Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, 137, 137-153.

Cognition 137 (2015) 137–153

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories

Micah B. Goldwater [a,*], Dedre Gentner [b,1]

[a] University of Sydney, School of Psychology, Brennan MacCallum Building (A18), Camperdown, NSW 2006, Australia
[b] Northwestern University, Department of Psychology, 2029 Sheridan Road, Evanston, IL 60208, United States

ARTICLE INFO

Article history:
Received 31 March 2014
Revised 1 December 2014
Accepted 2 December 2014
Available online 29 January 2015

Keywords:
Analogy
Causal learning
Categorization
Education

ABSTRACT

This research studies a relatively unexplored aspect of expertise – the ability to detect causal relational patterns in multiple contexts – and demonstrates learning processes that foster this ability. Using the Ambiguous Sorting Task (AST), in which domain information competes with causal patterns, we previously found that science experts spontaneously noticed and sorted by causal patterns such as positive feedback, while novices sorted primarily by content domain. We investigated two kinds of learning experiences that we claim are needed to achieve high fluency in detecting key cross-domain patterns. We found that *direct explication* of example phenomena increased people's accuracy in depicting the examples, but did not increase sensitivity to the causal patterns in new examples. However, *analogical comparison* between parallel examples did lead to greater propensity to detect the causal patterns across diverse examples. Combining within-example explication with between-example alignment led to the greatest gains in generalized sensitivity to causal patterns.

© 2015 Elsevier B.V. All rights reserved.

An important and understudied aspect of expertise is the ability to spontaneously notice key relational patterns in the flow of experience. For example, the Swiss inventor George de Mestral developed the idea of Velcro when he and his dog returned from an Alpen hike with masses of burrs in their clothes and fur. During the tedious process of removing the burrs, he began to focus on their extraordinary clinging power. After examining the burrs with a microscope, he came up with the idea for a reversible fastener, with stiff hooks (like the burrs) on one side and soft loops (like fur or fabric) on the other side. This ability to see beyond the routine irritation of dealing with burrs to a valuable insight suggests a creative mind at work. But it also underlines the importance of a *prepared* mind. De Mestral was a trained engineer, working in the machine shop of an engineering company and pursuing his own inventions on the side. (He received a patent for a toy airplane at the age of 12). His amassed experience in how things work gave him a rich internal vocabulary with which to interpret causal patterns that on the surface bear little resemblance to his everyday work.

Our question here is how people acquire generalized sensitivity to key causal patterns. We focus on causal patterns because of the importance and pervasiveness of causality in human cogition (Ahn, Kim, Lassaline, & Dennis, 2000; Mackie, 1980; Sloman, 2005). Having an abstract understanding of causal structure allows for deep connections to be made across domains. For example, both the melting of polar icecaps and the growth of economic pricing bubbles are governed by a *positive feedback* causal structure. An understanding of causal structure is critical for explanation and prediction (e.g., Lombrozo & Carey, 2006; Sloman, 2005; Thagard, 1989). It influences category

* Corresponding author. Tel.: +61 2 9351 5453; fax: +61 2 9351 2603.
E-mail addresses: micah.goldwater@sydney.edu.au (M.B. Goldwater), genter@northwestern.edu (D. Gentner).
[1] Tel.: +1 (847) 467 2035; fax: +1 (847) 491 7859.

http://dx.doi.org/10.1016/j.cognition.2014.12.001
0010-0277/© 2015 Elsevier B.V. All rights reserved.

organization (Ahn et al., 2000; Rehder & Burnett, 2005; Sloman, Love, & Ahn, 1998) and is reflected in linguistic structure (Fillmore, 1978; Jackendoff, 1983; Kuehne & Forbus, 2002; McCawley, 1968; Wolff & Song, 2003). For this reason, causal knowledge and reasoning has been a focus of cognitive science from the outset (de Kleer & Brown, 1981; Forbus, 1984; Hayes, 1979). There has been extensive research on how causal knowledge is represented, using formalisms such as qualitative process models (Forbus, 1985; Forbus, Nielsen, & Faltings, 1991) or causal Bayesian networks (e.g., Gopnik et al., 2004; Pearl, 2000; Waldmann, Hagmayer, & Blaisdell, 2006).

Empirical work on causality has examined how people determine the causal structure of a particular domain or phenomenon. While people typically find it quite difficult to infer complex causal patterns purely from observation (e.g., Lagnado & Sloman, 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003), this research has revealed kinds of experiences that lead people to infer causal relationships within a domain. These include being exposed to particular kinds of statistical relations among variables (Cheng, 2000), through direct causal interventions (e.g., Hagmayer, Sloman, Lagnado, & Waldmann, 2007) or to evidence of causal mechanisms (Ahn & Kalish, 2000; Ahn, Kalish, Medin, & Gelman, 1995; Lagnado, Waldmann, Hagmayer, & Sloman, 2007; Rehder & Hastie, 2001), and engaging in self-explanation (Chi, de Leeuw, Chiu, & LaVancher, 1994; Lombrozo, 2010) or counter-factual reasoning (e.g., Harris, German, & Mills, 1996).

Our focus here is on different. Instead of examining how people learn and use the causal relations that govern a particular phenomenon, we ask how people learn abstract categories of causal systems that apply across domains and phenomena. In our prior work we developed a sorting task aimed at assessing people's propensity to notice key causal patterns amidst competing information (Rottman, Gentner, & Goldwater, 2012). In this task, the *Ambiguous Sorting Task* (AST), subjects are asked to sort descriptions of causal phenomena into categories. They are given an array of five 'seed cards' to use in sorting (as well as an "Other" category). Each of the five seed cards depicts a different causal system and a different content domain from the others. This means that subjects are free to sort either by causal structure or by content domain. Likewise, the phenomena descriptions they are asked to sort vary both in their content domain (biology, economics, etc.) and in their causal structure (positive feedback, causal chain, etc.). (see Table 1 for the entire list.) The idea here is that people can achieve a successful sort simply by attending to the relatively obvious domain-level commonalities; thus, there is no need to seek some other sorting principle. However, if people spontaneously notice the causal commonalities, they may choose to use these instead.

To discover whether fluent knowledge of abstract causal patterns varies with expertise, we gave the AST to advanced physical science students and to social science students. There were two results of interest. First, students in social science and economics sorted primarily by content domain – evidence that the causal patterns were not obvious even to college students. Second, advanced physical science students – many of whom had taken courses in multiple science disciplines – sorted primarily by causal system – evidence that fluent perception of causal patterns increases with expertise (Rottman et al., 2012; and see Chi, Feltovich, & Glaser, 1981 for a similar result). To analogize to the de Mestral example, there was a shift from focusing on the domain level (removing burrs) to focusing on a cross-domain abstract principle (achieving an adhesive connection).

Our goal in the present research is to understand how people come to be sensitive to abstract causal patterns. We hypothesize that at least two kinds of learning experiences are needed for forming abstract causal categories. The first is experience and instruction on particular causal phenomena, as in much of the prior research on causal learning. Learning the causal structure of particular cases is important, but we suggest that by itself it is not enough. The second important contributor is a way to form abstract causal representations that apply across domains. We suggest that analogical comparison of cases in which the same causal system applies can achieve such abstractions.

These two contributors operate in different ways. The first contributor, encountering causal explanations for specific phenomena, seems likely to improve causal understanding of specific phenomena. But by itself it is unlikely to promote general causal abstractions. Rather, we hypothesize that comparing analogous phenomena from different domains is critical in promoting abstraction of the common causal structure. There is considerable prior evidence consistent with the idea that analogical comparison highlights common relations through a process of structural alignment (Falkenhainer, Forbus, & Gentner, 1989; Gentner & Markman, 1997), and that such an alignment permits learners to notice the common structure, which can then be applied more broadly. The idea that analogical comparison promotes transfer from specific learning contexts has received support from research with adults (Gick & Holyoak, 1983; Goldwater & Markman, 2011; Kurtz, Boukrina, & Gentner, 2013) and children (Christie & Gentner, 2010; Gentner, Anggoro, & Klibanoff, 2011) and by studies in educational contexts (Gentner, Loewenstein, & Thompson, 2003; Klahr & Chen, 2011; Rittle-Johnson & Star, 2009).

However, although these studies show that analogical comparison can improve transfer, in general it is not clear that it does so via abstracting the common structure. Analogical comparison can also improve the individual case representations, and this in itself could improve transfer (e.g. Gary, Wood, & Pillinger, 2012). Because our study includes a separate (and prior) assessment of the accuracy of the case representations, as described below, we can test whether there are effects of comparison on abstraction over and above its effects on representation accuracy.

Thus, we hypothesize that both causal explication and structural alignment will lead to increased sensitivity to causal structure, but for different reasons. Causal explication of the training examples will lead to better representation of the causal structure of the individual examples. Structural alignment of training examples will lead to abstraction of the causal pattern, and therefore to increased ability to perceive that pattern in other phenomena. Based on this reasoning, a further prediction is that

Table 1
Definitions and examples of five causal system categories from Rottman et al. (2012). Negative feedback was not used in the current experiments.

