Inference processes in causal analogies

Julie Colhoun (colhoun@northestern.edu)
Dedre Gentner (gentner@northwestern.edu)
Department of Psychology, 2029 Sheridan Road
Evanston, IL 60208 USA

ABSTRACT

In recent papers, Lee & Holyoak (2007, 2008a, 2008b) argue that extant models of analogy fail to explain how people draw inferences from causal analogies. In the current research, we argue that structure-mapping theory sufficiently explains the analogical inferences drawn from these causal analogies, and that, contrary to L&H's claims, the effect inference can indeed be evaluated by a post-analogical causal reasoning process. In Study 1, we present evidence that – consistent with SMT (Gentner, 1983), and counter to L&H – when relational inferences are considered, the inductive strength of these causal analogies matches their similarity. In Study 2, we provide evidence that, by analogical mapping, the base analog makes two contributions to the reasoner's knowledge about the causal system in the target, and argue that this analogically-constructed causal model is subsequently used to determine the presence of the effect. In an SME (Falkenhainer et al., 1989) simulation, we show that "outsourcing" the effect inference to a simple post-analogical calculation can match L&H's human data very closely. In short, although we agree with Lee & Holyoak that analogy is important for learning about causal systems, we maintain that analogy is a domain-general process. Models of analogical processing need not—and should not—subsume causal inferencing processes.

INTRODUCTION

We live in an uncertain world; daily we are confronted with situations in which we must reason about the unknown. Often we refer to similarity in service of this goal: What is that creature crossing my path? If it walks like a duck and quacks like a duck, we say, it probably is a duck. Drawing analogies between situations help us to better understand novel situations and to make predictions about them. To make inferences, we also use causal relations: when it's raining, we can infer that the pavement must be wet and slippery. These kinds of inductive reasoning give us the capacity to better understand and navigate uncertain situations.

Analogical reasoning confers the ability to determine similarity and to make inferences from one situation to another. Causal reasoning provides the ability to make inferences (predictions and diagnoses) about a given causal system or situation based on the particular generative and preventative causal relations at work in that system. They are alike in being informative; they are different in that analogy inherently applies to two systems and causal reasoning, to one system.

Furthermore, analogies often involve causal systems. Higher-order relations that govern analogy – those relations bind one relation to another ("skidding on the ice *caused* the car to spin off the road"), and thus give depth to a relational structure – are frequently causal relations. By analogy, we might think: if that car skidded on the ice, then perhaps another moving vehicle – a truck or a skateboard – could also skid on another slippery surface, such as wet leaves.

The purpose of this paper is to investigate the inferences made from causal analogies, and the processes which produce them. Our main goal is to examine the reasoning processes that produce the inferences drawn from causal analogies.

Analogical Inference

Analogical reasoning provides the ability to determine similarity and to make inferences from one situation to another. According to Gentner's (1983, 1989; Gentner & Markman, 1997) structure-mapping theory (SMT), analogical mapping is the process of establishing a structural alignment between two situations and then projecting inferences. The theory assumes structured representations in which the elements are connected by relations, and higher-order relations (such as causal relations) connect first-order statements (see Falkenhainer, Forbus, & Gentner, 1989; Markman, 1999). During the alignment process, possible matches are first found between individual elements of the two represented situations; these matches are then combined into structurally consistent clusters, and finally into an overall mapping. The resulting alignment consists of an set of correspondences between the elements and relations of the two situations, with an emphasis on matching systems of interconnected relational predicates (the systematicity principle). As a natural outcome of the alignment process, candidate inferences are projected from the base to the target. These inferences are propositions connected to the common system in one analog, but not yet present in the other. Thus, structural completion can lead to spontaneous unplanned inferences.

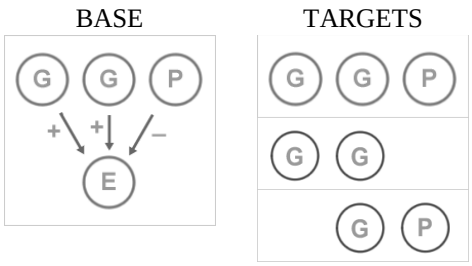
In general, models of analogy, including structure-mapping theory, (Gentner, 1983, e.g.) postulate that the more similar two analogs are, the greater their inductive strength. Lassaline (1996) explored causal analogies, and in particular the strength of inferences that result from various kinds of commonalities. In one study (Exp.2), she provided evidence that a greater number of *binding, non-shared causal relations* (those causal relations which are present in the base analog, and which are bound to, or take as an argument, an attribute shared by both analogs) leads to greater inductive strength. That is,

when a causal relation is present in the base, and its causal antecedent is shared by both analogs, people judge the relation's effect to be more likely.

Effect Inferences and Similarity

Intuitively it seems clear that a generative causal relation should increase the likelihood of the effect in the target. In recent research, Lee & Holyoak (2007, 2008a, 2008b) capitalize on the converse idea, that *preventative* causal relations should *decrease* the strength of an effect inference. In one study, they gave participants pairs of animals and asked for similarity judgments or inference ratings. The base animal consisted of three causal properties, one effect property, and three relations. Two of the causal properties are generative: each "tends to cause" the effect property. The third causal property is preventative, and "tends to prevent" the effect property. (See Figure 1.) Using these analog pairs, L&H show that when a preventative property in the base is also present the target analog, the ratings of the effect inference in the target decrease (vs. when the preventative property is not present in the base), but similarity between base and target increases. They thus show a dissociation between similarity and the effect inference. We will address this finding in Study 1.

