Boosting Analogical Arguments: The Effects of Goodness & Complexity on Everyday Arguments

Amy Bohan (amy.bohan@ucd.ie)
Mark T. Keane (mark.keane@ucd.ie)

Department of Computer Science, University College Dublin,
Belfield, Dublin 4, Ireland

Abstract

The use of analogical arguments is most often associated with political argumentation. However, our previous studies have found that analogical arguments are *not* as convincing as factual arguments. Should politicians rethink their rhetorical techniques ? In this paper, several possible criticisms of previous findings are considered, to determine whether analogical arguments might be considered more convincing than factual arguments. Two experiments that further investigate the use of analogies as arguments are reported. In Experiment 1, we replicate previous analogical/factual comparisons but use user-generated arguments for the materials, that are varied in terms of their pre-tested goodness. Experiment 2 investigates whether the complexity of the argument (i.e., the amount of information given in the arguments) might favor analogical over factual arguments. Finally, we outline a computational model of analogical arguments, which attempts to capture the effects found in these and previous experiments.

Introduction

Political argument is frequently peppered with analogical arguments; Saddam is a modern, genocidal Hitler not to be appeased, post-war Iran is a second Vietnam, and so on (c.f., Blanchette & Dunbar, 2001; Eemeren et. al., 2002). Yet, few studies have examined the cognitive basis of analogical argumentation. Although there is a substantial literature on the use of analogy in problem solving and reasoning (see e.g., Gentner, 1983; Keane, 1988; Keane et al, 1994; Holyoak & Thagard, 1995), this research has not been extensively applied to the communicative uses of analogy. Political analogies are fashioned to communicate key ideas and to convince. Yet there are few systematic studies of whether they are indeed more convincing than literal arguments. For example, perhaps people would be just as convinced by a factual argument simply pointing out that Saddam has committed genocide or that US losses in Iraq are substantial in personnel and material.

Keane & Bohan (2004) is one of the few studies to have explicitly asked people to rate the goodness of such analogical arguments and their factual equivalents (see Figure 1 for an example of the materials used). They found that people consistently find factual arguments more convincing than their analogical equivalents, suggesting that analogies are mere ornamentation. Interestingly, Keane & Bohan also showed that analogical arguments were rated better if people were encouraged to process the analogy more completely (e.g., with an explicit mapping task drawing out the correspondences between the domains). However, even these and other interventions never raised the convincingness of the analogical arguments above the level of the factual ones (see Keane & Bohan, 2004; Bohan & Keane, 2004, for details). For a discussion on factors affecting argumentation see Petty & Wegener (1999). However they do not explicitly compare analogical and factual arguments.

Trying to Save Analogy

There are a few possible criticisms of this previous work that could be advanced to save analogy. First, one might be concerned by the influence of people's beliefs on their assessments of the arguments. Keane & Bohan (2004) had people rate the arguments independently for their agreement/disagreement with the proposition and showed that their prior beliefs about the proposal did not influence their ratings of its goodness. That is, people can separate their own position on the proposition from their assessment of the goodness of the argument.

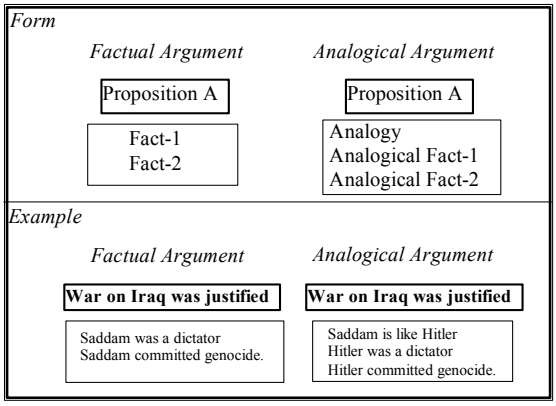

Figure 1: Abstract form and a gloss of a sample argument used in the experiments.

A second concern is that the analogies were not particularly good ones. Keane & Bohan's analogies all involved clear one-to-one mappings and so were, by definition, good analogies (see Gentner, 1983). However, Keane & Bohan also had participants rate these analogies for goodness and found that the majority of analogies were considered good rather than bad. So, this factor does not explain the results found.