Causal system	Definition	Example
Common effect	Many factors cause one effect	*Economics*: The Cost-of-Living Index is a measure of the cost of basic necessities in a particular city or country. Cost-of-Living is influenced by factors such as inflation, the price of consumer goods, and the local and federal tax rate. Because some cities are more expensive than others, companies will use the Cost-of-Living in part to determine salaries
Common cause	One factor causes multiple effects	*Biology*: In the human body, an allergic reaction can trigger inflammation, rash, asthmatic attack, and typical cold-like symptoms. This is why allergies can be so annoying! The use of an anti-histamine medication can significantly improve many of these symptoms
Causal chain	One causal factor X leads to an effect Y which in turns causes effect Z and so on	*Mechanical engineering*: The powertrain in cars is the system that transfers energy generated by the motor into turning the wheels. The motor turns gears in the gearbox which rotate the driveshaft which rotates the wheels. When a car is in neutral, no gears are engaged, and thus no power is transferred from the motor
Positive feedback	The output of a causal system (e.g., wherein X increases Y which increases Z) is then fed back into the system and further increases the original factor X such that the increasing of each factor goes on indefinitely	*Environmental science*: In the process of global warming, as the temperature of the earth rises, polar ice begins to melt. Water absorbs more heat from sunlight than ice. Consequently, as ice is turned into water, the temperature of the earth begins to rise even faster which in turn leads to increased ice-melt
Negative feedback	The output of a causal system is then fed back into the system, and decreases the original factor such that the system reaches equilibrium	*Electrical engineering*: A thermostat works by measuring temperature and turning on or off a furnace or air conditioner to reach a desired temperature. If the temperature is too cold, the thermostat will turn on the furnace until it becomes warm enough. Likewise, the thermostat on an air conditioner turns on when the house is too warm

combining the two processes will be more effective than either alone.

In the current research, we examined four causal system categories: *common cause, common effect, causal chain,* and *positive feedback*[2] (see Table 1). The basic idea in this study was to give people training on examples of the four causal systems discussed above and then compare their performance on a version of the Ambiguous Sorting Task (AST) with that of people who had received no training. In the training, we independently varied two kinds of experience: whether people received causal explications of the training examples, and whether they were encouraged to compare and align pairs of examples. This led to four between-subjects training conditions: both explication and alignment; explication but not alignment; alignment but no explication; and neither explication nor alignment. The training set for all four training conditions consisted of eight training examples – two each from each of the four causal system categories (positive feedback, causal chain, common cause, and common effect). One example was always from electrical engineering and the other, from political history.

We manipulated Explication as follows. In the Explication condition, each of the eight training examples was accompanied by a statement of the abstract causal schema (such as *positive feedback*) and a detailed explication of how the component events of the phenomenon fit into the roles of the causal schema. We contrasted this with simply receiving the label for the causal schema. (See Table 2 for an example.) After each training example, subjects were asked to draw a diagram depicting the causal structure of the phenomenon. This allowed us to gauge their understanding of the examples prior to engaging in the Alignment task.

These groups were now further divided: half received Alignment instructions, and half did not (see Fig. 1). Those in the Alignment condition were asked to compare pairs of training examples that had the same causal system but were from different content domains. To encourage a deep structural alignment, we adopted a task used by Kurtz, Miao, and Gentner (2001). Subjects first wrote what the two phenomena had in common, and then were asked to state the correspondences they perceived. Importantly, the Alignment manipulation occurred after subjects had read and diagrammed the eight training examples. Thus, half the subjects had received explications along with the examples and half had received only the examples.

In addition to the four training conditions, there was a No Training control condition, in which subjects simply performed the sorting task. All five groups – trained and untrained – performed the same AST at the end. Thus, the design of Experiment 1 was 2 (Explication vs. Label Only) × 2 (Alignment vs. No Alignment) + 1 (No Training) (see Fig. 1), with all variables manipulated between subjects.

The key dependent measure was performance on a version of the Rottman et al. (2012) Ambiguous Sorting Task.[3]

[2] We omitted negative feedback for reasons of design.

[3] Because one domain (electrical engineering) was used in the training materials, there were only four seed cards instead of five as in the Rottman et al. study.

Table 2
Example of a training phenomenon description (Electrical Engineering, Positive Feedback). In the Label condition, shown at top, subjects were given only the causal system label. In the Explication condition (bottom), subjects saw the causal system label, a statement of the causal schema and a description of how the roles of the schema are bound to parts of the phenomenon.

Audio feedback occurs when a microphone is placed too close to a speaker. Any noise that the microphone picks up is sent to an amplifier and then played through the speaker at a higher volume. The microphone then picks up the louder noise and sends it back to the amplifier and then onto speakers even louder. This can lead to a very loud squealing
This is an example of a *positive feedback* system
Audio feedback occurs when a microphone is placed too close to a speaker. Any noise that the microphone picks up is sent to an amplifier and then played through the speaker at a higher volume. The microphone then picks up the louder noise and sends it back to the amplifier and then onto speakers even louder. This can lead to a very loud squealing
This is an example of a *positive feedback* system. In a positive feedback system each causal factor increases the next causal factor and cycles back to increase an earlier causal factor. So, each causal factor both increases the others and is increased by them. In this example, the microphone is the first causal factor and it sends sound to the amplifier, a second causal factor, which amplifies and sends the sound to the speaker, resulting in the sound being played at an increased volume. But the speaker is also a cause because it sends this louder sound back to the microphone. The microphone then sends it to the amplifier, and so on

As in the original AST described above, subjects sorted a set of unlabeled examples into categories defined by a set of seed cards that differed both in causal structure and in content domain. Because this task allows people to sort either by causal patterns or by content domain (a highly salient dimension), it serves as an indication of people's sensitivity to causal structure. The key question is whether our training will lead to greater sensitivity to causal structure, and if so, how. To understand the specific roles of explication and alignment, we included a further dependent measure: performance in sketching the causal diagrams. Because these are drawn after subjects have read the training examples, but before the alignment manipulation, they allow us to ask whether receiving explication of the causal schemas (vs. causal labels only) increases subjects' accuracy in depicting the causal structure of the training examples, how the accuracy of the individual examples influences performance on the sorting task, and whether alignment improves causal sorting even for participants with accurate causal models.

To review, for Experiment 1a we hypothesize that both explication and structural alignment will lead to increased causal sorting, as compared to the No Training condition, but for different reasons. Causal explication of the training examples will lead to better representation of the causal structure of the examples, as assessed by the accuracy of the causal diagrams; and structural alignment of training examples will lead to abstraction of the causal pattern, and therefore to increased ability to perceive that pattern in other phenomena – over and above having correct causal understanding of the individual examples. Experiment 1b then evaluates whether the diagramming task itself influences the results.[4]

[4] We note that this method leaves open the possibility that our comparison task simply acts to alert participants to the relevance of causal information in this study – that is, it acts as a kind of demand characteristic. In addition, Experiment 1a does not rule out that simply revisiting the examples (without structural alignment) aids learning. We address these possibilities in Experiment 2.

1. Experiment 1

1.1. Methods

1.1.1. Subjects

Ninety-seven Northwestern University undergraduates participated for course credit: 20 in the No Training group, 20 in the Explication & Alignment condition, and 19 in each of the other three training conditions. The causal diagrams of two subjects were lost, so are excluded from analyses concerning the diagrams. This sample size was based on that used by Rottman et al. (2012). That study compared 12 physical science majors to 32 physical science novices and was able to successfully resolve effects. Thus, we used the average of their group numbers for each of our conditions.

1.1.2. Materials

Ambiguous Sorting Task. The materials for the AST were 16 example phenomena, composing a matrix of four causal systems (positive feedback, causal chain, common cause, and common effect) crossed with four content domains[5] (biology, economics, environmental science, and mechanical engineering) (see Fig. 2). These examples were taken from Rottman et al.'s (2012) study and their development and calibration is described in full there. By design, examples from the same domain had more lexical similarity than examples from the same causal system, and this was confirmed by their average pairwise LSA similarity scores.

Training. Eight additional examples were used in training – four from electrical engineering (one illustrating each of the causal systems) and likewise four from political history. The four electrical engineering examples were taken from Rottman et al. (2012); the four examples from political history were developed for this study.

1.1.3. Procedure

Fig. 1 displays the design and flow of the experimental procedure. In the No Training condition, subjects simply

[5] We used a 4×4 matrix of Causal system $\times$ Content domain, instead of a 5×5 matrix as in Rottman et al.'s study, and 16 cards to be sorted instead of 20. This was done so that we could use one of the original five domains (Electrical Engineering) as a training domain.

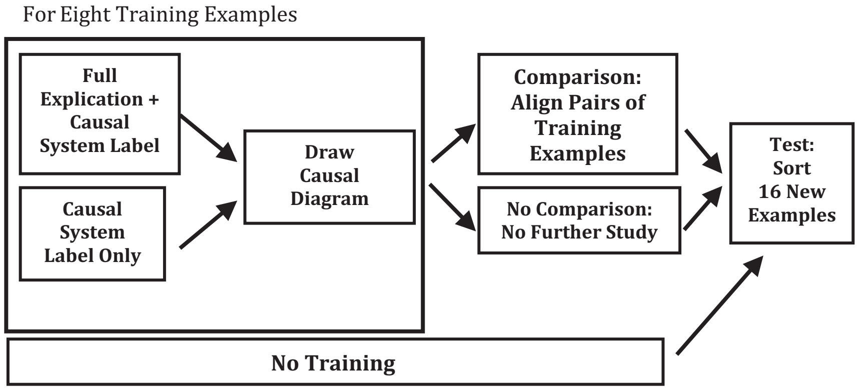

Fig. 1. Experiment 1a: design and experimental flow.

performed the Ambiguous Sorting Task. In the four training conditions, subjects went through training packets prior to performing the AST.

Training. The four training groups were first given the instructions "You will read a series of descriptions, and you will sketch a *causal graph* for each one. Causal graphs depict the cause and effect relationships in the description. . . ." They were given two simple examples: "Smoking increases the risk of lung cancer" and "Fluoride prevents cavities," each with the corresponding causal diagram (see Fig. 3). They then received the training packet: eight brief descriptions of phenomena – four from electrical engineering followed by four from political history. Each training example was given on its own page. Below the example description, the causal system label was given: e.g., "This is an example of a positive feedback system." All four training groups received these labels. In addition, those in Explication condition received an explication of the example that showed how the causal schema fit the particular example and how the example links to the causal roles in the schema (see Table 2). For all subjects, the bottom half of each page was blank, leaving space for the causal diagram. For the two No Alignment groups, these eight pages completed the training packet.

