

Should Politicians Stop Using Analogies? Whether Analogical Arguments Are Better Than Their Factual Equivalents

Mark T. Keane (mark.keane@ucd.ie)

Amy Bohan (amy.bohan@ucd.ie)

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

Abstract

In political argumentation, analogies are often used to convince an audience of one's views. For example, in political debates leading up to the Iraq War, one such analogical argument was that Saddam Hussein was like Hitler and therefore Saddam should be forcibly ousted. But are all analogical arguments really convincing? In this paper we investigate whether analogical arguments are actually more convincing than factual arguments. In Experiment 1 we asked people to rate analogical and factual arguments for various propositions and found that people considered factual arguments more convincing. In Experiment 2, we asked people to think more explicitly about the analogical mappings but still found that people considered the analogical arguments less convincing than the factual ones. These findings suggest that people are *not* more easily convinced by an analogical argument than a straight factual one, suggesting that perhaps politicians should re-consider their rhetorical tactics after all.

Introduction

Is the aftermath of the Iraq War like Germany post-WWII or Northern Ireland or, indeed, is it another Vietnam? In the furious political debate following the Iraq War, politicians on both sides have used different analogies to bolster their arguments. In science, analogies are often used to discover something new about natural phenomena, but in politics they are used to convince an audience of one's views. In this paper, we consider whether such analogical arguments are more convincing than their equivalent, factual arguments.

Though classical rhetoric has long advocated the use of analogy in argumentation (Plato, *Phaedo*, trans. 1871, 71c-d being a prime exponent of the craft) and political science regularly analyses the analogies used in political debate (Blanchette & Dunbar, 2001), we know of no studies that have systematically determined whether people actually find analogical arguments more cognitively convincing than their factual equivalents. This gap in the literature is all the more surprising when one considers the amount of research on the separate topics of argumentation and analogy. The nature of argumentation has been elaborated in a rich literature in philosophy, logic and psychology (e.g., Rips 2002; Voss & Van Dyke, 2001). Similarly, the nature of analogy has been empirically explored in many studies, supported by clearly articulated theory that has been modeled computationally (see Gentner, 1983; Holyoak & Thagard, 1995; Keane, 1997; Keane, Ledgeway & Duff,

1994; Hummel & Holyoak, 1997). Yet, the two areas have not been combined in a systematic study of their cognitive underpinnings. In the present paper, we attempt such a combination.

The Present Experiments

We propose a novel paradigm for assessing people's evaluation of arguments that pits analogical arguments and their factual equivalents against one another. In our experimental setup, people are presented with a proposition and a two-fact argument supporting this proposition (see Figure 1). They are then asked to rate how good they found this argument as a warrant or support for the proposition. For a given proposition, the argument was either two facts or two equivalent analogical facts.

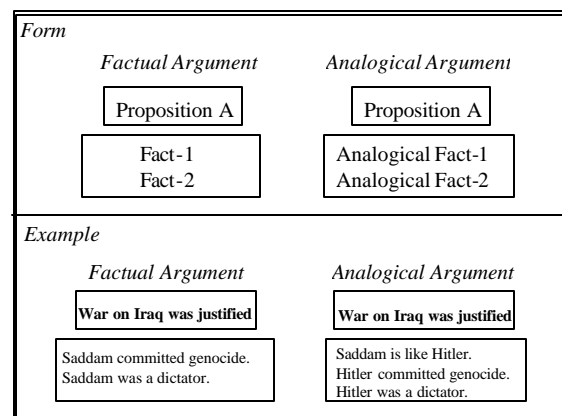


Figure 1: Abstract form and a gloss of a sample argument used in the experiments.

For example, the Iraq War argument suggests that going to war with Iraq was justified because Saddam was a dictator and had committed genocide in his country (see gloss in Figure 1). The analogical equivalent suggests that war on Iraq was justified because Saddam is like Hitler, and Hitler was a dictator and had committed genocide in his country. In this way, the analogical argument presents the same facts about Saddam but through the lens of a WWII analogy. This is the typical way in which politicians use analogies, suggesting a parallel in an analogous domain that supports their argument in a current domain.