A third possible criticism might be about the goodness of the arguments. If the arguments used in the study were not particularly good or representative ones in the space of possible arguments then perhaps people were not responding to them appropriately. Though this factor should affect factual and analogical arguments equally, it still remains an issue to be resolved. In Experiment 1, we explicitly manipulate the goodness of the arguments. We also selected the arguments from a set of arguments generated by participants, rather than relying on experimenter-designed ones (as used by Keane & Bohan, 2004).

Outline of Paper

In this paper, we report two experiments and a computational model that continues our program of research on argumentation by analogy. Experiment 1 uses participant-generated arguments to examine whether participant-generated goodness ratings of arguments is reflected in people's convincingness ratings. This experiment also re-examines the factual-analogical dimension to replicate earlier results. Experiment 2 further examines the issue of *complexity* systematically varying the amount of information given about the argument (factual or analogical). All previous work has used more complex arguments, typically a proposal and two supporting arguments (see Figure 1). Given our findings that encouraging more processing of the analogy resulted in it being perceived as better (see Keane & Bohan, 2004; Bohan & Keane, 2004), we hypothesized that perhaps analogical arguments might work best with simpler arguments (e.g., with a single predicate structure) rather than more complex arguments (e.g., several predicate structures). Finally, we outline a computational model that attempts to capture the effects found in these and previous experiments.

Experiment I

This experiment re-ran the factual-analogical manipulation previously examined by Keane & Bohan (2004), using participant-derived arguments rather than experimenter-designed ones. By analysing the frequency of arguments generated by people and ratings given to them, we operationally defined good and bad arguments for the propositions examined. Two different forms of analogy were used, close- and distant-domain analogies to assess whether this impacted the perceived goodness of the argument. As such, the experiment examined the effects of three variables: (i) argument type (analogical versus factual arguments), (ii) argument valency (good versus bad arguments) and (iii) domain type (close versus distant).

Method

Materials. Twelve topical propositions were selected (dealing with issues like health, drugs, war and so on) on the basis that they were topics about which most people have opinions. These propositions were used as targets from which to generate supporting arguments. Fifteen undergraduates at UCD were asked to *"come up with as many different arguments as you can for each of the propositions. The arguments should be in support of the proposition and should be short and to the point. In other words, if someone were trying to convince you of the proposition, what sort of argument would convince you?"*. A total of 632 responses were produced by the 15 participants, with an average of 3.5 arguments per proposition. Of those 632 arguments, 615 (97.3%) were factual arguments while only 17 (2.7%) were analogical.

These arguments were content analysed to identify the common arguments proposed by different people. From this content analysis it was found that each proposition has 6-11 distinct arguments. Fifteen new participants were given these arguments and asked to rank them in order of "*which is the best argument*." These orderings were then analysed by the mean ranks proposed by participants and the two best and two worst arguments were selected.

Analogies were developed for each proposition which had isomorphic structures and common relations. Analogical arguments for the four factual arguments (2 good and 2 bad) were then created based on the analogies. For example, for the proposition 'Many astronauts will die in the attempt to travel to Mars', the analogy "space travel is like polar travel" was created. The factual argument 'there are many unforeseen dangers in space travel' was therefore transformed into 'there are many unforeseen dangers in *polar travel*' as the analogical version. Six of the analogies used close domains and the other six used distant domain analogies. An example of a close domain analogy was 'Saddam is like Hitler', while an example of a distant domain analogy was the analogy above, 'space travel is like polar travel'.

Inspection of the initial ratings suggested that 2 materials did not meet the criteria and, hence, were dropped. The final material set had 10 propositions each of which had four arguments: 1 factual good, 1 factual bad, 1 analogical good and 1 analogical bad (see Table 1). Finally, half of the analogical domains were distant and half were close.