For the two Alignment groups, these eight pages were followed by four pages containing pairs of (previously seen) example phenomena – one from electrical engineering and political history – for subjects to align. The two examples shared a common causal system, though this was not mentioned. Each page asked what the two phenomena had in common, with a blank space below for the response. In addition, subjects were asked to complete a two-column table. One column listed the components of one example; the subject's task was to fill in the corresponding components of the other example.

Sorting. The procedure for the AST was adapted from that of Rottman et al. (2012). Each of the 16 phenomena to be sorted was printed on a file card. On the table before the subject there were four columns headed by the seed cards (the four yellow cells in Fig. 2), plus one column labeled "Other." Subjects were asked first to read the seed cards, and then to sort the other 12 cards into the columns "based on how well the description on the card goes with the initial card" heading each column. They could also place a card into the "Other" column if they did not think it fitted with any of the seed cards.

Because our goal was to pit the two sorting strategies (content domain and causal system) against one another, the four seed cards each differed from one another both in content domain and in causal category. Thus, as in Rottman et al. (2012), the columns could be taken to represent four different domains or four different causal systems, or some mixture of the two. See Table 3 for how two example seed cards can be matched with other cards based on either content domain or causal system. The 12 cards to be sorted were arranged in a semi-random order, with half the subjects receiving the reverse order. There was no time limit; subjects were told that they could work with the cards in any order and could rearrange previously sorted cards.

After sorting all 12 cards, subjects were asked to consider alternate sorting strategies and then to re-sort the examples. This was done to allow subjects to display sensitivity to both the causal similarities and the domain similarities. In particular, we were concerned that the deliberate ambiguity of the sorting task could result in underestimating causal knowledge. For example, subjects might recognize both causal similarities and domain similarities, but interpret the task to be about domain similarities. The second sort allowed subjects an opportunity to display both kinds of knowledge.

1.2. Results

We first consider the data from the first sorting. Each of the 12 cards (one "sorting") was coded according to whether it matched the seed card in its column by content domain or by causal system, or not at all. Overall, the average proportion of cards sorted into the "other" column or into a column that did not match the seed card either in domain or in causal system was .27. This did not differ reliably among the five groups (including the No Training group).

Error analysis. To further understand participants' strategies, we asked whether some causal categories were particularly confusable with one another. To do so, we analyzed their miss-sorts: cases in which a card was sorted in a column that differed from it both in causal category

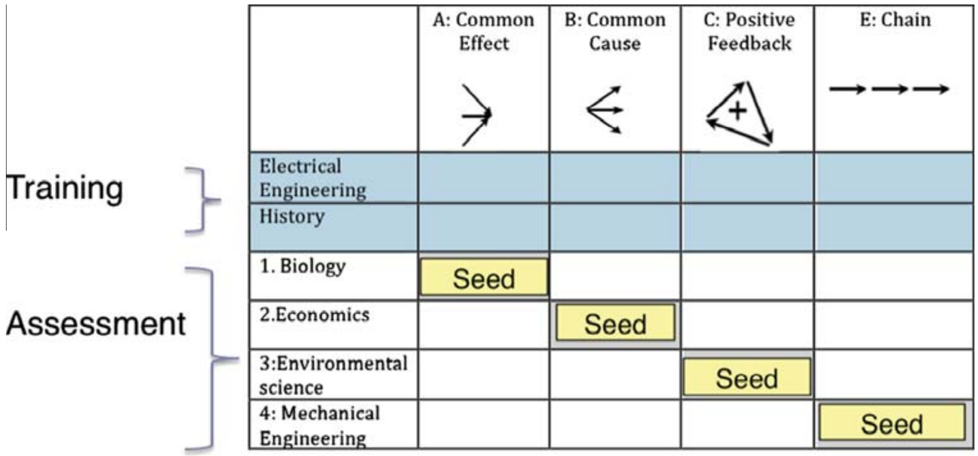

Fig. 2. Matrix of materials used in all studies, showing four causal systems × six domains: two domains (at top) used for training plus four domains (below) used in the sorting task. Seed cards are marked as such.

and in domain. There were six possible kinds of causal system miss-sorts (disregarding which card was the seed and which was the exemplar to be sorted). As shown in Table 4, the frequencies were not distributed randomly, $\chi^2(5) = 52.9$, $p < .001$; over half the miss-sorts were grouping a causal chain phenomenon with either a positive feedback or a common effect phenomenon.

We next compared the rate of causal sorting between the training conditions and the No Training control. The Explication/Alignment ($M = .62$), Explication/No-Alignment ($M = .37$), and Label-Only/Alignment ($M = .37$) conditions all showed more causal sorting than the No Training group ($M = .14$) all t's > 2.5, all p's < .05, all d's > .90.[6] The Label-Only/No-Alignment ($M = .29$) condition did not differ significantly from the No Training group $t(37) = 1.84$, $p = .07$, $d = .60$.

We now turn to the key findings: the rate of causal sorting across training conditions (Fig. 4). A 2 (Explication vs. Label-Only) × 2 (Alignment vs. No-Alignment) independent samples ANOVA revealed main effects of Explication, $f(1,73) = 5.26$, $p < .05$, $\eta^2_p = .067$, and Alignment, $f(1,73) = 5.0$, $p < .05$, $\eta^2_p = .064$, but no interaction, $f(1,73) = 1.15$, $p > .25$, $\eta^2_p = .016$. Planned comparison t-tests reveal that, as predicted, the Explication/Alignment group showed the highest rate of causal sorting, significantly greater than that of the other three training groups, all t's > 2.25, all p's < .05, all d's > .70.[7] Neither Explication nor Alignment by itself gave a reliable boost above the Label-Only/No-Alignment training condition (though each reliably outperformed the No Training control, as noted above).

To further illuminate the pattern of sorting, we display a scatterplot for each condition, representing all subjects by their proportions of causal and domain sorts along the Y and X axes, respectively (Fig. 5). These scatter plots display the obvious point that causal and domain sorting are negatively correlated. More importantly, they show to what degree subjects sorted by domain, by causal system, or by a mixed approach. They also reveal the degree to which a subject conformed to neither strategy (due to error sorts – that is, miss-sorts and "other" column sorts). The proportion of error sorts for a given point is 1 minus the sum of its X and Y values; thus, the closer a point is to the graphs' origin, the greater the number of errors in that subject's sorts.

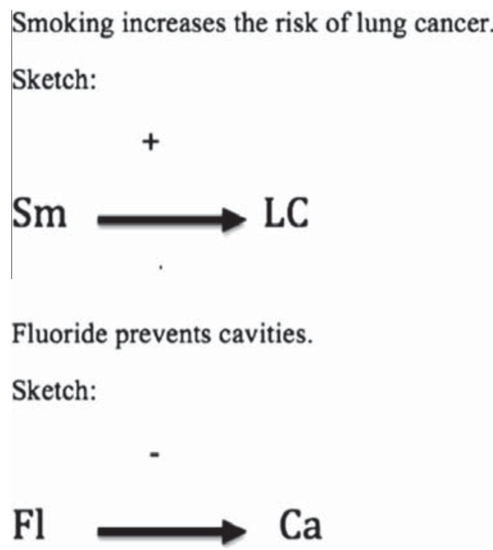

Fig. 3. Example diagrams used in causal sketch instructions.

Fig. 5 reveals there were very few hybrid strategies: out of 97 subjects, only 3 sorted over 30% by causal system and 30% by domain. That is, subjects tended to have either a causal or domain focus, and to differ in how successfully they implemented this focus (as shown by their points' distance from the origin). For example, comparing Explanation/Align with the other conditions in Experiment 1a, we see more points clustered near the causal axis, suggesting few domain sorters; and those points are higher on the causal axis, suggesting fewer errors.

[6] Compared to No Training: Explication/Alignment $t(38) = 5.15$ $d = 1.69$; Label-Only/Alignment $t(37) = 2.9$, $d = .95$; Explication/No-Alignment $t(37) = 2.86$, $d = .94$.

[7] Compared to Explication/Alignment: Label-Only/Alignment $t(37) = 2.23$, $d = .73$; Explication/No-Alignment $t(37) = 2.32$, $d = .76$; Label-Only/No-Alignment $t(37) = 3.1$, $d = 1.02$.

Table 3
Examples of two seed cards. Below each seed are shown one domain match and one causal system match.

Seed Card: (Environmental Science, Positive Feedback)
In the process of global warming, as the temperature of the earth rises, polar ice begins to melt. Water absorbs more heat from sunlight than ice. Consequently, as ice is turned into water, the temperature of the earth begins to rise even faster which in turn leads to increased ice-melt

Same-Domain Match	Same Causal System Match
Same-Domain Match: (Environmental Sci., Common Effect) Many processes are responsible for the level of CO_2 in the atmosphere such as fossil fuel burning, plant photosynthesis, and CO_2 absorption in oceans. When making public policy decisions about environmental issues, we must keep in mind the many complex factors that influence the level of CO_2 in the atmosphere.=	*Same Causal System Match: (Mechanical Engineering, Feedback)* In a nuclear reactor, water is used to keep the reaction under control. In the Three Mile Island incident, the water level was accidentally reduced while the nuclear reaction was already in progress. This allowed the reaction to increase and evaporate more water than normal. As the water level decreased at a faster and faster rate, the reaction increased causing a partial meltdown

Seed Card: (Economics, Common Cause)
The unemployment rate has many implications for society. It influences crime and suicide rates, general health conditions of the public, and the Gross Domestic Product Index. Because of its significant effects, many politicians focus on decreasing the unemployment rate to improve many aspects of the economy

Same-Domain Match	Same Causal System Match
Same-Domain Match: (Economics, Causal Chain) The US economy is heavily dependent on the availability of petroleum. Petroleum is used particularly for transportation of consumer goods by truck or airplane. Consequently, higher oil prices can significantly increase the price of transportation which is reflected in the price of consumer goods	*Same Causal System Match: (Mechanical Eng., Common Cause)* All water piping in a home comes out of one pipe connected to the city water grid. If this main pipe is damaged, the water throughout your house will be affected. This includes not only sinks and showers, but also toilets, washing machines, and dishwashers

Table 4
The sums (and proportions) of different kinds of mis-sorted phenomena (i.e., sorts that matched neither the causal system nor the domain of the seed card) for each experiment. Each category is defined by the two causal systems of the seed card and exemplar card sorted together, disregarding which card was the seed and which was the card to be sorted. The proportions are with respect to the number of miss-sorts for each experiment separately, except for the rightmost column, which displays the sums and proportions based on the aggregate of all experiments.