From a cognitive perspective, there are several reasons why analogical arguments might be more convincing than factual ones. Essentially, an argument is defined as “a course of reasoning aimed at demonstrating the truth or falsehood of something” (Kuhn, 1991, pg.12). That is, in any argument the aim is to convince an audience of the truth or falsehood of certain facts, that these facts in some way support or warrant your proposition and that, therefore, your proposition is justified as right or correct (Toulmin, 1958; Kuhn, 1991). Therefore if there is agreement that Saddam’s dictatorial powers and genocidal activities are bad and that these facts warrant the action of war as a response; then making war is, in some way, a necessary response to the facts given. One of the key steps in this process is getting the audience to accept the warrant as a necessary link between the facts and the proposition. Cognitive models of analogical thinking, show us that people use analogies to make high-level causal inferences about analogous domains (c.f., Keane, 1988). In this case, the comparison to Hitler provides a match to WWII where these dictator and genocide facts were viewed as essential reasons for military intervention. Thus, the analogy provides a previous case where the facts caused or strongly warranted military intervention, inviting the inference that war is therefore appropriate in Iraq too.

We report two experiments on the role on analogy in argumentation. These experiments used a wide variety of topical arguments from different domains covering alcohol abuse, military service, university entrance exams and traffic congestion policy. The analogies used also varied in the distance of the domains from one another; some involved close domains (e.g., Iraq War and WWII), others involved distant domains (e.g., Art and Foreign Languages). In the experiments, no single individual saw both the factual and analogical versions of a given argument. We also gathered people’s ratings of their *a priori* belief in the proposition (i.e., their agreement/disagreement with it) to check for any belief bias in their assessment of the argument. In Experiment 1, we made a direct comparison of people’s goodness ratings for the factual and analogical arguments to various propositions. In Experiment 2, we replicated this test with an intervention that encouraged people to reflect more on the analogy. To presage our findings, the evidence suggests that people are *not* more easily convinced by an analogical argument over a straight factual one, suggesting that politicians might indeed want to re-consider their rhetorical tactics.

Experiment 1

This experiment examined whether analogical arguments were deemed to be better (i.e., more convincing) than their factual equivalents for a variety of topical propositions. People were shown 10 different propositions (5 with factual arguments, 5 with analogical arguments) and asked to carry out two tasks on each: a belief task and an evaluation task. In the belief task, they were shown the proposition on its own and asked to rate their agreement/disagreement with it

on a 7-point scale. In the evaluation task, they were shown the proposition and the argument (factual or analogical) and asked to rate its goodness as an argument for the proposition on a 7-point scale. The order of these tasks was counterbalanced in two different conditions. If our politicians are right then the analogical arguments should be considered to be better than their factual equivalents.

Method

Materials. Ten propositions were created based on either currently debated topics (e.g., the Iraq War, the school examination system, societal effects of drugs, utility of GM foods) or long-standing debated topics (e.g., the introduction of the death penalty, military service, public funding of the arts). For each of these propositions, a two-fact argument was created based on the typical reasons used to support these propositions. Analogies were then developed that had clear one-to-one correspondences to the conceptual objects and relations used in the original facts. Eight different materials sets were made up from random selections of particular materials and arguments, such that each set contained 10 unique propositions, 5 of which had corresponding analogical arguments with the other 5 having corresponding factual arguments. This material-group variable is not reported in the results as an analysis of people’s ratings shows that it had no reliable effect on results found.

For every material set, two booklets were collated for the two tasks. The belief-rating booklet had a cover sheet explaining that people should rate how strongly they agreed/disagreed with the proposition on a 7-point scale, followed by 10 pages with a single proposition and rating scale shown on each page. The evaluation booklet had a cover sheet explaining that people should rate how good/bad they thought the argument was for the proposition on a 7-point scale regardless of their beliefs, followed by 10 pages with a proposition plus its corresponding (factual /analogical) argument and a scale shown on each page. The items in every booklet were randomly ordered for each participant.

Participants & Design. Thirty-two 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 2 task-order (belief-then-evaluation or evaluation-then-belief) one with argument-type being within-participants and task-order being between-participants.

Procedure. In the evaluation task, 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 judgement 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. In the belief task, the instructions and materials were presented in the same way, except that the proposition alone was presented and the instructions explained that people were to rate how strongly they disagreed/agreed with the proposition on the 1-7 agreement scale (1 being “strongly disagree”, 7 being “strongly agree” and 4 being “no opinion”).