Table 1: Example of the four types of arguments for the given proposition

Proposition		**The majority of books will be replaced by e-books because they are more convenient**
Factual Arguments	Good	E-books are more accessible to an internet audience than paper books.
	Bad	Online classes already exist
Analogy		E-books are like mp3's
Analogical Arguments	Good	Mp3's are more accessible to an internet audience than CD's.
	Bad	Online music libraries already exist

The set of materials were presented in booklet form with a cover sheet explaining the task to be carried out. Instructions asked participants to rate how convincing they thought the argument was in support of the proposition regardless of their beliefs, using a 7-point scale on the proposition. On each of the following pages the proposition-argument pairs was presented above the rating scale. Materials were randomly re-ordered for each participant and

no participant received different versions of the same argument.

Procedure. Participants read instructions that explained the 1-7 argument goodness scale (1 being "very bad", 7 being "very good" and 4 being "neither good nor bad"), and a sample proposition was shown with a factual argument and another shown with an analogical argument. The participants were asked to take their time over each decision and to make "an *objective* assessment of the arguments. That is, to make a judgment regardless of your agreement or disagreement with the proposition". Each proposition-argument pair was presented on a separate page with a marked space for participants to note their 1-7 goodness rating.

Results

Ratings of Arguments. A 2x2x2 general linear model ANOVA for unbalanced designs was carried out on the ratings data for the within-participant variables of argument-type, argument-valency and domain (see Table 2). All analyses of variance by participants and by items were performed by respectively treating participants (F_1) and sentences (F_2) as a random factor. These analyses revealed a main effect of argument-type, with the analogical arguments (M = 3.58) being rated worse than the factual ones (M = 4.1); F_1 (1, 536) = 14.55, $p < 0.0005$, MSe = 41.529; F_2 (1, 587) = 15.27, $p < 0.0005$, MSe = 42.034. A main effect of argument-valency was also found in the expected direction; F_1 (1, 536) = 34.30, $p < 0.0005$, MSe = 96.295; F_2 (1, 587) = 33.44, $p < 0.0005$, MSe = 90.849. There was also a reliable interaction between argument-type and argument-valency; F_1 (1, 536) = 6.11, $p < 0.014$, MSe = 17.206; F_2 (1, 587) = 6.36, $p < 0.012$, MSe = 17.329. This interaction essentially shows that good-factual (M = 4.66) arguments are significantly better than good-analogical (M = 3.81) arguments however there is no significant difference between bad-factual (M = 3.55) and bad-analogical (M = 3.35) arguments. No other reliable interactions were found.

Table 2: Average scores for argument-type, argument-valency and domain in Experiment I

	Factual		Analogical	
	Good	**Bad**	**Good**	**bad**
Distant	4.7	3.13	3.48	3.24
Close	4.63	3.96	4.15	3.45
average	**4.66**	**3.55**	**3.81**	**3.35**
average	**4.1**		**3.58**	

Discussion

This experiment confirmed the results found in previous studies, along with showing some new results. Firstly, it replicates the precedence people give to factual arguments over analogical ones, when participant-generated arguments are used. Second, it shows that independent ratings of the valency of the arguments (drawn from frequency and ratings data) are reflected in convincingness ratings. Finally, it shows that domain-distance does not matter in the assessment of analogical arguments.

Experiment II

There is one final possibility to consider that may save the role of analogy in argumentation. Other studies have shown that if people are encouraged to process the analogy (by directions to explicitly generate mappings) then analogical arguments are rated as better, though they never go above the goodness of factual ones (Bohan & Keane, 2004). It could be the case that analogies only work with quite simple arguments (e.g., a single proposition with a single argument), whereas we have only been testing them with more complex arguments (i.e., a proposition with 2 arguments).

So, in this experiment, we systematically varied the complexity of the factual and analogical arguments at three levels: complex (3 arguments), medium (2 arguments), and simple (1 argument). People were asked to carry out two tasks on each proposition: a belief task and an evaluation task (ala Keane & Bohan, 2004). These two tasks were counterbalanced given a 2 argument type (factual or analogical) x 3 complexity (complex or medium or simple) x 2 task-order (belief-then-evaluation or evaluation-then-belief) design.

