most interesting result involved the critical test items whose ratings should differ in case category level information was used. In this experiment the effect was again in the right direction but clearly weaker than in Experiment 1. The mean values of the averaged four critical items (two of them were recoded) were $M=59.6$ versus $M=49.6$ in the contrasting category conditions, $F(1,46)=3.92$, $p<.06$, $MSE=306.3$. As in the previous experiments the uncritical items whose ratings were not expected to differ across conditions yielded similar ratings ($M=64.3$ vs. $M=63.6$). The 2 (category conditions) × 2

(critical vs. non-critical items) analysis of variance did not show a significant interaction (p=.16) in this experiment. The ratings of the category-effect relations indicated that overall the differential relation of the two categories and the effect was learned, $F(1,47)=52.4$, $p<.01$, $MSE=595.2$. The mean ratings were 68.8 versus 32.7. However, a closer inspection of the data revealed that unlike in the previous experiments a considerable number of participants did not encode the differences on the category level. 14 out of the 48 participants gave equal ratings to the two categories. An analysis in which these cases were excluded showed that these 14 participants were mainly responsible for the decrease of the size of the effect for the critical items. The remaining 34 participants (14 in the size and 20 in the brightness condition) gave mean ratings of 43.6 versus 62.1 for the critical items, $F(1,32)=11.6$, $p<.01$, $MSE=245.3$, which closely corresponds to the results of Experiment 1. An analysis of variance that only included the data from these remaining participants yielded a significant interaction between category conditions, and type of items (critical vs. non-critical), $F(1,32)=7.55$, $p<.01$, $MSE=211.5$.

In summary, once again a considerable number of participants continued to use the old categorial scheme even though there was *no* semantic link between categories and the effect that suggested the usefulness of the categories for the causal context. However, we also found evidence that the semantic relatedness between categories and effect affects the likelihood of transfer. Unlike in the previous experiments a relatively large number of participants (ca 30%) seemed to have concluded that the category level information was not useful for the subsequent causal learning task, and therefore did not encode the relation between category level and effect.

Conclusions

Overall the three experiments provide clear evidence for the tendency to continue to use categories that have been acquired in previous learning contexts when learning about new causal relations. Identical learning experiences yielded different attributions of causal capacity depending on the categories that the learning exemplars were assigned to. This holds true even though there was no compelling reason that the old categories were useful, and in fact other possible category structures yielded stronger statistical relations between categories and the effect. In our view, these findings show that the relation between categories and causality is bi-directional. Categories not only reflect pre-existing knowledge of causal structures they also affect the acquisition of new causal knowledge. Like in scientific paradigms there is a tendency to continue to use old conceptual schemes at the potential cost of suboptimal predictability but with the computational gain of not having to use many categorization schemes in parallel. As in science there seem to be conditions, however, in which old paradigms tend to be abandoned. The results of Experiment 3 suggest that there is a tendency to acquire new knowledge from scratch when the semantic link between old categories and the new causal hypotheses is weak.

The inherent bi-directionality of the relation between categories and causality may, of course, extend to multiple steps in a dynamic process of theory revisions. Prior categories affect causal induction which in turn may create new causal categories that influence what kind of statistical structure new data exhibit. The present results show that the outcome of this dynamic process of theory development may be crucially dependent on how it started.

References

Barsalou, L. W. (1983). Ad-hoc categories. *Memory & Cognition, 11*, 211-227.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition*. Hillsdale, NJ: Erlbaum.

Clark, A., & Thornton, C. (1997). Trading spaces: Computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences, 20*, 57-90.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps. Analogy in creative thought*. Cambridge, MA: MIT Press.

Horwich, P. (Ed.)(1993). *World changes: Thomas Kuhn and the nature of science*. Cambridge, MA: MIT Press.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology, 32*, 49-96.

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*,1-54.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (Eds.)(1996). *The psychology of learning and motivation, Vol. 34: Causal learning*. San Diego: Academic Press.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 34: Causal learning*. San Diego: Academic Press.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124*, 181-206.