Table 1: Percentage of good arguments and mean goodness ratings for both experiments

<i>Measure Experiment</i>	<i>Analogical</i>		<i>Factual</i>	
	%Good	Mean Rating	%Good	Mean Rating
Expt. 1 belief-then- evaluation evaluation- then-belief <i>Mean</i>	21.3%	2.59	58.8%	4.3
	30%	3.11	42.5%	3.76
	25.6%	2.85	50.6%	4.03
Expt. 2 belief-then- evaluation evaluation- then-belief <i>Mean</i>	33.8%	3.43	68.4%	4.59
	43.8%	3.80	46.2%	3.90
	38.8%	3.6	57.2%	4.25

Results

Table 1 summarises the main results of the Experiment showing that the factual arguments were considered to be better than the analogical ones on several different measures.

Percentages of Good and Bad Arguments. A rough feel for people’s responses to the arguments can be gleaned by re-classifying their ratings into ordinal groups of good (> 4), bad (< 4) or indifferent (=4) according to how they rated the argument on the goodness scale. Overall, 320 arguments were evaluated in the experiment, 160 factual and 160 analogical. Of the factual arguments, 38.1% (61) were rated as bad and 50.6% (81) as good (the remainder being indifferent). Of the analogical arguments, 67.5% (108) were rated as bad and 25.6% (41) as good (the remainder being indifferent). Collapsing across the order conditions, this result was found to be reliably different using a Chi-squared analysis, $\chi^2(1) = 26.032$, $p < 0.0001$, $N=291$. However, an inspection of the percentages clearly shows that task-order has an impact too, in that more arguments were considered

to be good in the belief-then-evaluation conditions (40%) than in the evaluation-then-belief conditions (36%). Indeed, on the face of it, there appears to be an interaction between task-order and argument-type that is more easily revealed using the ratings measure.

Ratings of Arguments. A 2x2 ANOVA was carried out on the ratings data for the between-participant variable of task-order and within-participant variable of argument-type. 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 factual arguments ($M=4.03$) being rated as being better than the analogical arguments ($M=2.85$), $F_1(1, 286) = 40.02$, $p < 0.0005$, $MSe = 111.628$; $F_2(1, 307) = 40.30$, $p < 0.0005$, $MSe = 111.628$. There was also a reliable interaction between task-order and argument-type $F_1(1, 286) = 8.10$, $p < 0.005$, $MSe = 22.578$; $F_2(1, 307) = 7.20$, $p < 0.008$, $MSe = 19.938$. Planned pairwise comparisons revealed that the factual/belief-then-evaluation condition was reliably different to all the other conditions using Bonferroni adjustments ($ps < 0.0005$). None of the other comparisons were reliably different to one another.

The Impact of Belief on Evaluation. One of the key questions 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, there is little evidence of such a relationship. The correlation between participants’ belief ratings and their goodness ratings for the items is low and not reliable, using Pearsons product-moment correlation $r(319) = 0.36$, $p < 0.0005$.

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) people find factual arguments much better if they are first asked to rate their belief in the proposition.

The first of these findings should be a surprise for most politicians, as it shows that they might as well be using straight-forward factual arguments to present their views. In the next experiment, we explore whether this result may have occurred because people did not process the analogy sufficiently to draw out all its implications.

The second finding suggests that people can separate their belief in the proposition from their assessment of its goodness, when they are instructed to do so. In other words

that people can maintain a level of objectivity in evaluating arguments.

The third finding of a task-order effect was as unexpected as it is interesting. It shows that if someone rates their belief in a proposition and subsequently sees a factual argument for that proposition they consider it to be better than the same argument presented before they give their belief rating (this effect does not occur for analogical arguments). Why should this occur? One possibility is that when people are first asked to rate their agreement with the proposition, they must think of their own arguments for the proposition. Then, when they are subsequently shown some arguments for the proposition many participants may find them more convincing because they are similar to their own arguments. In contrast, when participants are first asked to evaluate the argument and proposition (before being asked for their belief) there is less opportunity to think of their own arguments, less opportunity to recognize similarities and, hence, less of a boost to the goodness rating of the argument. No parallel benefits are found for the analogical arguments because people do not readily think of their own analogical arguments when rating their belief in the proposition (see Gick & Holyoak, 1980; Keane, 1985, 1988, on people's tendency *not* to explore analogical possibilities without instructions to do so). In the next experiment, we attempt to replicate this task-order effect to determine whether it is robust.