Exp. 1a, $N = 97$	Exp. 1b, $N = 20$	Exp2, $N = 63$	Total, $N = 180$
Causal Chain/Positive Feedback			
60(.32)	14(.34)	34(.31)	108(.32)
Causal Chain/Common Effect			
45(.24)	4(.10)	24(.22)	73(.22)
Causal Chain/Common Cause			
32(.17)	15(.37)	23(.21)	70(.21)
Common Effect/Common Cause			
21(.11)	7(.17)	11(.10)	39(.12)
Common Effect/Positive Feedback			
17(.09)	0(.00)	5(.05)	22(.07)
Common Cause/Positive Feedback			
13(.07)	1(.02)	12(.11)	26(.08)

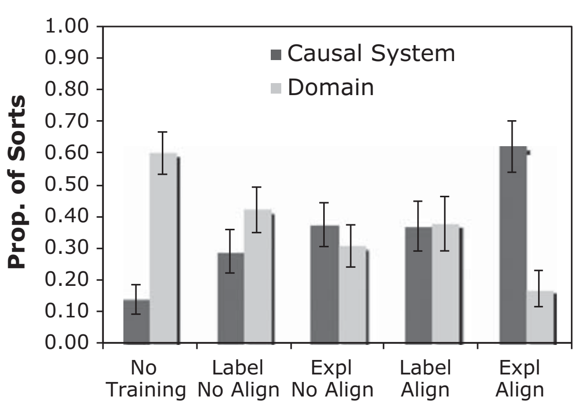

Fig. 4. Mean and standard error of proportions of causal and domain sorting in Experiment 1a for the first sort.

We next compared the first sorting to the second sorting. The second sorting was included to allow subjects to demonstrate causal sensitivity even if their first sorting had been by content domain. However, causal sorting was reliably less frequent in the second sort ($M = .14$) than in the first ($M = .36$), $t(96) = 4.92$, $p < .001$, $d = .50$, and domain sorting was correspondingly more frequent (2nd $M = .61$; 1st $M = .38$), $t(96) = 3.89$, $p < .001$, $d = .39$ (see Fig. 6). This pattern held for all training conditions; in the No Training condition, the rates of causal sorting did not differ between first ($M = .14$) and second ($M = .15$) sort, $t(19) = 0.18$). It appears that subjects who perceived the causal matches generally used them in the first sort; and these subjects were typically capable of using a domain strategy on the second sort. In contrast, those who used domain sorting on their first sort were unlikely to adopt causal sorting on the second sort.

Assessing the accuracy of causal representations. To test the hypothesis that causal explication would improve the representation of individual training examples, we assessed the accuracy of the subjects' diagrams. The diagrams were scored as correct or incorrect, according to whether they represented all and only the causal relations in the phenomenon description (see Fig. 7 for examples). All diagrams were scored by a research assistant. As a reliability check, the first author scored the diagrams from 25 of the 75 subjects (200 of the 600 diagrams). (Both raters were blind to condition.) There was disagreement on only 8 of these 200 diagrams; these disagreements were resolved by discussion.

Just over half of the subjects (41 out of 75) drew all eight diagrams correctly. As expected, explication improved subjects' accuracy: subjects in the Explication condition ($M = 7.37$ out of 8) drew more accurate diagrams than those in the Label-Only condition ($M = 6.54$), $t(73) = 2.23$, $p < .05$, $d = .55$. Because the means were close to ceiling we supplemented this analysis by using non-parametric bootstrapping to estimate their 95% confidence intervals (free from assumptions of normality). That is, we simulated sampling distributions by sampling with replacement 1000 times from each condition (creating distributions of 1000 means). From these simulated sampling distributions the 95% CI of the Explication condition was 6.79–7.79 (above the mean of 6.54 for the Label condition), while the 95% CI of the Label condition was 6.0–7.05 (below the mean of 7.37 for the Explication condition).

We next asked whether we could use the results of the causal-diagram task to clarify the effects of alignment. The idea is that if (as claimed) one effect of structural alignment is to create more abstract relational structures, then it should be most efficacious when subjects have accurate representations of both analogs. To test this, we asked whether structural alignment led to greater causal sorting for subjects who produced entirely accurate diagrams of the training cases than for those with some incorrect diagrams. A 2 (All-Correct Diagrams vs. Some Incorrect Diagrams) $\times$ 2 (Alignment vs. No-Alignment) independent samples ANOVA revealed main effects of Diagram Accuracy $f(1,71) = 12.63$, $p < .05$, $\eta^2_p = .151$, and Alignment $f(1,71) = 4.32$, $p < .05$, $\eta^2_p = .057$ but no interaction, $f(1,71) = 1.04$, $p > .30$, $\eta^2_p = .014$ (see Fig. 8). Planned comparison t-tests showed that for subjects with 100% diagram accuracy, Alignment ($M = .65$ causal sorts) led to significantly more causal sorting than No-Alignment ($M = .43$), $t(39) = 2.19$, $p < .05$, $d = .70$ – evidence for 'added value' of the alignment process even for those with perfectly correct understanding of the training examples. This effect did not hold for those with one or more incorrect diagrams (Alignment $M = .32$; No-Alignment $M = .24$), $t(32) = 0.78$, $p > .40$, $d = .27$.

To further clarify the effects of explication and alignment, we conducted a linear regression analysis with three categorical predictor variables: All Correct Diagrams vs. Some Incorrect,[8] Explication vs. Label-Only, and Alignment vs. No-Alignment to examine two critical hypotheses (1) The effect of explication was chiefly to improve the accuracy of the causal models for individual examples; and (2) Even when subjects were 100% accurate, alignment still improved performance. Diagram Accuracy (Beta = .334, $t = 2.91$, $p < .01$) and Alignment (Beta = .225, $t = 2.12$, $p < .05$) were shown to be reliable predictors of causal sorting, while Explication was not (Beta = .122, $t = 1.06$, $p > .25$). This analysis demonstrates that once perfect diagram accuracy is accounted for, there is no significant effect of explication on causal sorting; but alignment still adds value. This pattern suggests that the effect of explication occurs through promoting accurate understanding of the causal structure of the training examples. Consistent with this idea, the Sobel Test of mediation confirms that perfect diagram accuracy significantly mediates the effect of explication on causal sorting, Sobel statistic = 2.50, $p < .05$. Thus, the effect of Explication is mediated by increased accuracy of representation (as assessed by diagram accuracy).

Overall, the results provide support for our three main claims. First, explication of the causal schema improves people's representation of the training examples. Second, structural alignment of the training examples fosters abstraction of their common causal schema. Third, because *which* abstractions are produced by structural alignment depends on the subject's representations of the examples, the combination of explication and alignment leads to the greatest gain in causal insight.

However, despite these encouraging results, there are some issues that need to be considered before embracing the conclusions. First, in the analyses so far, we have treated the causal diagram task simply as a measure of the subjects' knowledge. However, the diagramming task could have contributed to learning the causal structures, by inducing subjects to explicitly represent them. To examine this possibility, and to replicate the central finding that combining Explication with Alignment leads to high rates of causal transfer, we re-ran that condition without causal diagramming.

2. Experiment 1b

2.1. Methods

2.1.1. Subjects

Twenty Northwestern University undergraduates participated for course credit.

2.1.2. Materials & procedure

The materials and procedure were identical to those of the Explication & Alignment condition of Experiment 1a, except that subjects did not draw causal diagrams and only sorted once.

2.2. Results

The results bear out the prior findings. In the Explication/Alignment condition, even without diagramming, subjects produced .58 causal sorts and .2 domain sorts – comparable to the .62 causal sorts and .17 domain sorts elicited by the same condition with diagramming from Experiment 1a: Comparing causal sorts, $t(38) = .41$, $p > .65$, $d = .13$. Fig. 9 shows the scatterplot of these results. Thus, there is no evidence that causal diagramming was crucial to improving the rate of causal sorting.

3. Experiment 2

The results so far bear out our hypothesis that structural alignment reveals common abstract relational

[8] We assessed accuracy as a binary variable (rather than a continuous variable) in order to address the question raised earlier, of whether alignment acts to abstract an already accurate relational representation.

patterns such as positive feedback and common-cause. This is consistent with our claim that structural alignment renders the shared causal structures more salient and promotes forming more abstract causal structures. However, we must consider another possibility. It could be that the effect of alignment is simply to alert subjects to the relevance of causal relations in this study. That is, the alignment experience could simply serve to shift subjects' attention toward causality, rather than to invite new causal insight (see Spalding & Ross, 1994). Even worse, perhaps the alignment experience just acts as a demand characteristic, implicitly telling subjects to focus on causal similarities rather than domain similarities.

To clarify the nature of the alignment advantage, in Experiment 2 we elicited the strategies subjects used in the sorting task. For example, if subjects in the alignment condition are more likely to refer to causality in their justifications, this will leave open the possibility that the alignment advantage in causal sorting can be explained by a strategy shift toward causal structure. Of course, it is possible that alignment will serve both to shift subjects' attention to causality and to foster causal abstractions (as revealed in the sorting data). In this case, we would see both an emphasis on causality in their verbal statements and an improvement in sorting ability, relative to other conditions.

