Blok, S. V., & Gentner, D. (2000, August). Reasoning from shared structure. Poster session presented at the Twenty-Second Annual Conference of the Cognitive Science Society, Philadelphia, PA.

Reasoning from shared structure

Sergey Victor Blok (s-blok@northwestern.edu)
Dedre Gentner (gentner@northwestern.edu)

Department of Psychology, Northwestern University,
Evanston, Illinois

Abstract

Two experiments contrasted the predictions of the similarity-coverage model of category-based induction with those of a structure-based account. We focused on the two theories' ability to account for the paradoxical fact that both monotonicities (increases in argument strength with the addition of premises) and non-monotonicities (*decreases* in argument strength with addition of premises) occur in human reasoning. The results are mainly in accord with the structure-based account and are inconsistent with the similarity-coverage account.

Introduction

Monotonicity and Induction

Humans routinely make inductive inferences, and the principles that guide these inferences have received a great deal of empirical attention (López, 1995; McDonald, Samuels & Rispoli, 1996; Osherson, Smith, Wilkie López & Shafir, 1990; Sloman, 1993). One principle that has both intuitive and empirical support is *monotonicity* – the principle that confidence in an inductive inference should increase with the number of supporting premises. For example, Osherson et al. showed that adults preferred Argument B over Argument A.

A. All FOXES have sesamoid bones,
All PIGS have sesamoid bones,
Therefore, all GORILLAS have sesamoid bones

B. All FOXES have sesamoid bones,
All PIGS have sesamoid bones,
All WOLVES have sesamoid bones
Therefore, all GORILLAS have sesamoid bones.

However, robust *nonmonotonicities* have also been documented. Osherson et al.'s participants chose Argument C over D.

C. All FLIES have sesamoid bones,
Therefore, all BEES have sesamoid bones.

D. All FLIES have sesamoid bones,
All ORANGUTANGS have sesamoid bones,
Therefore, all BEES have sesamoid bones.

Sloman (1993) and McDonald et al. (1996) have also documented nonmonotonic responding in adults. Even more strikingly, Lopez, Gelman, Gutheil & Smith (1992) showed nonmonotonicity effects very early in development; in fact, nonmonotonicity effects were reliably obtained earlier than monotonicity effects. People appear to believe that more premises make for a stronger argument, except when more premises make for a weaker argument. How can we reconcile these apparently contradictory phenomena?

Similarity-Coverage Model

A pioneering theory of argument strength is the Similarity-coverage model (SCM) of Osherson et al. (1990). The two components of SCM are *similarity* -- the extent of feature overlap between premise and conclusion categories -- and *coverage* -- the average similarity of the premises and the instances of the lowest level taxonomic category that includes both the premises and the conclusion. The similarity-coverage model predicts monotonicity when the additional premise is a member of the same lowest level superordinate category as the initial premises and the conclusion. It predicts nonmonotonicity when the additional premise is not a member of the lowest level superordinate category. Thus nonmonotonicity can be seen as a kind of dilution effect, as illustrated by Osherson et al's (1990) data in (1) and (2), respectively.

(1) a. ROBINS, *SPARROWS* / SEAGULLS >
b. ROBINS / SEAGULLS[1]

(2) a. ROBINS, *RABBITS* / SEAGULLS <
b. ROBINS / SEAGULLS

Argument (1) is monotonic; adding the extra premise *SPARROW* in (1a) adds an additional piece of premise support without diluting the category coverage, because it fits within the lowest-level category (BIRDS) that applies in the single premise case (1b). In contrast, the additional premise *RABBITS* in (2a) raises the lowest-level common category to ANIMALS, thus diluting the category coverage. Thus the SCM can successfully predict some instances of monotonicity.

[1] Research in this area typically uses so called "blank" or opaque properties – such as 'has sesamoid bones' to ensure that belief in the conclusion is derived from the premise statements, rather than from prior beliefs about the truth of the conclusion. We will omit property names from further examples.