Figure 1: Causal structure of base and target items



However, Lee & Holyoak make two apparently conflicting claims about how causal models and analogical inference interact. The first claim is that "analogical inference involves using the source analog to guide construction of a causal model of the target analog" (2008b, p1119), i.e., that "some form of analogical transfer can guide construction of a causal model appropriate for the target domain" (2008b, p1121). The competing claim is that "causal models guide analogical inference" (2008b, p1116). The former claim suggests that analogical inference plays the guiding role (by constructing the causal model); the latter claim implies that causal models have the guiding role (in analogical inference);

We agree with the first claim: specifically, we agree that analogical mapping (alignment and inference projection), guides the construction of a causal model in the target analog. With respect to L&H's first claim, we would not agree that causal models guide analogical inference. Rather, (1) as just noted, the causal model of the base domain is imported into the target via analogical inference; and (2) once the new inferences have been assimilated into the target, new causal inferences may be generated in the target domain. The causal model is also used to *evaluate* the analogical inferences after the mapping is completed. We will address these claims in Studies 2 and 3.

Lee & Holyoak conclude that because causal models guide analogical inference, the basic elements of causal models must be incorporated into models of analogy. We disagree, and provide experimental results in support of the integrity of the analogical process. We also provide a computational simulation (Study 3) using SME to demonstrate that the inference evaluation can indeed be outsourced to a post-analogical process.

Overview of Current Research.

Study 1A replicates Lee & Holyoak's (2008b) Experiment 1. Study 1B further examines the relationship between similarity, the effect inference, and the overall inductive strength of the analogy. Study 2 examines the contributions of analogical reasoning to the causal inference. Study 3 is a computational simulation to model the human data of Study 1. This simulation uses SME followed by a simple causal calculation operating on SME's output, as proof of concept that the causal inference evaluation *can* be "outsourced" to a postanalogical process.

STUDY 1A: Replication

As a check for consistency with our subsequent research, this study seeks to replicate Lee & Holyoak's (2008b) Experiment 1.

Method

Participants. Seventy Northwestern undergraduates participated to fulfill a course requirement. Half (n=36) were randomly assigned to the Similarity condition, and half (n=34) to the Inference condition.

Materials and procedure. Each participant received a set of nine descriptions of animal pairs (plus 3 filler items). Each of the nine test pairs included a base animal *that was described* as having four properties: one effect property (E); two generative properties (G1, G2) each of which "tends to cause" the effect property; and one preventative property (P) which "tends to prevent" the effect property. (See Figure 1.) These base animals thus followed the same structure as Lee & Holyoak's (2008b, Exp.1) stimuli. The target animal in the pair had either two generative features and one preventative feature (GGP), one generative feature and one preventative feature (GP), or two generative features (GG). (See Appendix A for sample stimuli.)

In total, 27 pairs were created using nine base animals and three target animals for each base. Each participant was randomly assigned three pairs of each target-type (3-GGP, 3-GP, 3-GG), with each of the nine base animals appearing exactly once. Pairs were arranged in three blocks; each block contained three

pairs (1-GGP, 1-GP, 1-GG), randomly ordered within the block. One filler item appeared after each block.

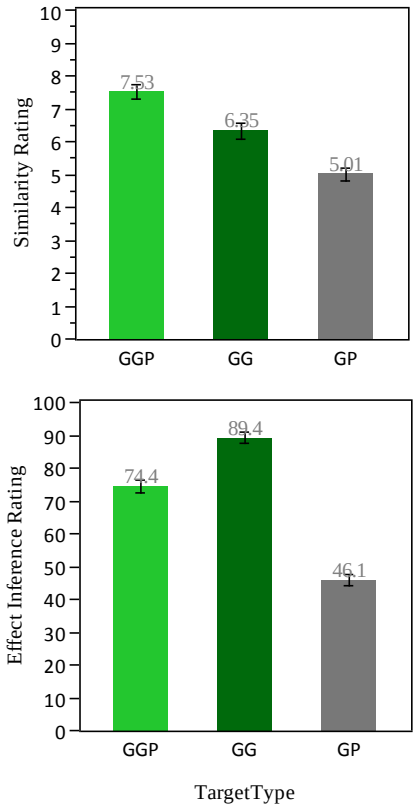
Participants in the Similarity condition were asked to rate the similarity of the animals in each pair; participants in the Inference condition were asked to provide an inference rating. Materials were presented and responses were collected using MediaLab on PC; instructions and experimental trials were self-paced.

Results

The results for both similarity ratings and inference ratings are shown in Figure 2. Similarity ratings and inference ratings were analyzed separately using one-way ANOVA repeated-measures design.

For the Similarity group, mean ratings differed significantly by target type, $F(2,286)=44.4$, $p<.0001$. Tukey's HSD contrasts ($q=2.36$) showed that the GGP targets ($M=7.53$, $SD=1.72$) were rated more similar to the base than were the GG ($M=6.35$, $SD=1.99$) or GP ($M=5.01$, $SD=1.34$) targets, and the GG targets were rated more similar to the base than the GP targets.

Figure 2: Mean Similarity and Effect inference ratings, by target type



In the Effect Inference group, the effect inference ratings also differed significantly by target type, $F(2,270)=164.6$, $p<.0001$. Tukey's HSD contrasts ($q=2.35$) showed that the effect inference ratings for the GG targets ($M=89.4$, $SD=11.7$) were significantly higher than for the GGP ($M=74.4$, $SD=15.3$) or GP ($M=46.1$, $SD=13.0$) targets, and the ratings for the GGP targets were significantly higher than for the GP targets.