Table 3: Argument Complexity for Proposition: *Paralysed people, as a result of a severed spinal cord, could be able to walk again*:

	Literal	**Analogy**
Simple	advances in medical science could allow the reconnection of severed spinal cords.	a severed spinal cord is like a severed fibre-optic cable. Advances in engineering allow the reconnection of severed fibre-optic cables.
Medium	advances in medical science could allow the reconnection of severed spinal cords, by bypassing the damaged section.	a severed spinal cord is like a severed fibre-optic cable. Advances in engineering allow the reconnection of severed fibre-optic cables, by bypassing the damaged section.
Complex	advances in medical science could allow the reconnection of severed spinal cords, by bypassing the damaged section and allowing communication to resume	a severed spinal cord is like a severed fibre-optic cable. Advances in engineering allow the reconnection of severed fibre-optic cables, by bypassing the damaged section and allowing communication to resume

Method

Materials. Ten sets of proposition-argument items were used in this experiment. The five best materials from our first series of experiments (see Keane & Bohan, 2004) were taken and the five best (good) arguments from Experiment 1 (described above). This ensured a mix of experimenter-generated and participant-generated materials. There were three versions of each item, varied by complexity at three levels: complex (3 arguments), medium (2 arguments),

simple (1 argument) – see Table 3. Booklets were structured as in Experiment 1.

Participants & Design. One hundred-and-twenty native English-speaking undergraduates at University College Dublin took part in the experiment. The order of the tasks was counterbalanced so that half the participants received the belief task before the evaluation task (belief-then-evaluation conditions) while the other half received the tasks in the opposite order (evaluation-then-belief conditions). So, the design was a 2 argument-type (factual or analogical) x 3 complexity (complex, medium and simple) x 2 task-order (belief-then-evaluation or evaluation-then-belief) one with argument-type and complexity being within-participants and task-order being between-participants.

Procedure. The evaluation task was identical to that used in Experiment 1 for rating the goodness of the arguments. The belief task asked participants to rate whether they agreed/disagreed with the proposition on the 1-7 agreement scale (1 being "strongly disagree", 7 being "strongly agree" and 4 being "no opinion"). Keane & Bohan (2004) had previously used this task to determine whether there was any relationship between people's *a priori* beliefs and their goodness ratings.

Table 4: Average scores for analogical and factual arguments in Experiment II

	Argument Complexity			
	Complex	**Medium**	**Simple**	**Average**
Analogy	4.165	4.135	4.045	**4.115**
Factual	4.68	4.715	4.261	**4.552**

Results

Table 4 shows that the factual arguments were considered to be better than the analogical ones overall. An effect of complexity was also found but there was no interaction between argument-type and complexity showing that analogical arguments do not especially benefit from simpler arguments.

Ratings of Arguments. A 2x3x2 ANOVA was carried out on the ratings data for the between-participant variable of task-order and within-participant variables of argument-type and complexity. All analyses of variance by participants and by items were performed by respectively treating participants (F_1) and sentences (F_2) as a random factor. These analyses revealed a main effect of argument-type with factual arguments (M=4.55) being rated as being better than the analogical arguments (M=4.12), F_1 (1, 1076) = 16.70, $p < 0.0005$, $MSe = 57.305$; F_2 (1, 1184) = 19.02, $p < 0.0005$, $MSe = 58.248$. There was also a main effect of complexity F_1 (2, 1076) = 3.50, $p < 0.030$, $MSe = 23.980$; F_2 (2, 1184) = 3.31, $p < 0.037$, $MSe = 20.270$. There was no other reliable main effects or interactions.

Pair-wise comparisons using a Bonferroni test were carried out to determine the locus of the complexity effects. In the analogy conditions, there was no reliable differences between the three levels of complexity. In the factual conditions however, simple and medium conditions were reliably different ($p < 0.0169$) as were the simple and complex conditions ($p < 0.0317$). This shows that the complexity effects mostly reflect differences between the factual conditions rather than the analogy ones.

The Impact of Belief on Evaluation. One of the key questions which we asked in previous experiments as in this one was whether people's prior beliefs in the proposition would have any impact on their rating of the goodness of the argument, even though we asked people to be as objective as possible. If people were rating the arguments in line with their beliefs then we should, for example, find that people gave high goodness ratings to arguments in which they strongly agreed with the proposition and low goodness ratings to arguments with which they strongly disagreed. However, as was found in previous studies, there is little evidence of such a relationship. Although the correlation between participants' belief ratings and their goodness ratings for the items is moderate, using Pearsons' product-moment correlation $r(1198) = 0.423$, $p < 0.0005$, we do not believe it to be high enough to suggest people's subjective assessments of the arguments. To date, in all previous experiments, there has not been any correlation between bias and argument assessment.