However, as Sloman (1993) noted, there are other instances of nonmonotonicity that are not explainable by dilution of category coverage. His participants found (3b) to be stronger than (3a).

(3) a. CROCODILE, KINGSNAKE / ALLIGATOR
b. CROCODILE / ALLIGATOR

Even though the lowest level taxonomic category (REPTILE) does not change across these arguments, nonmonotonicity[2] is observed. Sloman acknowledges, however, that his own feature-based induction theory is also unable to explain nonmonotonicities.

Structure-Based Induction

We propose a *structure-based induction* approach that uses structural overlap instead of overall similarity or feature overlap to predict argument strength. Our model is very different from the previous theories in that we explicitly assume that the evaluation of argument strength is accomplished by a process of aligning the representations of the premise(s) and the conclusion.

Specifically, we assume that the perceived strength of an induction from premise to conclusion depends on the goodness[3] of the common schema. For the one-premise case, this idea is closely related to similarity in Osherson et al.'s account and with feature overlap in Sloman's account. But when there are multiple premises, we postulate a *premise comparison process* whereby a common schema is derived from the premises. This schema is then aligned with the representation of the conclusion statement.

This variant of the *progressive alignment hypothesis* (Kotovsky and Gentner, 1996; Kuehne, Gentner & Forbus, 2000; Kuehne, Forbus, Gentner & Quinn, 2000) states that carrying out a comparison involves alignment of structured representations (e.g. Gentner & Markman, 1997).

There is evidence that structure-mapping theory captures some important aspects of inductive reasoning. Wu and Gentner (1998) told participants that a conclusion had attribute a_1. They were also told that two different premise kinds P_1 and P_2 also had a_1. Participants were then given the option of inferring an attribute form P_2 that was causally connected to a_1 or an attribute from P_1 that was not causally connected to a_1. Results indicated that people strongly preferred to reason from a causal base (P_2) over an attribute base (P_1). See Clement & Gentner (1996) and Lassaline (1996) for related findings.

The SBI view makes several specific predictions. First, it predicts that *monotonicity* (in at least the weak sense) will result when the additional premise is alignable with the other premises and the conclusion. Second, conversely, *nonmonotonicity* should result when the additional premise is not alignable with the premises (even if it is alignable on other grounds with the conclusion).

These two assertions predict the monotonicity of argument (1) and the nonmonotonicity of (2). A further point is that the predictions of the SBI model do not rely on taxonomic category structure. Neither monotonicity nor nonmonotonicity are influenced by whether the additional premise belongs to the lowest common category that includes the premises and the conclusion. Thus SBI explains Sloman's example (3) above by noting that the goodness of alignment between the premise CROCODILE and the conclusion ALLIGATOR is diminished by first aligning CROCODILE with KINGSNAKE.

The third prediction of SBI is that the properties inferred depend on the particular aligned schema. That is, people base their inferences (even about nominally blank properties) on the specific alignment between premises and conclusion, and not on a general sense of similarity. Because the quality of the premise-conclusion alignment determines both the specific properties people are willing to infer *and* the argument strength, we expect a strong association between these two (see Heit & Rubinstein, 1994, for a related proposal).

Experiment 1. Two vs. three premises

In this experiment, we varied category coverage and alignability in order to contrast the predictions of the similarity-coverage model and the structure-based approach. We used five variants of each argument: a two-premise item plus four kinds of additional premises that were added to make three-premise arguments (Table 1).

Table 1. Sample base two-premise item and the additional premise in the four variant conditions in Experiment 1.

ROBINS, EAGLES, ... / BATS

Coverage

Alignment	C+	C–
A+	**SEAGULLS**	**AIRPLANES**
A–	**DOGS**	**TV's**

The premises and the conclusion of the two-premise arguments shared a common relational schema, such as *flight* or *underwater habitat*. The three-premise arguments were constructed by adding an additional premise to the two-premise arguments. There were four types of additional premises, constructed according to a 2x2 design of *alignability* with the two-premise schema and *category coverage* – i.e., whether the additional premise belonged to the lowest level category spanning the two premises

[2] Monotonicity can be interpreted in the strong sense of *increasing monotonicity* or in the weaker sense of *non-decreasing monotonicity*. Note that even the latter, weaker sense is violated by these examples.