A further goal of Experiment 2 was to test another possible explanation of the alignment advantage. Subjects in the Alignment condition saw each of the eight phenomena twice, whereas those in the No-Alignment condition saw them only once. Thus the Alignment effects could have stemmed from greater exposure, rather than from comparison processes *per se*. To test this possibility, in Experiment 2, we ran another condition in which subjects also revisited the training examples but were simply asked to "list the important factors" for each example phenomenon individually, rather than to compare examples. Thus, Experiment 2 had three conditions: Explication/Alignment, Explication/No-Alignment, and Explication/List-Factors. The first and second conditions were identical to the conditions of the same name in Experiment 1a.

3.1. Methods

3.1.1. Subjects

Sixty-three Northwestern University undergraduates participated in exchange for course credit: 19 in Explication/No-Alignment, 21 in Explication/Alignment, and 23 in Explication/List-Factors.

3.1.2. Materials and procedure

The materials were identical to those in Experiment 1. All subjects first read the eight training examples and their explications (and diagrammed each in turn). For the Explication/Alignment and Explication/No-Alignment conditions, the remaining procedures were as in Experiment 1, with one exception. After both the first and the second sorting, subjects were asked to write down the strategies they had used in sorting the descriptions. The instructions read, "There are many ways to sort these descriptions. Could you please write out on this page an explanation for why you sorted them the way you did?"

In the Explication/List-Factors condition, after reading and diagramming the first four training examples (from electrical engineering), subjects were instructed to write out descriptions of the examples on four additional pages. They were encouraged to look back at the descriptions if needed. To guide their description, each page read [for example] "Consider the description of Audio Feedback. What were two factors in that situation?" Subjects responded on the same page. They then repeated this process for the four political history examples, first reading descriptions with explications and diagramming, and then listing factors. Then they performed the AST and reported their sorting strategies as in the two other conditions.

3.2. Results

As shown in Fig. 10a, Explication/Alignment ($M = .65$) elicited more causal sorts than either Explication/No-Alignment ($M = .37$) or Explication/List-Factors ($M = .37$), $f(2,\ 60) = 4.44$, $p < .05$, $\eta^2{}_p = .13$; planned comparison *t*-tests: Explication/Alignment vs. Explication/No-Alignment $t(38) = 2.48$, $p < .05$, $d = .81$; Explication/Alignment vs. Explication/List-Factors $t(42) = 2.63$, $p < .05$, $d = .81$. Explication/No-Alignment elicited the same proportion of causal sorts as did Explication/List-Factors. Thus the effect of alignment cannot be reduced to simply receiving more experience with the examples.[9] These results again affirm that combining explication of individual cases with alignment across cases produces the greatest gain in causal insight. As in Experiment 1a, there were no gains in causal sorting during the second sorting attempt (see Fig. 10b).

Fig. 11 shows the scatterplots for the first sort. As in prior studies, hybrid strategies were quite rare, with only a single subject sorting more than 30% via both domain and causal system. As in Experiment 1a, the Explication/Alignment condition contrasts with the two non-alignment conditions in showing both a *greater focus* on causal commonalities over domain commonalities (as shown by more points near the causal axis) and *better* causal sorting (as shown by more points higher on the causal axis). In the next set of analyses, we probe further into these differences by analyzing the strategies the subjects reported using.

Strategies reported. Subjects were scored as having a causal strategy if they (1) used the term "cause" or "effect" (in noun, verb, or adjective form) or (2) used a causal system label such as "positive feedback system." Subjects were scored as having a domain strategy if they explicitly labeled the domains, either with single names such as "electricity," or "the body" or with a longer description, such as "all relate to processes of the human body." See Table 5 for typical examples of each kind of strategy.

With this coding scheme, 58 of 63 responses were unambiguously classified. Four of the 63 responses contained both kinds of language, but were classified as

[9] In an unreported experiment, we also ran a condition combining Label-Only with List-Factors. List-Factors had the same effect in this experiment as reported here.

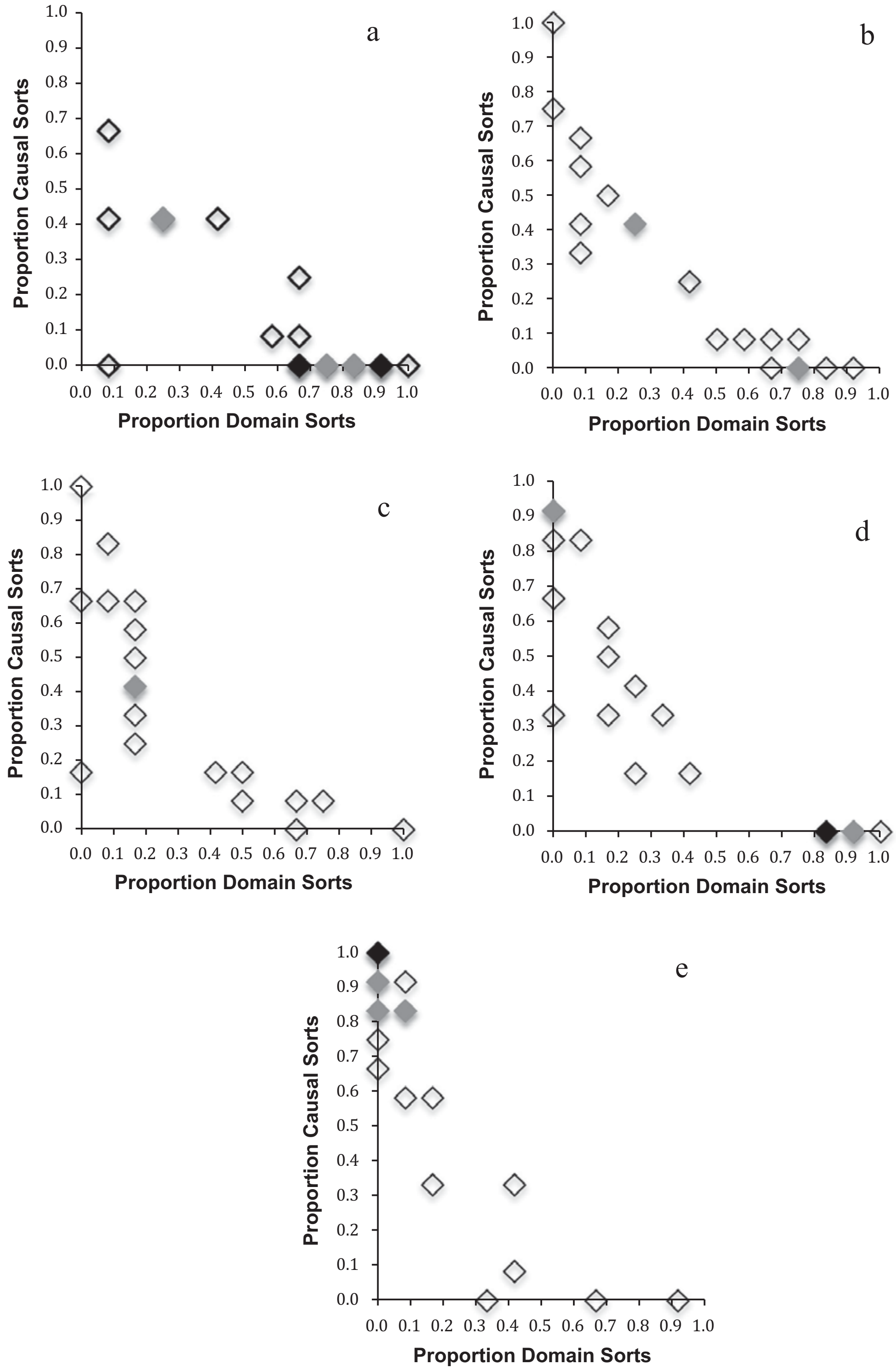

Fig. 5. Experiment 1a: Scatterplots showing rates of causal sorting (*Y* axis) and domain sorting (*X* axis) for each condition of Experiment 1a. White diamonds represents a single subject's score; grey diamonds represent two subjects' scores; black diamonds' represent three subjects' scores. The conditions are (a) No training; (b) Label/No-Alignment; (c) Explication/No-Alignment; (d) Label/Alignment; (e) Explication/Alignment.

domain strategies because the causal language was used only to connect the phenomenon to the domain: e.g., "The theme of this column is the effects (immediate and long-term) that human activity can have on the environment." One response did not fit either criterion, but was scored as causal because it explicitly linked sorting to the

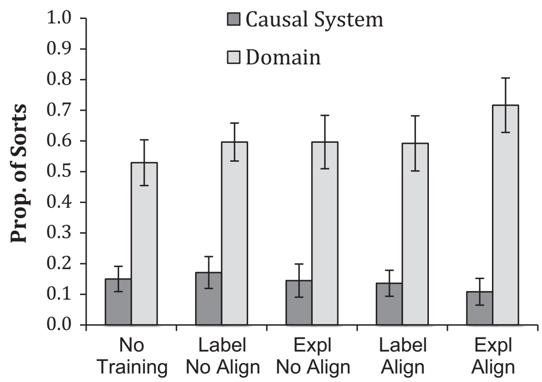

Fig. 6. Experiment 1a: Mean and standard error of proportions of causal and domain sorting for the second sort.

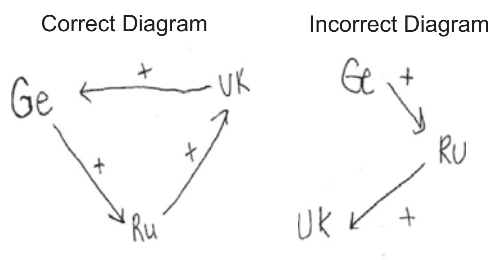

Fig. 7. Results of Experiment 1a. A correct and an incorrect depiction of the positive feedback system governing the Pre-WW1 arms race. The incorrect diagram represents a causal chain instead of a positive feedback system.