The pattern of similarity ratings (GGP > GG > GP) does not match the pattern of ratings for the Effect inference (GG > GGP > GP). Although the GGP Target was rated most similar to the base, the Effect Inference Rating was highest for the GG Target. This pattern replicates Lee & Holyoak's Experiment 1 (2008b).

Discussion

These results show that when a shared feature was eliminated[1] -- when two features (GG) where shared, rather than three (GGP) -- similarity decreased, but the effect inference increased. This replicates Lee & Holyoak's finding.

At first glance, this pattern seems to pose a major challenge to theories of analogy, as Lee & Holyoak (2007, 2008a, 2008b) point out. Most models, include SMT, predict that the inferential strength of an analogy should correlate with the similarity of the analog pairs. These results seem to suggest a dissociation between inferential strength and similarity.

However, the only inference tested in these experiments (our Study 1, Lee & Holyoak's studies, Lassaline's study) is the effect inference. Although Lee & Holyoak claim that "the ultimate goal of analogical inference is to predict the presence or absence of some outcome in the target" (2008b, p 1112), we argue that there are multiple goals of analogical inference – not the least of which is *understanding*. Each inference projected from base to target represents new information that may be true of the target. Such inferences include not only the inferred presence of some specific outcome, but also the inferred presence of whole chunks of relational structure. These inferences yield a better understanding of the target system, they help us explain why certain conditions occur or exist in the target, and they provide a basis for extrapolating new information – i.e., *learning* about the target.

Furthermore, according to Structure-mapping Theory, it is *relational* similarity – shared structure, consisting of interconnected relations – that provides support for candidate inferences, by structural completion (Clement & Gentner, 1991; Markman & Gentner, 2000; Gentner & Kurtz, 2006; see also Blok & Gentner, 2000, for a further discussion of inferences and the goodness of the common schema). In L&H's stimuli, there is very little shared structure; nothing is known about the similarity of Animals A

[1] Lee & Holyoak state that their experiments 1 & 2 "reduced structural overlap by eliminating a shared relation" (2008b, abstract). Strictly speaking this is inaccurate: none of the stated relations are explicitly shared by the base and target. Rather, they reduced similarity by eliminating a shared *feature* which was connected to a non-shared, binding relation. In our assessment, this manipulation effectively reduced the support for that relational candidate inference.

and B, apart from a few shared features. The specific relations in question are stated as present in the base only; they are not shared by the target. Thus, the similarity between the two animals is almost entirely feature-based. Any causal relations in the target must be projected from the base as candidate inferences.

For these reasons, we argue that the inductive strength of an analogy should be measured by all its candidate inferences, and not solely a single effect inference. Despite the limited structural overlap, we predict that when all the inferences are considered, the inductive strength of these analogies should reflect the pattern of similarity ratings.

To test this claim, we gave participants the same animal pairs as in Study 1A, but gave them a list of 4-5 possible inferences, and asked them to endorse those inferences that are "probably true" of the target animal. Our prediction is that when these endorsements are taken together, the resulting overall inductive strength of the analogies will parallel the pattern of similarity ratings.

STUDY 1B: Relational Inferences

Method

Participants. Twenty-two Northwestern undergrads participated to fulfill a course requirement. Three additional participants were excluded for failing the catch trials.

Materials and procedure. As in Study 1A, each participant received a set of nine descriptions of animal pairs (plus 3 filler items). The same 27 animal pairs were used, and as in Study 1A, each participant was randomly assigned three pairs of each target-type (3-GGP, 3-GP, 3-GG), with each of the nine base animals appearing exactly once. Pairs were arranged in three blocks; each block contained three pairs (1-GGP, 1-GP, 1-GG), randomly ordered within the block. One filler catch trial appeared after each block.

For each animal pair, participants were asked to select, from a list of possible inferences warranted by the analogy, the inferences that were "probably true" of the target animal. (See Appendix A for sample stimuli.) For the GGP trials, the list included three relational inferences and the single effect inference. For the GG and GP trials, the list included three relational inferences, the effect inference, and one antecedent inference (e.g., for the GG trials, the "P" antecedent was included in the inference list). Thus, each participant responded to four inferences for each of three GGP trials, five inferences for each of three GG trials and three GP trials. The study was self-paced and administered using MediaLab on PC.

Results

Table 1 shows the mean proportions of inferences endorsed, by target type. Results were analyzed using a one-way repeated-measures ANOVA.

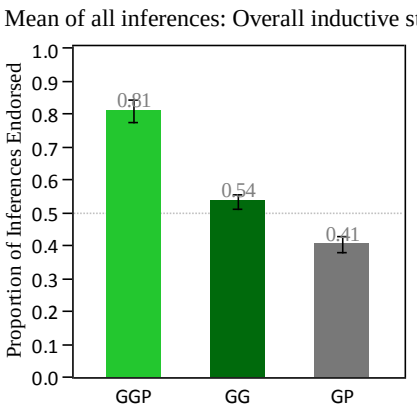
Table 1: Mean proportion of inferences endorsed, for each inference and target type[2]

	GGP	**GG**	**GP**
	Mean (SD)	Mean (SD)	Mean (SD)
Antecedent	n/a	0.0 (0.0)	.045 (.156)
G1-relation	.879 (.318)	.848 (.32.1)	.788 (.334)
G2-relation	.894 (.280)	.848 (.32.1)	.197 (.351)
P-relation	.818 (.367)	.121 (.28.3)	.818 (.321)
Effect	.652 (.430)	.864 (.28.5)	.182 (.321)
Overall Avg	**.811 (.259)**	**53.6 (17.6)**	**40.6 (19.2)**

The mean proportion of endorsements for all inferences within each target type are shown in

Figure 3a. As predicted, this measure of inductive strength replicates the pattern of Similarity ratings in Study 1A. The mean proportion of inferences endorsed differed significantly by target type, $F(2,174)=142.3$, $p<.0001$. Tukey's HSD contrasts ($q=2.36$) showed that this measure was significantly higher for the GGP targets than for the GG targets, and significantly higher for the GG targets than for the GP targets.