[3] We will use the term *goodness* of the common schema as a shorthand for *structural evaluation*; it depends on the size and depth of the common schema.

and the conclusion (hereinafter abbreviated spanning *category*).

For example, given the two-premise argument ROBIN, EAGLE / BAT, the aligned schema presumably involves flight and the spanning category is ANIMAL. The four kinds of additional premises are as follows:

1. **A+C+** type: Alignable with the 2-premise schema (*High Alignment*) and a member of the lowest-level spanning category (*High Coverage*).
   e.g., ROBINS, EAGLES, *SEAGULLS* / BATS
2. **A–C+** type: Not alignable with the 2–premise schema (*Low Alignment*), but a member of the spanning category (*High Coverage*). [4]
   e.g., ROBINS, EAGLES, *DOGS* / BATS
3. **A+C–** type: Alignable with the 2–premise schema (*High Alignment*), but not a member of the spanning category (*Low Coverage*).
   e.g., ROBINS, EAGLES, *AIRPLANES* / BATS
4. **A–C–** type: Not alignable with the 2–premise schema (*Low Alignment*), nor a member of the spanning category (*Low Coverage*).
   e.g., ROBINS, EAGLES, *TV'S* / BATS

Method

37 Northwestern University undergraduates were presented with 40 inductive arguments, one at a time on a computer, and asked to rate them according to "how well the conclusion follows from the premises." There were eight sets, each with five argument types (8 two-premise arguments plus 4 x 8 = 32 three-premise arguments).

For example,

Fact:
All ROBINS have property F.
All EAGLES have property F.
Therefore,
All BATS have property F.

After rating all the arguments, participants were given a printed packet with the forty arguments they had just rated and were asked to write down their best guess about the property associated with each argument. They were also given the option of skipping any items for which no property had come to mind.

Predictions

Table 2 summarizes the predictions of the two models. The structure-based induction model predicts monotonicity for alignable types (A+C+ and A+C–) and nonmonotonicity for non-alignable types (A–C+ and A–C–), relative to the two-premise arguments. The similarity-coverage model predicts monotonicity for high coverage types (A+C+ and A–C+), and nonmonotonicity for low coverage types (A+C– and A–C–).

Table 2. Summary of predictions of the two models

Theory	Prediction
SCM	**A+C–, A–C– ≤ 2P ≤ A+C+, A–C+**
SBI	**A–C+, A–C– ≤ 2P ≤ A+C+, A+C–**

Another line of prediction concerns the subjects' guesses about the blank properties. According to the structure-based view, the same process of structure-mapping that gives rise to the goodness of the common schema also gives rise to its specific content. Thus we predict (1) people's *confidence* in their property guesses will increase with their subjective argument strength; (2) the *uniformity* of property guesses will increase with their subjective argument strength; and (3) both the confidence and the uniformity of property guesses will be greater for alignable types than for non-alignable types.

Results

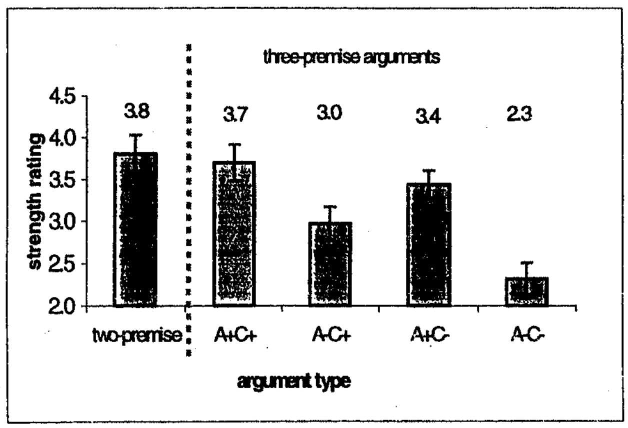

Figure 1. Argument strength ratings for five argument types in Experiment 1 (error bars are 95% confidence intervals).