Discussion

This experiment reveals three main findings: (i) analogical arguments are not considered to be better than their factual equivalents; (ii) people's *a priori* agreement/disagreement with the proposition does not affect their subsequent evaluation of the goodness of an argument for that proposition; (iii) the complexity of arguments effects the rating of the argument, but there is no evidence to suggest that this effect specifically favors analogical arguments. In short, analogy is not saved by these results.

The Analogical Argument Analysis Model

We have developed an initial model, called the Analogical Argument Analysis Model (AAAM, pronounced triple-A model), to capture the results of all of these experiments. The fundamental proposition underlying AAAM is that the more processing that is done on an analogy, the better it will be perceived as an argument in support of a proposition (see Figure 2 for a schema of the model's components). AAAM takes as input a proposition-argument pair and outputs a goodness score of the convincingness of the argument in support of the proposition. AAAM has two main modules: (i) the Argument Analysis module that processes the arguments, and (ii) the Analogy module that performs the mapping between the two analogical domains. The Analogy module has the *eagerness parameter* that influences the amount of time given to processing the analogy[1]. The

[1] The analogy is a standard structure-mapper IAM (Keane & Brayshaw, 1988; Keane et al., 1994)

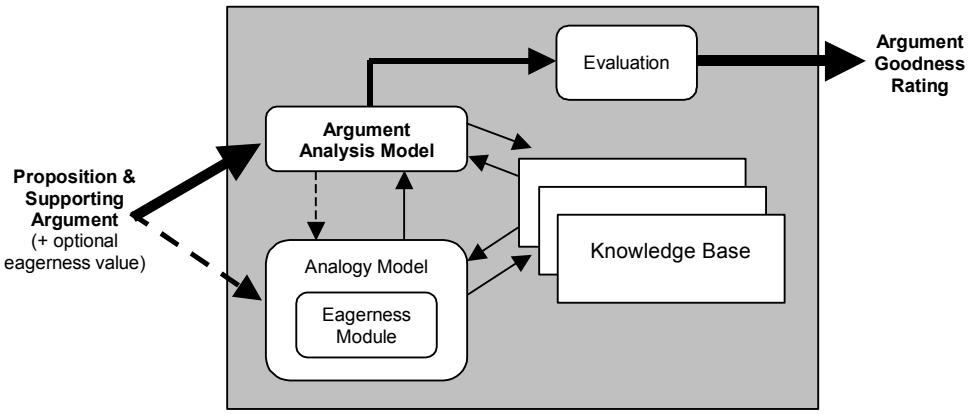

Figure 2: Outline of Analogical Argument Analysis Model

processing of the arguments by the argument analysis module evaluates the number of causal steps from the arguments to the proposition and also the complexity of the argument itself, as our findings suggest that medium and complex arguments are significantly better than single fact factual arguments. This data is passed to the Evaluation component which calculates the goodness rating of the argument on the basis of the number of causal steps plus the complexity score. The AAAM has a knowledge base (KB) that represents the causal relations between the arguments and the proposition and is used by both modules (see Figure 3).

The Argument Analysis Module:

- Identifies whether a factual or analogical argument has been input

If (an analogical argument => mapping module is invoked)

The Analogy module:

- Identifies source and target and searches the knowledge-base for the two domains
- Verifies mapping by examining predicates and higher-order predicates
- if mapping task is given as input, mapping is fully explored
- The depth to which the mapping is explored is dependent on the eagerness parameter which may terminate processing prematurely
- When mapping is terminated or completed the identified arguments are passed back to Analysis module
- A mapping score which returns the extent of the mapping is returned to the evaluation component

Else

- Searches the knowledge-base for proposition and argument specifications
- Analyses causal steps from argument to proposition; the fewer causal steps from argument to proposition the better the goodness rating
- Determines complexity of argument and explores relevance/logic of additional facts (if any); the more relevant facts the better the goodness rating.
- Returns the number of causal steps plus complexity score to evaluation component