Figure 3a: Mean proportion of all inferences endorsed, by target type



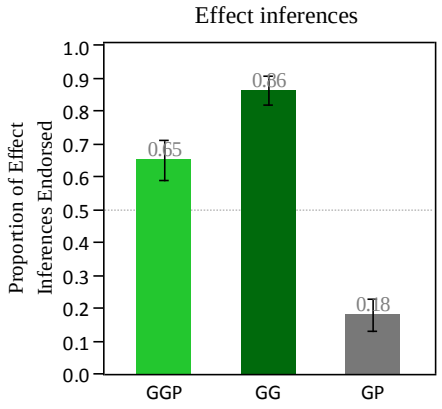
As a consistency check, the mean Effect inferences (shown in Figure 3b) follow the same pattern as in Study 1A.[3] The proportion of effect inferences endorsed differed significantly by target type, $F(2,174)=142.3$, $p<.0001$. Tukey's HSD contrasts ($q=2.35$) showed that the effect inference endorsements were significantly more likely for the

[2] For each target type, the participant had 3 opportunities – in 3 trials – to endorse each type of inference. This table shows the mean proportion of endorsements for each inference type. For example, for the GGP target-type, participants, on average, endorsed approximately 2/3 (0.65) of the Effect Inferences.

[3] The Effect inference for the GP target in Study 1B looks lower than in Study 1A; our explanation for this is that because this task asked for categorical endorsements rather than ratings, the participants whose ratings would have been below 50% (as most of them were in Study 1A) did not endorse the inference.

GG targets than for the GGP or GP targets, and significantly more likely for the GGP targets than for the GP targets.

Figure 3b: Mean proportion of inferences endorsed, for Effect inference, by target type



Discussion

This pattern of overall inductive strength parallels the similarity pattern found in Study 1A. Consistent with structure mapping theory, adding a shared element increased inductive strength, just as it increased similarity.

The inferential strength for the effect inference decreased, as it did in Study 1A. However, assuming that there are multiple goals of analogical inference, there seems no reason to imagine that any one of the individual candidate inferences is a more important measure of inductive strength than the others. Thus, we argue, the mean of all the potential inferences warranted by the analogy is a truer measure of an analogy's inductive strength. The results of this overall measure of inductive strength, taken together with the Similarity results in Study 1A, suggest that inductive strength does not dissociate from similarity.

However, the question remains: if the effect inferences are not fully explained by models of analogy, then how are these causal inferences processed?

According to SMT (Gentner 1983, 1989; Gentner & Markman, 1997), the relations in the base are projected to the target as candidate inferences during the mapping process. Lee & Holyoak's suggestion that analogical inference is used to construct the causal model in the target (see also Gentner, 2001), is consistent with SMT on this point. In the analogies in these studies, the inferred relations are causal, and form the causal model in the target analog. In other words, the relational inferences that participants made in Study 1B constitute the causal model in the target.[4]

We further extend L&H's claim: we argue that by analogical reasoning, the base analog in these studies makes *two* contributions to knowledge about the causal system in the target.

The first contribution, as Lee & Holyoak suggest, is the structure of the causal model. The second contribution the base may make to information about the causal system in the target is essentially a proxy for the posterior probability that E will occur when G1, G2, and P are present [or, p(E|G1, G2, P)]. In models of causal reasoning, prior probabilities (base rates, such as p(G1) – the probability that G1 will occur) and posterior probabilities (such as p(E|G1)) are typically used to calculate the probability of an outcome under various conditions. (Pearl, 2000; Griffiths & Tenenbaum, 2005; see Glymour, 2001 for a discussion of the probability assumptions of several models of causal inference).

In Study 2, we examine these contributions. If the causal model in the target is constructed by analogical mapping, then the causal model constructed in the target should be effectively the same as if the relations were given in the target in the first place. If that model is then used to determine whether the effect is present in the target, then people's effect inference ratings should not differ based on whether the relations are given in the base or target. For example, given an analogy where the base contains two features (e.g., G and P), and the target includes the same two features, the effect inference ratings when the relations are given in the *base* should not differ from effect inference ratings when the relations are given in the *target*. These should also not differ from a no-analog causal inference task, where the effect inference is made about one animal which has the two features and the two relations (and no base is given).

However, if the base contributes information about the effect's prior probabilities, then knowing that the effect is present in the base should lead to higher effect inference ratings than when the effect's status is not known.

STUDY 2: Contributions of Base to Target's Causal System

To examine the contributions that the base analog makes to the causal system in the target, we vary the information given in the base. Specifically, we vary whether a base analog is given, whether the stated causal relations are in the base or target, and whether the relations given in the base are explicitly accompanied by the effect, and ask participants for effect inferences for each of the targets.

[4] In these studies, the stimuli are very sparse, and the target analogs contain no stated relations. The entire causal structure is therefore projected from the base. In general, according to SMT, the relational inferences are a completion of the shared structure, in addition to any other structure (relations) already present in the target. Thus, if the target were to contain its own causal relations, then any relations inferred from the base would be incorporated into the target's existing causal structure. (As a related side note, candidate inferences in general could be projected in either direction; not only from base to target, but also from target to base (e.g., Bowdle & Gentner, 1997))

Our first claim is that the base analog contributes the relations to the target by analogical mapping. Our corresponding prediction is that effect inferences should not differ based on whether the causal relations are given in the base or in the target analog. Furthermore, these inference ratings should also not differ from those when only one animal (the target, and no base) is given.