Experiment 2

In Experiment 1, we found that people failed to be convinced by analogical arguments relative to their factual equivalents. This result could be due to the amount of cognitive processing people have to carry out on analogical arguments as opposed to factual arguments. In the analogical case, they must understand the analogical arguments, map the corresponding objects and relations between the two domains, then apply the mappings to the proposition's domain and, finally, evaluate it. In the factual case, they merely have to understand the argument, relate it to the proposition and evaluate it. Maybe participants in Experiment 1 did not bother to draw the analogy and, hence, marked these arguments down. We should note that this explanation is somewhat implausible as we know from the literature that people readily appreciate and understand analogies (see Keane, 1988; Holyoak & Thagard, 1995). So, in this experiment, we explicitly asked people to report their mapping of key objects between the two domains to ensure that the analogy was being properly processed. We also ran the task-order manipulation again to see if it could be replicated.

Method

Participants, Materials & Design. Thirty-two native English-speaking student volunteers at University College Dublin took part in the experiment. The materials were the same as those used in Experiment 1, as were the grouping of material sets and organization of task booklets. As before,

the design was a 2 argument-type (factual or analogical) x 2 task-order (belief-then-evaluation or evaluation-then-belief) one, with argument-type being within-participants and task-order being between-participants.

Procedure. The procedure was as in Experiment 1, except for one change to the analogical argument materials. In each case where an analogical argument was presented, people were shown two boxes listing three key objects from each domain of the analogical argument (as shown in Figure 2). The participants were asked to draw lines between the corresponding objects in the analogy. For example, in the Saddam Hitler analogy, Hitler corresponds to Saddam, Germany corresponds to Iraq etc. They were asked to perform this mapping before they rated the analogical argument in the evaluation task.

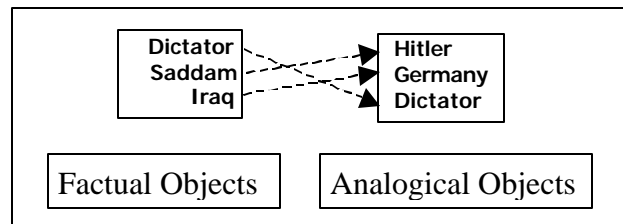


Figure 2: Example of the object-mapping task used in Experiment 2

Results and Discussion

Table 1 summarises the main results of the experiment showing that the factual arguments were considered to be better than the analogical ones on several different measures. The pattern of findings replicates those found in Experiment 1, with a strengthening of the effects being found.

Percentages of Good and Bad Arguments. Re-classifying people's responses into the ordinal groups of good (> 4), bad (< 4) or indifferent ($=4$) we found that (i) of the 159 factual arguments evaluated 35.2% (56) were rated as bad and 57.2% (91) as good (the remainder being indifferent), (ii) of the 160 analogical argument evaluated 52.5% (84) were rated as bad and 38.8% (62) as good (the remainder being indifferent). Collapsing across the task-order conditions, this result was found to be reliably different using a Chi-squared analysis, $\chi^2(1) = 11.093, p < 0.0009, N=293$.

Ratings of Arguments. A 2x2 ANOVA was carried out on the ratings data for the between-participant variable of task-order and within-participant variable of argument-type. 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 factual arguments ($M=4.25$) being rated as being better than the analogical arguments ($M=3.6$), $F_1(1, 255) = 14.17, p < 0.001, MSe =$

37.154; $F_2(1, 279) = 7.06, p < 0.016, MSe = 38.465$. There was also a reliable interaction between task-order and argument-type $F_1(1, 255) = 8.63, p < 0.006, MSe = 22.623$; $F_2(1, 279) = 10.02, p < 0.005, MSe = 23.646$. Planned pairwise comparisons revealed that the factual/belief-then-evaluation condition was reliably different to all the other conditions using Bonferroni adjustments ($ps < 0.005$). None of the other comparisons were reliably different to one another. So, again, we replicate the task-order x argument type interaction found in the previous experiment. The main effect of task order was not reliable.

The Impact of Belief on Evaluation. Again, we found as in Experiment 1, that there is little evidence to suggest that people's prior beliefs in the proposition affected their assessment of the argument. The correlation between participants' belief ratings and their goodness ratings for the items is low and not reliable, using Pearson's product-moment correlation $r(318) = 0.243, p < 0.0005$.

Conclusions from Experiments 1 and 2. So, again we find that the analogical arguments were considered to be less convincing than the factual ones, even when we ensure that people have mapped the analogy appropriately. However, it is noteworthy that, relative to Experiment 1, their analogical arguments seem to be rated as slightly better (e.g., 38.75% are considered good arguments in Experiment 2, relative to 25.63% in Experiment 1).