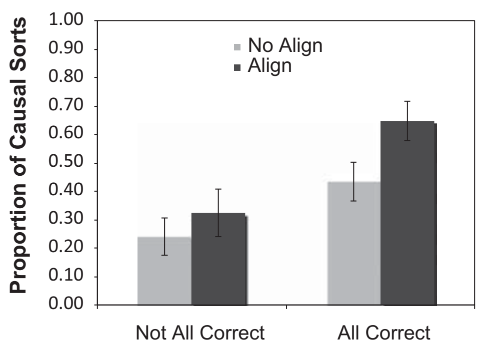

Fig. 8. Experiment 1a: Effects of structural alignment and causal diagram accuracy (all diagrams correct vs. one or more diagram incorrect) on causal sorting (means and standard errors of proportions).

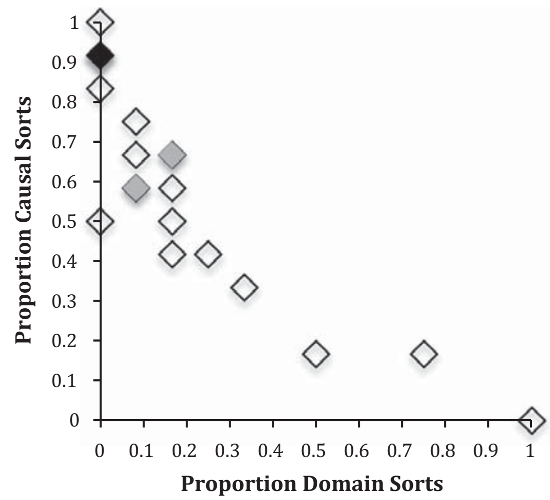

Fig. 9. Experiment 1b: Scatterplots showing rates of causal sorting (*Y* axis) and domain sorting (*X* axis). White diamonds represents a single subject's score; grey diamonds represent two subjects' scores; black diamonds' represent three subjects' scores.

causal graphs that had been diagrammed previously: "I sorted them based on the system of graphs..." (see Table 5).

One motivation for the strategy analysis was to discover whether the Alignment advantage in the prior studies stemmed simply from increased attention to causal factors, rather than from promoting causal abstractions. If this were the case, we would expect more causal focus in the strategies reported by the Alignment group. We found that the majority of reported strategies (44 out of 63) were causal, with no significant difference across the three conditions: Explication/Alignment (17 out of 21); Explication/No-Alignment (14 out of 19); Explication/List-Factors (13 out of 23). The difference between Explication/Alignment and Explication/List-Factors did not reach significance, $\chi^2(1) = 3.02$, $p = .08$. Of course, it is possible that aligning exemplars did led to greater attention to causality, and that a larger study would show this. However, even if this turns out to be the case, the next set of analyses show that the advantage of alignment extends beyond increased attention to causal relations.

We next analyzed the actual sorting patterns of the subjects who reported causal strategies. This showed that, despite their verbal emphasis on causality, subjects differed in their rate of causal sorting (see Fig. 12). The Explication/Alignment condition ($M = .77$) elicited more causal sorts than Explication/No-Alignment ($M = .51$) and Explication/List-Factors ($M = .63$), $f(2,41) = 6.31$, $p < .01$, $\eta^2_p = .23$; planned comparison t-tests: Explication/Alignment vs. Explication/No-Alignment, $t(29) = 3.26$, $p < .01$, $d = 1.21$; Explication/Alignment vs. Explication/List-Factors $t(28) = 2.13$, $p < .05$, $d = .81$.

We also analyzed the kinds of non-causal responses made by subjects who reported causal strategies, including both domain-based responses and errors.[10] While the Explication/No-Alignment ($M = .17$) condition gave rise to more domain sorts than Explication/Alignment did ($M = .06$), Explication/List-Factors did not ($M = .08$), $f(2,41) = 3.54$, $p < .05$, $\eta^2_p = .14$. However, both Explication/List-Factors ($M = .29$) and Explication/No-Alignment ($M = .32$) elicited more error sorts (miss-sorts and 'other' responses) than did Explication/Alignment ($M = .16$): $t(28) = 2.38$, $p < .05$, $d = .91$; $t(29) = 3.07$, $p < .01$, $d = 1.15$ (respectively). In sum, among the subjects who reported a causal strategy, Alignment elicited the most causal sorts and fewest error sorts.

[10] An error was defined as sorting a card in a column that did not match in either causal system or domain or placing a card in the "other" column.

Next, we examined the minority of subjects ($n = 19$) who reported a domain strategy (due to the small number per condition, and a lack of reliable differences across conditions, we combined these subjects across conditions). They mostly sorted by domain – ranging from .50 to 1.00 domain sorts – with few causal sorts (.02 causal, .77 domain, and .21 errors). (In contrast, only a single subject reporting a causal strategy sorted primarily by domain (at .58).) Interestingly, domain-strategy subjects were less accurate overall in their diagrams ($M = 6.68$ out of 8) than causal-strategy subjects ($M = 7.63$), $t(60) = 3.54$, $p < .01$, $d = .99$. We again estimated the confidence intervals of each condition via non-parametric bootstrapping. The 95% CI for the domain-strategy subjects was 6.0–7.26 (below the 7.63 mean of the causal-strategy subjects), and the 95% CI for the causal-strategy subjects was 7.41–7.81 (above the 6.68 mean of the domain-strategy subjects).

Next we analyzed the reported strategies for the second sorting attempt ($N = 57$; six subjects who reported strategies for their first attempt did not report their second strategy). We classified their strategies as to whether they were a modification of their initial strategy (causal or domain) or a switch from one to the other. As in the prior studies, the likelihood of switching from causal to domain strategy was far greater (33 switched while 7 modified) than the reverse (6 switched while 11 modified), $\chi^2(1) = 12.30$, $p < .001$.

In summary, subjects mostly espoused a causal strategy. However, among those subjects who expressed a causal strategy, subjects who aligned examples were better able to implement this strategy. Specifically, subjects in the Explanation/Alignment group produced more causal sorts and fewer error sorts than those in the other two groups – supporting the idea that aligning parallel causal phenomena led subjects to form abstract, portable causal system schemas. Additionally, analysis of the reported strategies on the second sort confirms that, as in the prior studies, domain strategies are readily available to people who sort causally; but the reverse does not hold.

4. General discussion

Expertise is often marked by an increased sensitivity to important abstract patterns in the world. Our goal here was to shed light on the kinds of learning processes that foster this generalized sensitivity. We focused on high-level causal system categories such as positive feedback and common-effect – an important set of cross-domain abstractions. Recognition of these causal systems allows people to understand and predict key phenomena in the world, from global warming to economic bubbles. We know from prior research that these categories are not salient to college students without advanced training (Rottman et al., 2012). Our interest here is in what kinds of learning experiences allow people to spontaneously recognize these patterns in the world.

Of course, in our laboratory situation we were not able to achieve the richness and complexity of real-world experience. To approximate this situation, we used the Ambiguous Sorting Task (AST), in which subjects are free to sort as they choose, and in which a highly salient competing organization – that of content domain – is present. Our findings indicate that the AST technique was successful in providing a competing organization. In the No-Training condition, as in Rottman et al.'s study, we found very low levels of spontaneous causal sorting; only 14% of examples were sorted causally. Thus, the AST achieved its goal in that the high-level causal patterns did not stand out. This sets the stage for asking whether and how alignment and schema-explication can improve causal insight.

There are three main findings. First, as predicted, receiving explications of how the training examples fit the causal schemas improved subjects' causal representations of the examples. This was evidenced by (1) analysis of subjects' causal diagrams, which showed that explication led to more accurate representations of the training examples; and (2) regression and mediation analyses indicating that the positive effects of causal explication on transfer stemmed from its having increased the accuracy of subjects' representations of the training examples (Experiment 1a). Second, structural alignment increased causal sorting, even for those who drew accurate diagrams for every training example – evidence of its value in promoting abstractions that go beyond the examples given. Third, combining causal explication with analogical abstraction led to the greatest degree of causal insight, as assessed by greater causal sorting in the AST. Further studies showed that the results cannot be attributed to the diagram task *per se* (Experiment 1b) nor to additional exposure to the pairs of examples (Experiment 2), nor simply to a general attentional shift toward causality among alignment subjects (Experiment 2).

4.1. Analogical processing in learning and transfer

The consistent pattern across the set of experiments adds to research showing that analogical comparison can lead to relational insight and thereby facilitate transfer to new situations. Further, our findings join other recent work in arguing for an expanded role for analogical processing in

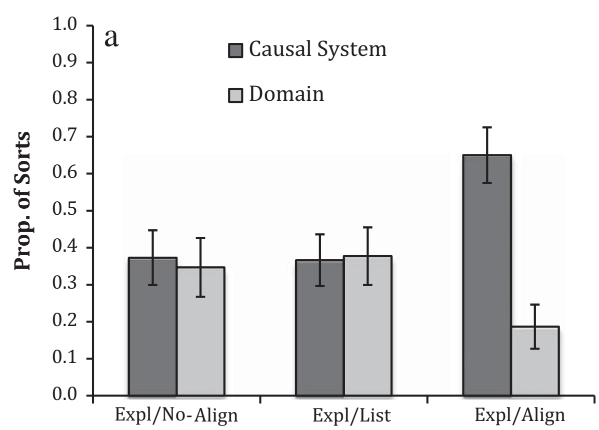

Fig. 10a. Results of Experiment 2, first sort. Mean and standard error of proportions of causal and domain sorting, showing Explication/No-Alignment, Explication/List-factors, and Explication/Alignment.