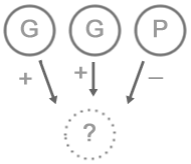
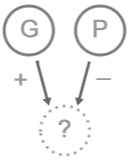
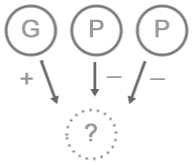
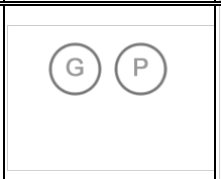
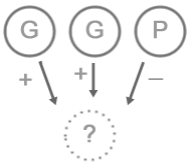
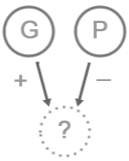
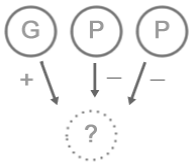
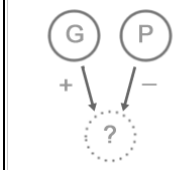
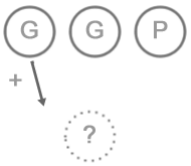
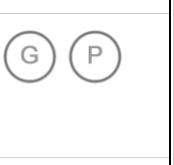
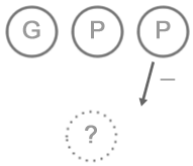
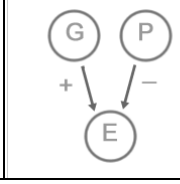
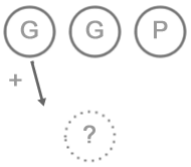
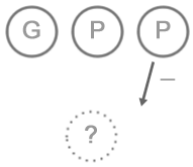
To test this, we created three sets of stimuli. what differs among these sets is the location of the G1 and P1 relations. In the Analogy-BaseRelations condition, the GP relations are given in the base. In Analogy-TargetRelations conditions, the relations are given in the target animal. In the No-Analogy condition, there is only a single animal (no base) which contains all the relations. (These variations are shown in

Table 2.) If the base analog contributes the relations to the target by analogical mapping, then the causal model constructed for each target should be the same for these three groups, and so, for each target, the effect inference ratings should not differ between these groups.

Our second claim is that when the effect is explicitly present in the base analog, the base contributes information about the combined strength of the stated relations (G,P) in producing the effect. This is essentially the posterior probability, p(E|GP). Our corresponding second prediction is that when the effect inference is included in the base, the effect inference ratings for the GP and GPP target-types should be higher than when the effect is not stated to be present (Analogy-BaseRelations, Analogy-TargetRelations, and NoAnalogy conditions). To test this, we created a fourth set of stimuli (Analogy-BaseRelations+E condition) in which the GP relations are given in the base – as in the Analogy-BaseRelations condition – and added an explicit statement that the effect is present in the base analog. (See Appendix A for sample stimuli.)

Thus we predict that, for the GP target e.g., the Analogy-BaseRelations+E group should give higher

Table 2: Stimuli variations for Study 2

Group	BASE	GGP Target	GP Target	GPP Target	
No Analogy	*(no base)*				No analogy; single animal only
Analogy w/Target Relations					Structurally uninformative analogy
Analogy w/Base Relations					Structurally informative analogy
Analogy w/Base Relations +Effect					Structurally informative analogy; Effect present in base

E inference ratings than the Analogy-BaseRelations group.[5]

Method

Participants. Thirty-two undergraduates participated to fulfill a course requirement or for nominal compensation. Participants were randomly assigned to one of four conditions (see Table 2). Four additional participants were excluded for failing the catch trials.

Materials and procedure. Each participant received a set of three descriptions of animals (plus filler items). The animals were adapted from the ones used in Study 1. The Analogy-TargetRelations (n=9),
Analogy-BaseRelations (n=11), and Analogy-BaseRelations+E (n=11) groups received three descriptions of animal pairs (base and target); the No-Analogy group (n=11) received three descriptions of single animals (the target). We created four sets of stimuli. These variations are shown in
Table 2.

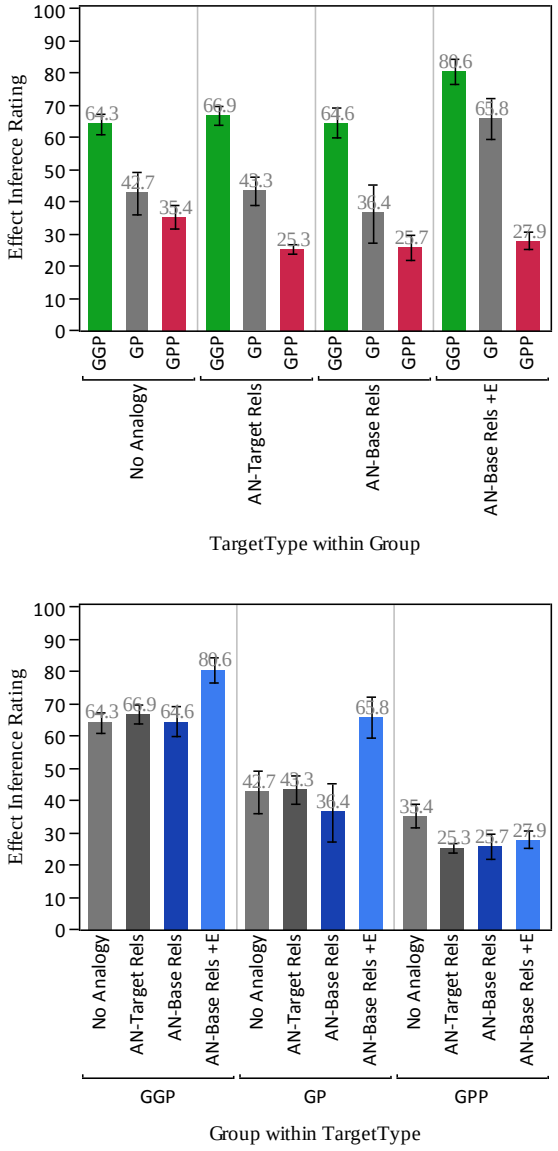
In the first trial, all groups received the GP target; in the second trial, the GGP target; and in the third trial, the GPP target. In each trial, participants were asked to judge the presence of the effect in a target animal *(*E.g., " How likely is it that animal S has scaly skin? (on a scale of 0-100)"). For each target-type, the base and target descriptions varied by group, as shown in Table 2The experiment was self-paced and was administered using MediaLab on PC.