General Discussion

The results of these experiments suggest that politicians should stop using analogies, as they do not seem to provide much more than a sugar coating on the convincingness of a straight, factual argument. Overall, we have shown several novel findings about the use of analogy in argumentation. First, we have seen that analogical arguments are generally not considered to be as good as factual arguments. Second, we have seen that it is very hard for analogical arguments to challenge the goodness of factual arguments (in other experiments we have found that even when the full factual argument is given along with the analogical argument, the evaluations do not go higher than the plain factual arguments). Third, we have found that factual arguments' ratings can be boosted if people are asked to reflect on the proposition in advance of rating them. Finally, we have seen that people can separate their beliefs in a proposition from their evaluation of an argument to that proposition, showing a noteworthy objectivity in their evaluations. To traditional rhetoricians this evidence may seem unwelcome and unconvincing. In the remainder of this section, we consider three main objections that might be raised to our findings.

The Arguments Were Not Very Good. One argument against the evidence would be to maintain that the arguments used were not very good; that if you had better arguments then different results would be found. Unfortunately, we do not have data on how many people in

a population find a given argument to be good or bad relative to some proposition, so it is hard to judge whether our arguments are in some way unrepresentatively poor. What we do know is that people only found 35-38% of our factual arguments to be bad (50%-60% of these arguments being considered good). On the face of it, using the "you can fool some of the people all of the time..." adage these figures appear to be reasonable levels of goodness. As such, we would argue that there is no obvious deficiency in the arguments used. Furthermore, we should also note that many of the arguments used were ones that people have used to support these propositions in everyday life.

Maybe Our Analogies Are Not Very Good. If one admits that the arguments are adequate, then a further objection could be that the analogies were, in some way, inadequate. Again this is a hard objection to assess given that we have little idea of the space of possible analogies used in argumentation. What we can say is that all of the analogies used conform to what is deemed to constitute an analogy in the literature; they involve one-to-one mappings, they involve matching relational structure and they suggest inferences by analogy connecting the arguments and the proposition (c.f., Gentner, 1983; Hummel & Holyoak, 1997; Keane et al., 1994). But, what if some are, in some way, better than others.

To explore this possibility, we presented a separate group of 16 participants with a mixture of 10 analogies and non-analogies asking them to rate the goodness of the analogies on a 7-point scale. Of the 10 materials used in the experiment only one received a bad goodness rating (i.e., < 4), all of the remainder being rated as being good (with mean ratings from 4.25 to 5.25). Overall, people reliably distinguished the analogies ($M = 4.6$) from the non-analogies ($M = 2.5$), using a dependent t-test, $t(157) = 8.10, p < 0.0005$. So, the failure of the analogical arguments cannot be attributed to the poorness of the analogies.

Are There Other Ways in to Improve Analogies? A final objection is that we have not appropriately intervened to boost the analogy. We saw that asking people to plot the object mapping improves their goodness ratings for the analogy arguments. Perhaps there is some other intervention that might boost them further. It is unclear to us what this intervention might be. However, this objection in a sense misses the point. If we did find some intervention that promotes analogical arguments is it quite likely to be quite artificial. In the cut and thrust of political debate the facts of the matter are generally known (though may not be stated explicitly) and the analogy is provided to be understood on the spot (without, for example, asking people to specify the object mappings).

Acknowledgments

This work was funded in part by grants from University College Dublin and Science Foundation Ireland under Grant No.03/IN.3/I361 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.
- Gentner, D (1983). Structure-mapping: A theoretical framework for analogy, *Cognitive Science*, 23, 155—170.
- Gick, M.L., & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
- Holyoak, K. J. & Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*. The MIT Press, Cambridge, MA, 1995
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representation of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427-466.
- Keane, M.T. (1997). What makes an analogy difficult?: The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23, 946-967.
- Keane, M.T. (1998). *Analogical Problem Solving*. Ellis Horwood Ltd, Chichester.
- Keane, M.T., Ledgeway, T.& Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, 18, 287-334.
- Rips, L.J. (2002). Circular Reasoning. *Cognitive Science*, 26, 767-795
- Kuhn, D. (1991). *The skills of argument*, Cambridge University Press.
- Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition*, 49, 67-96.
- Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory and Cognition*, 26(5), 965-978.
- Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.
- Voss, J.F., & Van Dyke, J.A. (2001). Argumentation in Psychology: Background Comments. *Discourse Processes*, 32(2&3) 89-111.
- Zwaan, R.A., Magliano, J.P., & Graesser, A.C. (1995). Dimensions of situation-model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386-397.