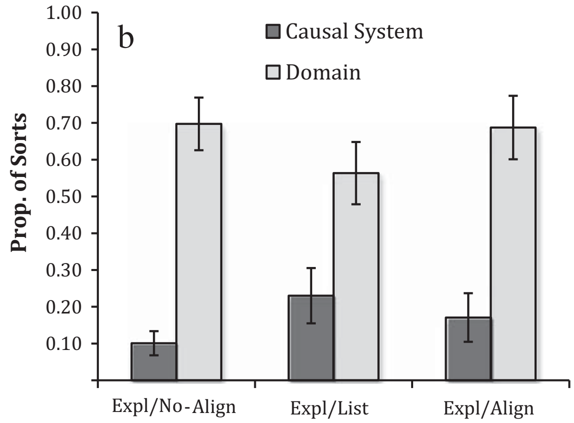

Fig. 10b. Results of Experiment 2, second sort. Mean and standard error of proportions of causal and domain sorting, showing Explication/No-Alignment, Explication/List-factors, and Explication/Alignment.

learning and transfer (Day & Goldstone, 2012; Doumas & Hummel, 2013; Gentner, 2003; Gentner, 2010). In much of the initial work on analogical transfer, the focus was on direct mapping and inference projection from one example to another (e.g. Gick & Holyoak, 1980; Reed, 1987; Reeves & Weisberg, 1994; Ross, 1984). But as Day and Goldstone (2012) and Lave (1988) note, this "two-problem" experimental design, wherein transfer is assessed by determining whether a given example influences a single subsequent (analogical) case, greatly restricts the scope of what prior knowledge a participant can bring to bear on a new situation.

Our goal in developing the AST is to go beyond this method – to study the more generalized ability to spontaneously notice important patterns. Our findings add to evidence that the effects of analogical processing go well beyond inference projection (Gentner, 2010; Kurtz et al., 2001). In our studies, people who received a combination of explication and alignment showed a generalized sensitivity to the key causal patterns that emerged from their training. This kind of learning serves as a source of insight that can influence the way we interpret further information (e.g., Day & Goldstone, 2012; Doumas & Hummel, 2013; Fernbach & Sloman, 2009; Gentner & Stevens, 1983; Schwartz, Sears, & Bransford, 2005). We conjecture that the mutual alignment of examples during learning

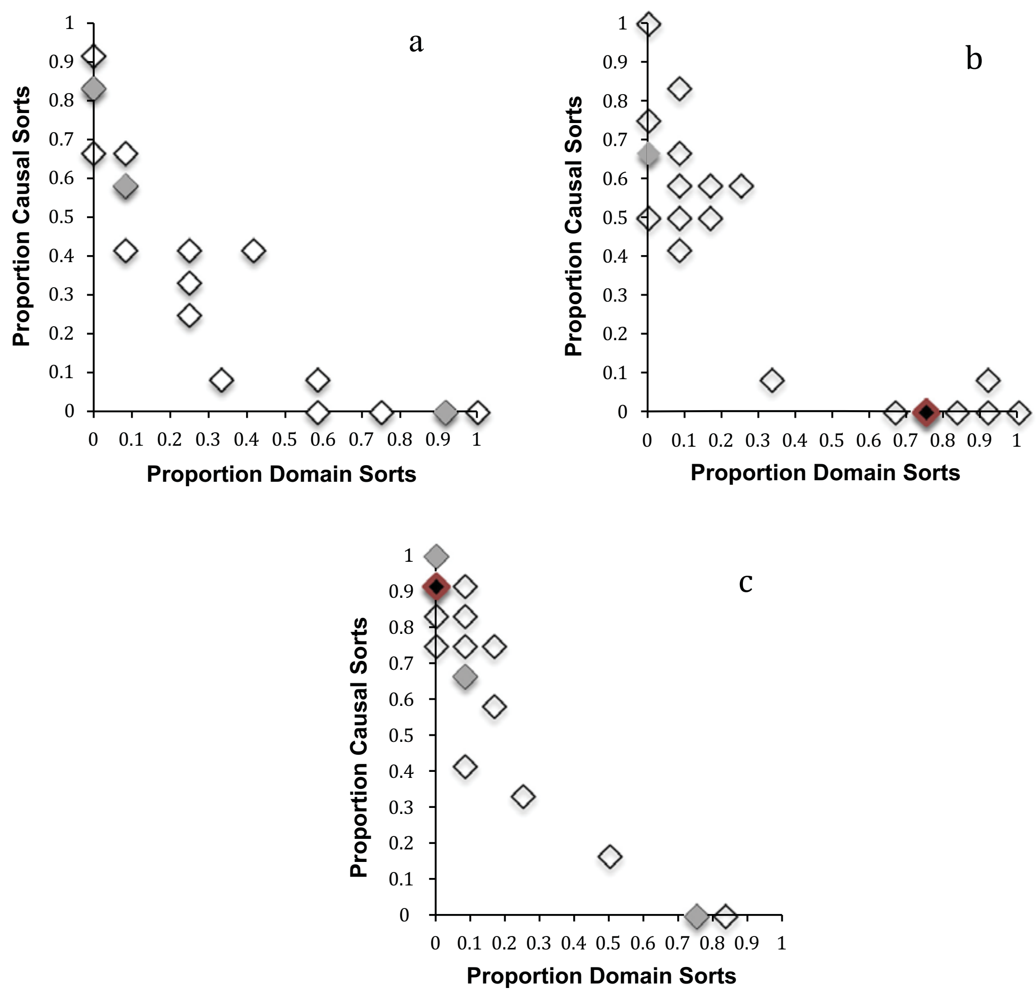

Fig. 11. Experiment 2: Scatterplots relating causal sorting rate to domain sorting rate for each condition. White diamonds represents a single subject's score; grey diamonds represent two subjects' scores; black diamonds' represent three subjects' scores; and black diamonds with red border represent four subjects' scores. The conditions are (a) Explication/No-Alignment; (b) Explication/List-Factors; (c) Explication/Alignment.

Table 5
Examples of sorting strategy descriptions, showing three typical causal strategies and three typical domain strategies, as well as three that did not neatly fit our criteria.

Example Causal Strategy Descriptions
"I looked at how certain things affected other things. For example, I grouped together structures where one thing had many different effects, and I grouped together situations where there was a cyclical chain of effects."
"The first column described a steady escalation, similar to positive feedback. All the cards sorted into this column described a similar escalation concerning the topic on that card. The second column described many factors influencing one outcome (common outcome). The third column described a chain of results, where one event causes a string of events which lead to a cause that may not have been originally anticipated when looking at the beginning cause and final event. The 4th column described one event creating many reactions (common cause)."
"I sorted them by whether they explained a common cause, a chain reaction, a common effect, or a positive feedback relationship."
Example Domain Strategy Descriptions
"I sorted the descriptions based on the broad topic they closely relate to. The first category I labeled as mechanical as the description, mention certain things such as powertrain, gears, piping, nuclear reactor and car engines. The second category is based on the economy as there were terms in the descriptions such as cost of living index, unemployment, and stocks. The third category is based on ecology and the environment. There were key words in the descriptions that relate to the category such as global warming, greenhouse gases and non-native species. The fourth category is based on human bodily functions as there are key words such as firing of neuron, allergic reaction, and heart disease."
"I sorted all the cards pertaining to the economy or money together because they all seemed to be related to the US economy, which is mentioned in the first, base card. I grouped the cards related to CO_2 and global warming together because they all included information about the environment and its safety. I grouped the cards about biological processes together because they were all about the human body, though they were not specifically about drug overdose like in the top card. The cards in the "other" section were not about water, piping, the economy, the human body, or the environment."
"I split the cards into 4 categories: Medical, Environmental, Economy, Mechanical"
Example Ambiguous Responses
"1: The card on how water distribution is does not really tie into economic trends, environmental issues or personal health. 2: The theme on these cards is the interconnectedness of many economic indicators. 3: The theme of this column is the effects (immediate and long-term) that human activity can have on the environment. 4: The theme of this column is personal health. 5: At the end of this exercise, the theme of both cards can be automobile propulsion." (Classified as a domain strategy)
"I sorted everything by their relevance to the main topic. For instance, under the topic about drug tolerance I included the description about allergies and factors contributing to heart disease because they pertain to how the human body is affected by different substances. If I could not find a strong enough connection between a topic and description, I put it in the "other" column." (Classified as a domain strategy)
"I sorted them based on the system of graphs from part one of the experiment today. I noticed that the categories were similar to the systems, so I looked for this trend in the cards and was able to sort them through that." (Classified as a causal strategy)

creates a kind of internal vocabulary that allows for the recognition of abstract relational patterns in novel situations.

We have provided evidence that analogical processing supports deriving causal systems abstractions that aid in transfer. We suggest that this kind of higher-order knowledge of causal systems can make people better able to detect such patterns in novel experiential situations. Consistent with the hypothesis, Fernbach and Sloman (2009) found that people improved in their ability to discover complex causal structures – specifically, causal chains – from observation when they were pretrained with schematic causal information. In combination with the current results, this supports the possibility that alignment-based pretraining might support later causal structure discovery.

A second open question concerns the mechanism that underlies analogical abstraction. We hypothesize that alignment improved their existing structured knowledge in two ways: by dropping non-common surface features from the representations and by refining the representations of the relations themselves. Both the symbolic AI models built upon the Structure-Mapping Engine (Falkenhainer et al., 1989; Kuehne, Forbus, Gentner, & Quinn, 2000; Lovett, Tomai, Forbus, & Usher, 2009) and the symbolic-connectionist architectures of LISA (Hummel & Holyoak, 1997) and DORA (Doumas, Hummel, & Sandhofer, 2008) offer computational descriptions of feature-shedding and relational re-representation (see Gentner & Forbus, 2011 for a recent review). However, it remains an open question as to how to distinguish the two processes in behavior, as each would lead to improvements in transfer.

4.2. Category learning and representation

The current studies add to the growing literature on relational categories. These are categories whose membership is determined not by common intrinsic properties of the members, but by common relational structure (Corral & Jones, 2014; Gentner, 2005; Gentner & Kurtz, 2005; Gentner et al., 2011; Goldwater, Markman, & Stilwell, 2011; Jung & Hummel, 2011; Markman & Stilwell, 2001; Rehder & Ross, 2001; Tomlinson & Love, 2010). We view the causal system categories studied here as examples of relational categories. This is consistent with research showing the connection between causal reasoning and category representation, which finds that the causal relational structures binding features of category exemplars are critical for determining category membership and category-based induction (e.g., Ahn et al., 2000; Forbus, 1984; Rehder, 2003; Rehder & Burnett, 2005; Sloman et al., 1998).