Results and Discussion

Table 3 shows the mean effect inference ratings for each group, by target type. Results were analyzed using two-way ANOVA with repeated-measures on the target-type factor.

To test the first prediction, that within each target type there would be no differences by location of the causal relations, the results for the No-Analogy, Analogy-TargetRelations, and Analogy-BaseRelations groups were analyzed using two-way ANOVA with repeated-measures on the Target-type factor. As predicted, there was no effect of Group, $F(2,56)=.55$, n.s., and no significant interaction, $F(4,56)=.77$, n.s.[6] There was a main effect of Target Type, $F(2,56) = 56.1$, $p<.0001$. Tukey's HSD revealed that the effect inference ratings for the GGP target were significantly higher than for the GP target, and the ratings for the GP target were significantly higher than for the GPP target. Thus, each of these three groups yielded the same patterns of results (see Figure 4).

Figure 4: Mean Similarity and Effect inference ratings, by Target Type and Group



Note that for the GGP and GPP targets, the Analogy-BaseRelations group was given some relations in the base, and some in the target. Although the 'model construction' task involves merging the base's (G,P) relations with the target's relations (e.g., G2), this group's effect inferences for GGP and GPP targets did not differ from the No-Analogy and Analogy-TargetRelations groups. This null result, together with the main effect of target-type, is consistent with our claim that the base's relations are used to construct a causal model in the target, and that that constructed model is used in making causal inferences about the target.

[5] Note that the Analogy-BaseRelations+E group's ratings may also look different from the ratings Study 1 because the base analogs are different: GP+E in this study; GGP+E in the prior study.

[6] Although the Analogy-BaseRels group's ratings for the GP appear to be potentially lower than the No-Analogy () and Analogy-TargetRelations groups, post-hoc t-Tests are not significant, $t(76)=1.0$, and $t(76)=1.1$, respectively. Nevertheless, this potential trend may warrant further exploration. / Altenrative analysis: Welch ANOVA within GP target-type, $F(2,28) =0.3$, n.s.

Table 3: Inference ratings by condition and target type

		GGP		GP		GPP	
Condition	N	Mean (SD)	vs. Chance (50)	Mean (SD)	vs. Chance (50)	Mean (SD)	vs. Chance (50)
No Analogy	11	64.6 (14.9)	t(10)=4.6, p<.01	36.4 (29.4)	t(10)=1.1, n.s.	25.7 (12.9)	t(10)=3.8, p<.01
AN-Target Rels	9	66.9 (8.6)	t(8)=5.9, p<.001	43.3 (13.2)	t(8)=1.5, n.s.	25.3 (4.8)	t(8)=15.3, p<.001
AN-Base Rels	11	64.3 (10.4)	t(10)-3.3, p<.01	42.7 (21.5)	t(10)=1.5, n.s.	35.4 (12.8)	t(10)=6.3, p<.001
AN-Base Rels+E	11	80.6 (12.9)	t(10)=7.9, p<.001	65.8 (21.4)	t(10)=2.5, p<.05	27.9 (9.0)	t(10)=8.1, p<.001
all groups	42	69.2 (13.6)		47.2 (24.5)		28.7 (11.0)	

In short, this result is consistent with the hypothesis that a causal model constructed by analogy (in the Analogy-BaseRelations group) is equivalent to a causal model explicitly given in the target analog (as in the Analogy-TargetRelations group) and to a causal model presented in a target without a base analog (No-Analogy group).

To test the second prediction – that the explicit presence of the effect in the base would increase effect ratings in the target – all four groups were included in a two-way ANOVA with repeated-measures on the Target-type factor. As predicted, when the Analogy-BaseRelations+E group was included in the analysis, the Group by Target-type interaction was significant, $F(6,76)=2.8$, $p<.05$. There was also a significant effect of Group, $F(3,76)=4.4$, $p<.01$, and a significant effect of Target Type, $F(2,76)=83.5$, $p<.0001$.

Planned simple-effects contrasts revealed that within the GP target-type, as predicted, the Analogy-BaseRelations+E group rated the effect inference higher than did the Analogy-BaseRelations, Analogy-TargetRelations, and NoAnalogy groups; and, as predicted, there were no differences between the latter three groups. (Furthermore , when compared with chance (50%), only the Analogy-BaseRelations+E group is significantly above chance for the GP target.) Within the GGP target-type, planned contrasts again showed that the Analogy-BaseRelations+E group rated the effect inference higher than the other groups, which did not differ from one another. Within the GPP target-type, there were no differences between any of the groups.