Argument Strength Ratings.[5] Figure 1 shows the mean ratings across items. As predicted by the structure-based account, monotonicity (in the weak, though not the strong form) held when the additional premise was alignable. That is, there were no significant differences in judged strength between two-premise arguments (M = 3.81; SD = 1.31) and either the A+C+ type (M = 3.70; SD = 1.28) or the A+C– type (M = 3.44 ; SD = 1.25) , $t(36) = 1.46$, $p >$ .008, $t(36) = 2.49$, $p >$ .008 respectively. Also as predicted, nonmonotonicity held when the additional premise was nonalignable. Arguments of the A–C+ type (M =

[4] The extra premise for the A-C+ type always belonged to the same superordinate as the conclusion. This had the effect of giving the A-C+ type the highest relative coverage of any of the 3-premise arguments, as defined by the similarity-coverage model. Importantly, the A-C+ type had higher coverage than the A+C+ type, providing a very strong test of the alignment model against the coverage model.

[5] We performed six planned comparisons on the mean argument strength for each subject within a type, setting the two-tailed Bonferroni corrected alpha value at 0.008.

2.97; SD = 1.15) and the A–C– type (M = 2.32; SD = 1.14) were rated reliably lower than two-premise arguments, $t(36) = 4.34$, $p < .008$, $t(36) = 5.96$, $p < .008$, respectively.

There were no significant differences based on category coverage. In the crucial comparison of the two models, we found that A+C– arguments were rated reliably stronger than the A–C+ type, $t(36) = 3.01$, $p < .008$, suggesting that alignability, not category coverage, best predicts the effects of adding a third premise to a two-premise argument.

Property Guesses. To test the relation between argument strength and likelihood of listing a property (confidence), we scored the listings on whether a participant chose to guess a property. There were 1241 guesses and 439 (26% of the total) "no guess" responses. The highest proportion of guesses was elicited by the two-premise and the A+C+ argument types (94% and 90%, respectively). The A+C– argument type also elicited a high proportion of guesses (85%). The A–C+ and A–C– types elicited substantially fewer property guesses (65% and 36%, respectively). Overall, the proportion of property guesses closely mirrored the argument strength ratings, $r = .82$, $p < 0.0001$.

To test our predictions concerning property uniformity, we rated the content of the property guesses. When subjects were presented with alignable arguments (i.e., two-premise, A+C+ or A+C– types), subjects almost unanimously provided guesses specific to the hypothesized common schema. When presented with a non-alignable argument, subjects tended to provide general and haphazard guesses and tended to disagree about the nature of the blank property. To test this intuition, we asked two naive raters to score the property listings on the basis of *coherence*. Confirming our hypothesis, alignable arguments elicited highly focused patterns of property guesses, while non-alignable ones displayed little agreement between subjects, as observed by our independent raters. Mean coherence rating across the forty different arguments were correlated with argument strength at $r = 0.599$, $p < 0.0001$.

Discussion

The results of Experiment 1 largely bear out the predictions of the structure-based induction model. The effect of adding a premise to a two-premise argument depends entirely on whether the third premise is alignable with the schema that holds in the two-premise argument. If the third premise is alignable, the argument strength remains constant; if the third premise is nonalignable, the argument strength decreases. The predictions of the similarity-coverage model were not borne out for either monotonicity or nonmonotonicity. The SCM predicts monotonicity if the third premise belongs to the lowest-level spanning category of the two-premise argument; and nonmonotonicity when the third premise forces an increase in the level of the spanning category. Neither prediction held.