Evaluation:

- Evaluates argument rating based on:
  - # casual links; less => better goodness score

if (factual argument)

complexity score => better goodness score

else

analogical argument => no change

- mapping score; higher => better goodness score

An example of KB entry for the proposition 'Paralysed people, as a result of a severed spinal cord, will be able to walk again' is illustrated in Figure 3.

```
Prop:
will_walk(paralysed(people)),
cause(severed(spine), paralysed(people)),

Assumptions:
severed(spine) => ¬walk(paralysed(people),
reconnect(severed(spine)) =>
¬(paralysed(people)) => walk(people)

Arguments:
 i.   cause(advances(medical_science),
      reconnect(severed(spine)))
ii.   cause(research, reconnected(spine))
.ii.  advances(medical_science) =>
      cause(bypass(damaged_spine_section)
      , reconnect(severed(spine)))
iv.   bypass(damaged_spine_section) =>
      resume(communication, nerve_ends)
 v.   resume(communication, nerve_ends) =
      reconnect(severed(spine))
vi.   motivate(recognition, research) AND
      motivate(economic_rewards,
```

Figure 3: Knowledge base record representing the causal relations between the arguments and proposition

General Discussion

Overall, we have shown that analogical arguments are not as convincing as factual arguments. This finding has been consistently replicated across all of our previous studies. One possible criticism of a study of argumentation is that the arguments used were not very good. However, we have shown that the same results are found with participant-generated arguments, and that the appraisal of the quality of the arguments is also reflected by participant's ratings. These findings of consistent appraisal of the quality of the arguments supports the criticism of the quality of the analogies in the arguments, as there was consensus between participants regarding the 'good' and 'bad' analogical arguments. In other words, participants were able to distinguish good and bad arguments with the same analogy, therefore the analogy itself is not to blame for the ratings.

We have also shown that argument complexity has no bearing on the rating of analogical arguments. So whether the analogical argument contains a simple one-fact structure or a more complex multi-fact structure, people rate them similarly, and less well then factual arguments. However, complexity does boost the factual scores; simple arguments are rated poorer than multiple fact arguments. This finding suggests that while the additional facts improve understanding of the factual arguments, the underlying interpretation of the analogical arguments is missed altogether, so no amount of additional facts make up for the lack of processing of the analogical mapping. Finally, we have shown that these effects can be captured in an effective procedure, implemented in our model.

Acknowledgments

This work was funded in part by grants from University College Dublin to the first author.

References

Blanchette, I. & Dunbar, K. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition 29(5)*, 730-735.

Bohan, A, & Keane, M. T (2004). Towards a Computational Model of Analogical Arguments, *In Proceedings of CMNA IV, 4th Workshop on Computational Models of Natural Argument at ECAI 2004 Edited by Floriana Grasso, Chris Reed and Giuseppe Carenini,* Aug 23-25, pp 65-67.

Eemeren, F.H. van, Grootendorst, R. & Snoeck Henkemans, A.F. (2002). *Argumentation; Analysis, Evaluation, Presentation*. Lawrence Erlbaum Associates. New Jersey.

Falkenhainer, B. Forbus, K and Gentner, D.(1989) "The Structure Mapping Engine: Algorithm and Examples", *Artificial Intelligence, 41*, 1-63,

Gentner, D (1983). Structure-mapping: A theoretical framework for analogy, *Cognitive Science, 23,* 155—170.

Holyoak, K. J. & Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought.* The MIT Press, Cambridge, MA, 1995

Keane, M.T. (1988). *Analogical Problem Solving*. Ellis Horwood Ltd, Chichester.

Keane, M.T. & Bohan, A (2004). Should Politicians Stop Using Analogies? Whether Analogical Arguments Are Better Than Their Factual Equivalents. *In Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.* Aug 5-7. Hillsdale, NJ: Erlbaum.

Keane, M.T., Ledgeway, T.& Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18,* 287-334.

Petty, R. E & Wegener, D.T (1999). The Elaboration Likelihood Model: Current Status and Controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology*. New York: Guilford Press.

Rips, L.J. (2002). Circular Reasoning. *Cognitive Science,* 26, 767-795