These patterns suggests that the interaction is driven entirely by the Analogy-BaseRelations+E group. This is consistent with our hypothesis that the presence of E in the base provides a clue to the combined strength of the G & P causal factors (i.e., the posterior probability, p(E|G,P).

The finding that the effect inference ratings differ only to the extent that E is explicitly present in the base – and not based on the location of the relations, or indeed on whether an analogy is performed at all – supports our claim that, by analogical mapping, the base makes two specific contributions to knowledge about the causal model in the target (i.e., causal structure and a clue to combined causal strengths).

STUDY 3: Computational Simulation of "Outsourcing"

As discussed earlier, Lee & Holyoak claim that "the missing theoretical mechanism for dynamic inference evaluation cannot be simply outsourced to some postanalogical module" (2008b, p1121). We argue that the evaluation of the effect inference can indeed be "outsourced" to a post-analogical process. As proof of concept, we used the stimuli from Study 1 (which have the same structure as the stimuli used in L&H's Study 1, 2008b) as input to SME (the computational model of structure-mapping theory), and then applied a simple algorithm to SME's output to simulate the post-analogical inference evaluation.

Method

The mapping process has been operationalized in the Structure Mapping Engine (SME; Falkenhainer, Forbus & Gentner, 1989), a computational model that instantiates Gentner's (1983) Structure-mapping theory. This system operates in a local to global fashion, first finding all possible local matches between the elements of two potential analogs. It combines these into structurally consistent clusters, and then combines the clusters (called kernels) into the largest and most deeply connected system of matches. Other propositions connected to the common system in one analog become candidate inferences about the other analog. Each of these candidate inferences receives a support score. Finally, SME computes a structural evaluation score estimating the systematicity of the structural match (see Forbus, Gentner & Law, 1995).

For this simulation, we created propositional representations of the base and target stimuli used in Study 1A. Using blank features, one base item and three target items were created, to form three pairs of animals (GGP, GG, and GP). SME was used to create mappings of these pairs. When multiple mappings were generated for a pair, we selected the one with the highest structural evaluation score (SES). In each mapping, SME computes candidate inferences. (These inferences included relational and attribute inferences, and were conceptually similar to those in

the lists used in Study 1B.) For each candidate inference, SME generated a support score, reflecting the degree of support that the analogy provides for the inference.

In the second phase of the simulation, we used these candidate inference support scores as input for a simple algorithm to calculate the effect inference. This phase essentially represents the causal model that is constructed by the mapping, and that is used to evaluate the effect inference. Conceptually, the algorithm gives the proportion of the total causal forces that produce the effect. Specifically, it uses the number of generative causal relations divided by the total number of (generative and preventative) causal relations to determine the probability of the effect. One way of thinking about this ratio is as votes – it's the proportion of the total votes that are in favor of the effect.

$$(1) \qquad \frac{G}{G+P} = E$$

For each candidate inference, we take SME's support scores for the candidate inferences and enter them into the equation, so the equation for these stimuli becomes:

$$(2) \qquad \frac{G_1 + G_2}{G_1 + G_2 + P} = E$$

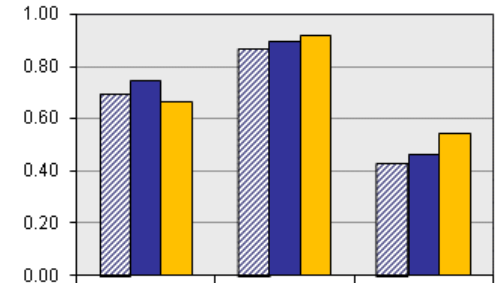
This evaluation algorithm uses SME's support scores to estimate the final effect inference rating for each analog pair. In this way, we use the causal model constructed by the analogical mapping (i.e., the candidate inferences) to evaluate the Effect inference.

Results and Discussion

The results of the computational simulation of Study 1 are shown in Figure 5. As predicted, the results of the simulation closely match the human data, both from our Study 1A and from L&H's Experiment 1 (2008b).[7] These results demonstrate that the inference evaluation can indeed be outsourced to a post-analogical process, and that a two-phase process simulation using SME followed by a simple causal calculation can closely match the human data.

We do not claim that this simple equation is necessarily the precise evaluation algorithm that people use; other, more complex algorithms may yield equivalent results.[8] We only argue that this existence proof supports our claim that causal inference evaluations are handled by a post-analogical process. Clearly, many further simulations may be run to further test this claim.

Figure 5: Effect inferences for Simulation, compared with human data from Exp. 1a and L&H (2008b, Exp1).



	GGP	GG	GP
HUMAN - L&H	0.694	0.866	0.425
HUMAN - Exp 1a	0.744	0.894	0.461
SIMULATION	0.667	0.917	0.542

GENERAL DISCUSSION

Three studies addressed the questions of how causal analogies are processed. In Study 1, we tested structure mapping theory's prediction that the overall inductive strength of causal analogies parallels the similarity ratings of analogous pairs. We found that although the effect inference dissociates from similarity (replicating Lee & Holyoak, 2008b), the overall inductive strength of the analogy – consistent with structure-mapping theory – does follow the same pattern as the similarity ratings.

In Study 2, we examined the claim that in causal analogies, the base makes two particular contributions to the causal system in the target. The results suggest that the effect inferences made from a causal analogy do not differ from those made from a single example, except to the extent that the causal analogy may provide a clue to the conditional probability of the effect, given the causal antecedents ([p(E|G,P)].