The most direct contrast between the models is to compare A–C+ items (low alignability but high coverage) with A+C– items (high alignability but low coverage). Participants found the A–C+ premise sets to be a far weaker inductive base than A+C–, with its specific schema. For example, argument (4a) was weaker than (4b):

(4a) ROBINS, EAGLES, DOGS / BATS

(4b) ROBINS, EAGLES / BATS

Thus increasing in the number of premises even while holding coverage constant can result in nonmonotonicity if the alignment is diminished. Indeed, (4c) is judged stronger than (4a), despite clearly having poorer coverage

(4c) ROBINS, EAGLES, AIRPLANES / BATS

In short, our nonmonotonicity findings support the claims of the structure-based framework over those of the coverage model.

The property guess findings were also consistent with the predictions of the structure-based framework. There was a strong connection between considering an argument strong and having a clear idea of what property was being inferred. This observation is consistent with our claim that the process at work here is an alignment process that results in a specific common schema.

Overall the results are encouraging. However, one point requires discussion. We found evidence of *nondecreasing* monotonicity but not of *increasing* monotonicity. There was no *increase* in argument strength for any argument type. This contrasts with Osherson et al.'s (1990) report that strength increased from two- to three-premise arguments. We suspect much of the difference stems from the fact that, whereas we used a single-argument rating task, Osherson et al. used a choice task. Comparing arguments to choose the stronger could have led to heightened contrast between the two- and three-premise arguments.

Structure-mapping does not predict a steady increase in argument strength as additional premises are added.[6] However, it does predict an increase when going from one-premise to two-or-more-premise arguments (always provided the added premise(s) are alignable), because alignment highlights the common structure (Gentner & Wolff, 2000). To test this prediction, we asked subjects to rate single-premise arguments matched to the multi-premise arguments used in Experiment 1. This will allow us to compare (albeit across experiments) the strength of one-premise vs. three-premise arguments.

A second motivation for Experiment 2 was to rule out a possible confound, namely, that the gain in strength for the additional premises was simply due to an increase in overall similarity (or feature overlap, on Sloman's (1993) account) brought about by the additional premise, rather than by interactions among the premises as claimed by the structural account.

[6] This is because progressive alignment cannot *increase* the size of the common schema. Thus if increases in argument strength do occur when, say, 11 premises are increased to 12, the explanation must lie with other factors beyond alignment.

and the conclusion (hereinafter abbreviated _spanning category_).

For example, given the two-premise argument ROBIN, EAGLE / BAT, the aligned schema presumably involves flight and the spanning category is ANIMAL. The four kinds of additional premises are as follows:

1. **A+C+** type: Alignable with the 2-premise schema (*High Alignment)* and a member of the lowest-level spanning category (*High Coverage*).
   e.g., ROBINS, EAGLES, *SEAGULLS* / BATS
2. **A–C+** type: Not alignable with the 2–premise schema (*Low Alignment*), but a member of the spanning category (*High Coverage*).[4]
   e.g., ROBINS, EAGLES, *DOGS* / BATS
3. **A+C–** type: Alignable with the 2–premise schema (*High Alignment*), but not a member of the spanning category (*Low Coverage*).
   e.g., ROBINS, EAGLES, *AIRPLANES* / BATS
4. **A–C–** type: Not alignable with the 2–premise schema (*Low Alignment*), nor a member of the spanning category (*Low Coverage*).
   e.g., ROBINS, EAGLES, *TV'S* / BATS

Method

37 Northwestern University undergraduates were presented with 40 inductive arguments, one at a time on a computer, and asked to rate them according to "how well the conclusion follows from the premises." There were eight sets, each with five argument types (8 two-premise arguments plus 4 x 8 = 32 three-premise arguments).

For example,

Fact:
All ROBINS have property F.
All EAGLES have property F.
Therefore,
All BATS have property F.

After rating all the arguments, participants were given a printed packet with the forty arguments they had just rated and were asked to write down their best guess about the property associated with each argument. They were also given the option of skipping any items for which no property had come to mind.

Predictions

Table 2 summarizes the predictions of the two models. The structure-based induction model predicts monotonicity for alignable types (A+C+ and A+C–) and nonmonotonicity for non-alignable types (A–C+ and A–C–), relative to the two-premise arguments. The similarity-coverage model predicts monotonicity for high coverage types (A+C+ and A–C+), and nonmonotonicity for low coverage types (A+C– and A–C–).