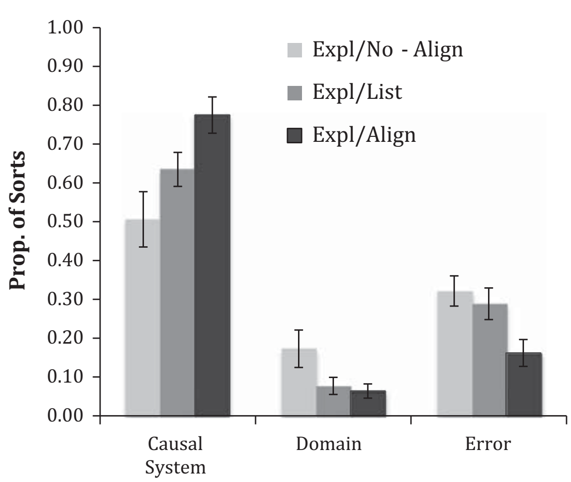

Fig. 12. Experiment 2. Means and standard error of proportions of causal, domain, and error sorting for subjects who reported a causal sorting strategy, showing Explication/No-Alignment, Explication/List-factors, and Explication/Alignment.

Our finding that analogical comparison can support the abstraction of such causal structures is consistent with past work showing that comparing examples supports forming relational categories (Christie & Gentner, 2010; Gentner et al., 2011; Goldwater & Markman, 2011; Jung & Hummel, 2011; Kotovsky & Gentner, 1996; Kurtz et al., 2013; Tomlinson & Love, 2010). Further, our finding that alignment was most beneficial when exemplar relational representations were accurately represented accords with those of Doumas and Hummel (2013) who found that alignment aids learning category-defining higher-order relations (i.e., relations among relations) only when the lower-order relations are properly represented.

Another link between our findings and the relational category literature is that we found a shift from focusing on relatively concrete domain-level commonalities in the No-training condition to noticing common causal patterns among the subjects after full training. This kind of "relational shift" in category understanding – from an early focus on concrete features to a later focus on relational structure – is well-attested in cognitive development (Gentner & Rattermann, 1991; Imai, Gentner, & Uchida, 1994; Richland, Morrison, & Holyoak, 2006), as well as in novice-expert studies of adults (Chi et al. 1981; Jee et al., 2014; Proffitt, Coley, & Medin, 2000; Shafto & Coley, 2003; Stains & Talanquer, 2008). Further, there is evidence that analogical comparison can promote this shift from object focus to relational focus (Christie & Gentner, 2010; Gentner et al., 2011; Kotovsky & Gentner, 1996; Namy & Gentner, 2002).

4.3. Analogical learning in education

These findings have implications for educational practice. It appears that a combination of explication plus structural alignment can lead to greater attention to and ability to recognize causal patterns, mirroring the pattern found in development of expertise. Indeed, the contrast between the No Training group (14% causal sorting) and the maximal-training group (Explication/Alignment, 62% causal sorting) parallels the contrast found by Rottman et al. (2012) between science novices (23% causal sorting) and physical science majors (44% causal sorting). Of course, we would not claim that a brief training session can substitute for systematic study and experience. Nevertheless, these results join with current research in suggesting that providing more support for analogical comparison improves education in science and mathematics (e.g., Chen & Klahr, 1999; Gentner et al., 2015; Jee et al., 2014; Richland, Zur, & Holyoak, 2007; Rittle-Johnson & Star, 2009; Schwartz, Chase, Opezzo, & Chin, 2011; Thompson & Opfer, 2010; see Alfieri, Nokes-Malach, & Schunn, 2013 for a meta-analysis).

4.4. Conclusion

This research shows that explicating and aligning causal phenomena fosters abstraction of the causal structure and increased sensitivity to causal systems in novel phenomena. These results add to the literature on causal reasoning by focusing on the recognition of common causal structure across disparate phenomena. More generally, these results add to our understanding how abstract knowledge is acquired over experience.

Acknowledgements

This study was supported by ONR Grant (N00014-92-J-1098) to the second author. This research was supported by National Science Foundation grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC). Preparation of the paper was supported in part by an award to the second author from the Alexander von Humboldt Foundation and by a fellowship at the Hanse-Wissenschaftskolleg, Delmenhorst, Germany.

References

Ahn, W. K., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge: MIT Press.
Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation vs. mechanism information in causal attribution. *Cognition, 54*, 299–352.
Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361–416.
Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113.
Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120.
Cheng, P. W. (2000). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.
Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.
Chi, M., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.
Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356–373.
Corral, D., & Jones, M. (2014). The effects of relational structure on analogical learning. *Cognition, 132*(3), 280–300.
Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47*(3), 153–176.
de Kleer, J., & Brown, J. S. (1981). Mental models of physical mechanisms and their acquisition. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 285–309). Hillsdale, NJ: Erlbaum.
Doumas, L. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PLoS ONE, 8*(6), e63889.
Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological review, 115*(1), 1.
Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1–63.
Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of experimental psychology: Learning, memory, and cognition, 35*(3), 678.
Fillmore, C. J. (1978). On the organization of semantic information in the lexicon. Donka Farkas, Wesley M. Jacobsen & Karol W. Todrys (Eds.), *Parasession on the lexicon* (pp. 148-173). Chicago: Chicago Linguistic Society.
Forbus, K. D. (1984). Qualitative process theory. *Artificial Intelligence, 24*, 85–168.
Forbus, K. D. (1985). The role of qualitative dynamics in naive physics. In J. Hobbs & R. Moore (Eds.), *Formal models of the common sense world*. Palo Alto: Tioga Press.
Forbus, K., Nielsen, P., & Faltings, B. (1991). Qualitative spatial reasoning: The CLOCK project. *Artificial Intelligence, 51*(1–3).
Gary, M. S., Wood, R. E., & Pillinger, T. (2012). Enhancing mental models, analogical transfer, and performance in strategic decision making. *Strategic Management Journal, 33*(11), 1229–1246.
Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775.
Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure-mapping and relational language support children's learning of relational categories. *Child Development, 82*, 1173–1188.
Gentner, D., Levine, S. C., Dhillon, S., Ping, R., Bradley, C., Poltermann, A., & Honke, G. (2015). Rapid learning in a children's museum via analogical comparison. *Cognitive Science.* [in press].
Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.
Gentner, D. (2005). The development of relational category knowledge. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 245–275). Hillsdale, NJ: Erlbaum.
Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the lab* (pp. 151–175). Washington, DC: APA.
Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408.
Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.
Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London: Cambridge University Press.
Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Erlbaum.
Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.
Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38.
Goldwater, M. B., & Markman, A. B. (2011). Categorizing entities by common role. *Psychonomic Bulletin and Review, 18*(2), 406–413.
Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition, 118*, 359–376.
Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 3–32.
Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. *Causal Learning: Psychology, Philosophy, and Computation*, 86–100.
Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*(3), 233–259.
Hayes, P. J. (1979). The naive physics manifesto. In D. Michie (Ed.), *Expert systems in the micro-electronic age* (pp. 242–270). Edinburgh: Edinburgh University Press.
Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*(3), 427.
Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45–75.
Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
Jee, B., Gentner, D., Uttal, D., Sageman, B., Forbus, K., Manduca, C., et al. (2014). Drawing on experience: How domain knowledge is reflected in sketches of scientific structures and processes. *Research in Science Education*.

Jung, W., & Hummel, J. E. (2011). Progressive alignment facilitates learning of deterministic but not probabilistic relational categories. In *Proceedings of the 33rd annual conference of the cognitive science society*.

Klahr, D., & Chen, Z. (2011). Finding one's place in transfer space. *Child Development Perspectives, 5*, 196–204.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797–2822.

Kuehne, S. E., & Forbus, K. D. (2002). Qualitative physics as a component in natural language semantics: A progress report. In *Proceedings of the 24th annual conference of the cognitive science society*, August.

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1303–1310.

Kuehne, S., Forbus, K., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. In *Proceedings of the 22nd annual meeting of the cognitive science society* (pp. 770-775).

Kurtz, K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*(4), 417–446.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 856.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). New York, NY: Oxford University Press.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology, 61*, 303–332.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*(2), 167–204.

Lovett, A., Tomai, E., Forbus, K., & Usher, J. (2009). Solving geometric analogy problems through two-stage analogical mapping. *Cognitive Science, 33*(7), 1192–1231.

Mackie, J. L. (1980). *Cement of the universe: A study of causation*. Oxford, England: Clarendon Press.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence, 13*(4), 329–358.

McCawley, J. D. (1968). The role of semantics in a grammar. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 125–169). New York: Holt, Reinhart and Winston.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131*(1), 5–15.

Pearl, J. (2000). *Causality: Models, reasoning and inference* (Vol. 29). Cambridge: MIT press.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 811–828.

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 124–139.

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115*(3), 381–400.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264–314.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*(3), 323–360.

Rehder, B., & Ross, B. H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1261–1275.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94*, 249–271.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316*(5828), 1128–1129.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371–416 (Ross, 1987).

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science, 36*(5), 919–932.

Schwartz, D. L., Chase, C. C., Opezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*, 759–775.

Schwartz, D. L., Sears, D., & Bransford, J. D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich, CT: Information Age.

Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naïve similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory & Cognition, 29*, 641–649.

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford: Oxford University Press.

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

Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1251.

Stains, M., & Talanquer, V. (2008). Classification of chemical reactions: Stages of expertise. *Journal of Research in Science Training, 45*, 771–793.

Steyvers, M., Tenenbaum, J., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453–489.

Thagard, P. (1989). Explanatory coherence. *Behavior and Brain Sciences, 12*, 435–502.

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*(6), 1768–1786.

Tomlinson, M. T., & Love, B. C. (2010). When learning to classify by relations is easier than by features. *Thinking & Reasoning, 16*, 372–401.

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science, 15*(6), 307–311.

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*, 276–332.