Taken together, these findings are consistent with / support our claims that (1) the causal model in the target analog is constructed by analogical inference, and that (2) the base contributes information about the combined strength of the causal factors in producing the effect.

Study 3 tested the prediction that a computational simulation using SME (which implements structure-mapping theory), followed by a post-analogical algorithm, can match human effect inferences. The results of this simulation, bolstered by the results of Studies 1 and 2, support our claim that inference evaluation can occur post-analogically.

These findings are important for a few reasons. First, they support the hypothesis that analogical reasoning provides an important method for learning about novel systems, and particularly for understanding the causal structure of a novel system. Second, these findings are consistent with the

[7] We also ran this simulation on Lee & Holyoak's stimuli from their Exp. 2 (2008b), with similar results. Reporting those results is beyond the space constraints of this paper.

[8] In fact, a more complex algorithm could make better use of the posterior probability proxy described in Study 2.

predictions of SMT, that similarity is an important contributor to overall inferential strength.

Our assertion is that analogy does not explain everything, nor should it. If other reasoning processes explain causal inferences adequately, even when reasoning from causal analogies, there's no parsimonious reason to suppose that analogical processing models should be adapted to do their job. In sum, we maintain that analogy is important for learning about novel causal systems, but models of analogy need not subsume causal inferencing processes.

ACKNOWLEDGMENTS

We thank Ken Forbus, Scott Friedman, Katherine James, Jason Jameson, Kate Lockwood, and all members of the Cognition & Language Lab. This research was funded by Department of Education, IES grant R305B040098 to the first author and ONR Grant N00014-02-1-0078 to the second author.

REFERENCES

Blok, S. & Gentner, D. (2000). Reasoning from shared structure. *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society*, 621-626.

Bowdle, B., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology, 34*(3), 244-286.

Clement, C . A., & Gentner, D . (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15*, 89-132.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1–63.

Forbus, K. D., Gentner, D., &Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141–205.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). London: Cambridge University Press.

Gentner, D. (2001). Exhuming similarity. Commentary on Tenenbaum, Joshua B. & Griffiths, Thomas L. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 669.

Gentner, D., & Kurtz, K. (2006). Relations, objects, and the composition of analogies. *Cognitive Science, 30*, 609-642.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.

Glymour, C. (1002). *The mind's arrows: Bayes nets and graphical causal models in psychology.* Cambridge, MA: MIT Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354-384.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295–355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure. *Psychological Review, 104,* 427–466.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 754-770.

Lee, H. S., & Holyoak, K. J. (2007). Causal models guide analogical inference. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 1205-1210.

Lee, H. S., & Holyoak, K. J. (2008a). Absence Makes the Thought Grow Stronger: Reducing. Structural Overlap Can Increase Inductive Strength. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 297-302.

Lee, H. S. and Holyoak, K. J. (2008b). The Role of Causal Models in Analogical Inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1111–1122

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science, 21*, 373–418.

Markman, A. B., & Gentner, D. (2000). Structure-mapping in the comparison process. *American Journal of Psychology, 113,* 501–538.

Pearl, J. (2000). *Causality: models, reasoning, and inference.* New York: Cambridge University Press, 1-40.

Wu, M., & Gentner, D . (1998). Structure in category-based induction. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 1154-1158.

APPENDIX A: SAMPLE STIMULI

Study 1

GGP target-type:

Animal A has enzyme aliesterase, neurotransmitter tyrosine, hormone TSH, and exceptional hearing.
For animal A, enzyme aliesterase tends to cause exceptional hearing;
neurotransmitter tyrosine tends to cause exceptional hearing;
and hormone TSH tends to prevent exceptional hearing.

Animal B has enzyme aliesterase and neurotransmitter tyrosine.

Study 1A, similarity:
How similar are animals A and B? (on a scale of 0-10)

Study 1A, inference:
For animal B, what percentage have exceptional hearing? (on a scale of 0-100)

Study 1B, multiple inferences:
Which of the following are probably true of Animal B? Please check all that apply.
[ ] Enzyme aliesterase tends to cause exceptional hearing
[ ] Neurotransmitter tyrosine tends to cause exceptional hearing
[ ] Hormone TSH tends to prevent exceptional hearing
[ ] Has exceptional hearing

Study 2

GPP target-type for the Analogy-BaseRelations+E group:

Animal R has blocked oil glands, filaggrin protein, and scaly skin.
For animal R, blocked oil glands tend to cause scaly skin, and filaggrin protein tends to prevent scaly skin.

Animal S has blocked oil glands, filaggrin protein and a marker chromosome.

For animal S, a marker chromosome tends to prevent scaly skin.

How likely is it that animal S has scaly skin? (on a scale of 0-100)

Study 3

Base and GGP case representations used as input to Structure Mapping Engine.

```
(Case Base)
(knownSentence (isa a1 Animal))

(knownSentence (isa a1 f1))
    ; f1 is an attribute of a1.
(knownSentence (isa a1 f2))
(knownSentence (isa a1 f3))
(knownSentence (isa a1 e1))

(knownSentence (causes-Underspecified
       (isa a1 f1) (isa a1 e1)))
; A1 having feature F1 causes A1 to
have feature E1.
(knownSentence (causes-Underspecified
       (isa a1 f2) (isa a1 e1)))
(knownSentence (prevents-Underspecified
       (isa a1 f3) (isa a1 e1)))


(Case Target_GGP)
    (knownSentence (isa a2 Animal))
    (knownSentence (isa a2 f1))
    (knownSentence (isa a2 f2))
    (knownSentence (isa a2 f3))
```