Table 2. Summary of predictions of the two models

Theory	Prediction
SCM	**A+C–, A–C– ≤ 2P ≤ A+C+, A–C+**
SBI	**A–C+, A–C– ≤ 2P ≤ A+C+, A+C–**

Another line of prediction concerns the subjects' guesses about the blank properties. According to the structure-based view, the same process of structure-mapping that gives rise to the goodness of the common schema also gives rise to its specific content. Thus we predict (1) people's *confidence* in their property guesses will increase with their subjective argument strength; (2) the *uniformity* of property guesses will increase with their subjective argument strength; and (3) both the confidence and the uniformity of property guesses will be greater for alignable types than for non-alignable types.

Results

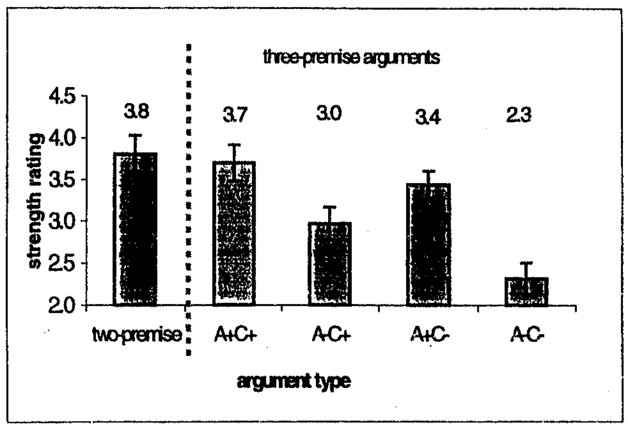

Figure 1. Argument strength ratings for five argument types in Experiment 1 (error bars are 95% confidence intervals).

Argument Strength Ratings.[5] Figure 1 shows the mean ratings across items. As predicted by the structure-based account, monotonicity (in the weak, though not the strong form) held when the additional premise was alignable. That is, there were no significant differences in judged strength between two-premise arguments (M = 3.81; SD = 1.31) and either the A+C+ type (M = 3.70; SD = 1.28) or the A+C– type (M = 3.44 ; SD = 1.25) , $t(36) = 1.46$, $p > .008$, $t(36) = 2.49$, $p > .008$ respectively. Also as predicted, nonmonotonicity held when the additional premise was nonalignable. Arguments of the A–C+ type (M =

[4] The extra premise for the A-C+ type always belonged to the same superordinate as the conclusion. This had the effect of giving the A-C+ type the highest relative coverage of any of the 3-premise arguments, as defined by the similarity-coverage model. Importantly, the A-C+ type had higher coverage than the A+C+ type, providing a very strong test of the alignment model against the coverage model.

[5] We performed six planned comparisons on the mean argument strength for each subject within a type, setting the two-tailed Bonferroni corrected alpha value at 0.008.

Experiment 2. Single-premise arguments

Participants evaluated single-premise arguments. For each argument, the premise was the *additional premise* used in Experiment 1. For example, for the BATS item in Table 1, the four arguments tested in Experiment 2 were

(A+C+)′[7] SEAGULLS / BATS
(A–C+)′ DOGS / BATS
(A+C–)′ AIRPLANES / BATS
(A–C–)′ TELEVISIONS / BATS

The first question is whether, as predicted by structure-mapping, single-premise arguments will be weaker than their three-premise alignable counterparts in Experiment 1. The second question is whether the relative strengths of the three-premise arguments in Experiment 1 are mirrored by the strengths of the corresponding single premises (Thus undermining our premise-comparison account.)

Method

16 Northwestern University undergraduates saw 32 single-premise arguments (8 items x 4 types) and rated them for strength. The procedure was identical to that in Experiment 1, except that the arguments were given in printed form, rather than on a computer.

Results

We contrasted the mean argument strengths by argument type between Experiments 1 and 2.[8] Figure 2 presents the mean strength ratings across argument types.

As predicted by structural framework, among alignable types, there was a reliable advantage for three-premise over one-premise arguments (a difference of 1.28, $t\,(51) = 3.90$, $p < 0.001$). For non-alignable types, this difference was 0.53, $t\,(51) = 1.84$, $p > 0.05$, n.s. Also, as predicted, planned comparisons within alignable types revealed reliable differences between the three-premise A+C+ ($M = 3.70$, $SD = 1.28$) and the single-premise (A+C+)′ types ($M = 2.49$, $SD = 0.88$), $t(51) = 3.43$, $p < 0.005$. A reliable contrast was also observed between the three-premise A+C– ($M = 3.44$, $SD = 1.25$) and the single-premise (A+C–)′ type ($M = 2.08$, $SD = 0.85$), $t(51) = 3.95$, $p < 0.001$.

Planned contrasts for the non-alignable types revealed a non-reliable difference between the three-premise A–C+ type ($M = 2.97$; $SD = 1.15$) and the single-premise (A–C+)′ type ($M = 2.85$; $SD = 1.22$), $t(51) = 0.34$, $p > .70$, n.s.

So far, the results are consistent with the structural account. However, a reliable difference was also observed between the three-premise A–C– type ($M = 2.32$; $SD = 1.14$) and the single-premise (A–C–)′ type ($M = 1.38$; $SD = 0.54$), $t(51) = 3.15$, $p < 0.005$. This result is not predicted by the structural account.

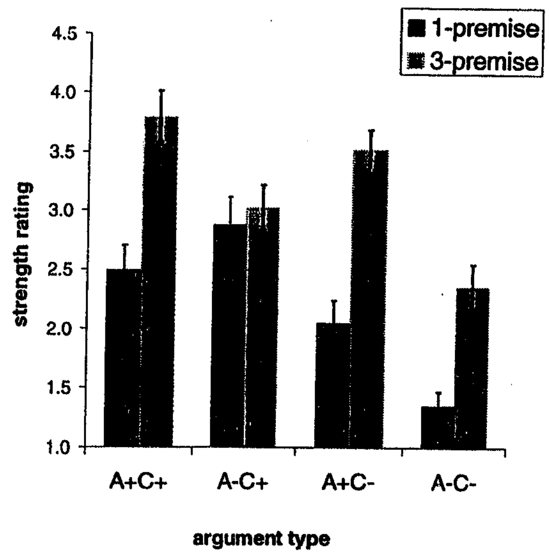

Figure 2. Argument strength ratings for four argument types in Experiment 2 and Experiment 1 (Error bars are 95% confidence intervals).

Turning to the second question, we found that the pattern of strength among single-premise arguments could not account for the three-premise results in Experiment 1. Indeed, the mean strength in the one-premise arguments was significantly *higher* for the nonalignable premises than for the corresponding alignable premises.[9] This is the *opposite* direction from what happened in Experiment 1, where there was an alignability advantage for the three-premise versions of these arguments. This means that the alignability advantage in Experiment 1 cannot result simply from independently accruing similarity or feature overlap across the premises.

Discussion

Our hypothesis that alignable three-premise arguments would exhibit strong monotonicity relative to their single-premise counterparts was supported. For both of the alignable types (A+C+ and A+C–), three-premise arguments received higher ratings than their respective single-premise counterparts.

[7] We will refer to single-premise versions arguments by adding a prime to the three-premise symbol: e.g., (A+C+)′.

[8] Because the sample sizes across the two experiments were not equal, we also performed a set of more conservative non-parametric analyses, which revealed the same pattern.

[9] That is, the mean strength of (A-C+)′ arguments was significantly higher than for (A+C+)′ arguments ($M = 2.87$, $SD = 1.21$; $M = 2.49$; $SD = 0.85$, respectively, $t(16) = 2.665$, $p < 0.025$. (A-C+)′ arguments were also rated reliably higher than the (A+C-)′ arguments ($M = 2.05$; $SD = 0.83$), $t(16) = 3.407$, $p < 0.025$.

General Discussion and Conclusion

These experiments offer support for the structure-based model of induction. The alignment approach predicts both nonmonotonicities and monotonicities accurately. When the additional premises are alignable, argument strength increases between one- and multiple premises, and is weak-monotonic from two- to three- premises. Strong monotonicity holds for alignable added premises.

Osherson et al's (1990) similarity-coverage model predicts monotonicity except when the additional premise forces a taxonomically higher spanning category. But the results of Experiment 1 showed nonmonotonicity even when category coverage was constant, as well as weak monotonicity despite a decrease in coverage. Across the board, (weak) monotonicity was observed between two- and three-premise cases for just those cases where the additional premise was alignable. The pattern in Experiment 2 was similar: With one exception, monotonicity between one- and three-premise arguments was observed only for alignable arguments.

Further evidence that argument strength judgments involve thinking about the specific relational schema, as opposed to overall similarity, comes from the property listings in Experiment 1. When given alignable third premises, subjects not only rated the arguments as strong, they also had clear opinions on what "Property P" might have been, and those guesses were highly uniform. These findings are consistent with there being a specific schema that emerged from the alignment.

What is the broader significance of these findings? First, premise comparison process must be a part of argument strength models. We have documented both (A) and (B) occurring *simultaneously*:

(A) P1, P2, P3 / C > P1, P2, P4 / C [Exp. 1]

(B) P3 / C < P4 / C [Exp. 2],

Since the same premises are added to both sides in going from B to A, this reversal cannot be explained in terms of accruing overall similarity or total feature overlap. It requires an explanation in terms of premise interactivity. People are not integrating individual premise-conclusion argument strengths (e.g. "P1/C + P2/C + P3/C") but aligning premises to determine what aspects of the premises *as a set* are relevant to the argument.

The evidence for premise interactivity presented here poses a challenge to the feature-based induction theory (Sloman, 1993). As an important theoretical alternative to the coverage model, the feature-based theory assumes that instead of computing category coverage, people are assessing total feature overlap between the premises and the conclusion. Monotonicity is predicted because the addition of a premise must either increase total feature overlap or maintain it. The addition of a premise can never decrease total feature overlap, so nonmonotonicities cannot be predicted. The systematic nonmonotonicities we have observed, as well as the evidence of premise interactivity, are inconsistent with the current formulation of the feature-based induction model.

Sloman (1993) has suggested an extension to the feature-based model -- a premise comparison mechanism that weighs common features of the premises more heavily than unique attributes. This might allow the feature-based model to predict some nonmonotonicities. However, it is unclear which common features of the premises will be weighted over others. An important forte of the structure-based model is that it *constrains* similarity by treating matching attributes that play similar roles in their respective concepts as more similar than matching attributes that do not (Medin, Goldstone & Gentner, 1993). Thus, inductive inferences are appropriately constrained.

Acknowledgments

This research was supported by the National Science Foundation and the Office of Naval Research. We thank Ken Kurtz, Sven Kuehne, Jeff Loewenstein and Kathleen Braun for their assistance at all stages of this project. We are also grateful to an anonymous reviewer for helpful comments on an earlier draft of this paper.

References

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

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

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(2), 1-12.

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., Gentner, D. & Quinn, B. (2000). SEQL- Category learning as incremental abstraction using structure mapping. *Proceedings of the 22nd meeting of the Cognitive Science Society.*

Kuehne, S.E., Gentner, D. & Forbus, K.D. (2000). Modeling infant learning via symbolic structural alignment. *Proceedings of the 22nd meeting of the Cognitive Science Society.*

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

Lopez, A., Gelman, S. A., Gutheil, G., Smith, E.E., 1992. The development of category-based induction. *Child Development 63, 1070-1090*

McDonald J., Samuels M. & Rispoli J. (1996). A hypothesis-assessment model of categorical argument strength, *Cognition, 59,* 199-217

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185-200.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231-280.

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