A Negative Effect of Evaluation Upon Analogical Problem Solving

Christopher R. Bearman (C.Bearman@lancaster.ac.uk)
Psychology Department, Lancaster University, LA1 4YF, UK

Thomas C. Ormerod (T.Ormerod@lancaster.ac.uk)
Psychology Department, Lancaster University, LA1 4YF, UK

Linden J. Ball (L.Ball@lancaster.ac.uk)
Psychology Department, Lancaster University, LA1 4YF, UK

Abstract

Evaluation is generally considered to enhance problem solving and is strongly correlated with increasing expertise. Moreover, manipulations that increase the active processing of source problems generally promote analogical transfer of solution principles. Therefore, we expected that an instruction to evaluate the information given in problem and solution exemplars would enhance analogical problem solving. However, in Experiment 1, evaluation was found to have a detrimental effect on transfer compared with control groups instructed to summarize source problems and solutions, even when participants received additional instructions to memorize source problems for later recall. In Experiment 2, the impairing effects of instructions to evaluate were not reduced by making participants engage in evaluation when solving the target problem, a test of a 'transfer appropriate processing' explanation. We propose that instructions to evaluate lead participants to focus upon some elements of source problems and solutions at the expense of other elements required for effective transfer.

Evaluation, the set of processes involved in judging problem-related information under dimensions such as importance, value, appropriateness and scope, is generally seen as fundamental to human intellectual performance. It is central to the TOTE (Test-Operate-Test-Exit) unit proposed by Miller, Galanter and Pribram (1960) as the basis of exploratory problem-solving behaviour, and is considered to be beneficial when seeking solutions to complex real-world problems (Klein, 1999). For example, Ashton and Kennedy (2002) found that evaluation instructions led to a reduction in the use of temporally salient but otherwise inappropriate data when making auditing decisions. Evaluation is also invoked in accounts of expertise in domains such as management (Easton, 1992), medicine (Nezu & Nezu, 1995), and design (Darses, 2002; Ormerod & Ridgeway, 1999). Indeed, the proportion of problem-solving time devoted to evaluative activity has been shown to increase with domain expertise (Chi, Feltovich & Glaser, 1981; Schoenfeld, 1992).

Like evaluation, analogy is also seen as fundamental to problem solving (Gick & Holyoak, 1983; Anderson, 1989). The power of analogy as a strategy for problem solving lies in the transfer of abstract solution concepts from a source problem to a superficially different but conceptually similar target problem. Studies of real–world domains have shown its importance, both for domain experts (e.g., Blanchette & Dunbar, 2000) and novices (e.g., Bearman, Ball & Ormerod, 2001). However, laboratory studies have identified limitations upon spontaneous analogising (Dunbar, 2001). In particular, participants often fail to transfer a conceptual solution in the absence of an explicit hint that source and target problems are related (e.g., Gick & Holyoak, 1980). When spontaneous analogising does occur, it is frequently based upon superficial similarities (e.g., Keane, 1987; Holyoak & Koh, 1987).

There are, however, grounds to expect that direct instructions to evaluate might enhance analogical transfer. Studies have shown benefits from manipulations that encourage the 'active' processing of information given in a source problem and its solution. For example, Lewis and Anderson (1985) had participants actively generate a hypothesis about a piece of missing information needed to solve a geometry problem. After receiving feedback about their hypotheses, participants were more likely to select the correct solution to a target problem from a set of alternatives. Similarly, encouraging participants to generate solutions to a source problem facilitates transfer (Gick & Holyoak, 1980; Needham & Begg, 1991). To the extent that evaluation involves the active processing of source problems and solutions, it may also enhance analogical problem solving.

Instructions to evaluate may well increase active processing of source information, but they may also change other aspects of the problem-solving process. For example, evaluation may focus participants' attention upon some aspects of the source problem and solution at the expense of others, which in turn might adversely affect memory for source information. Thus, in the present study, as well as assessing participants' solution performance before and after they received a hint to analogise, we also measured the extent of their recall of source information.

Experiment 1

Experiment 1 was conducted to test the hypothesis that instructions to evaluate would influence analogical problem solving. This hypothesis is based upon an assumption that active processing enhances analogical transfer. However, there is evidence to suggest that *any* additional processing of source information can enhance analogical transfer. For example, instructions to memorize source material have been shown to increase solution rates (Needham & Begg, 1991). Therefore, we also investigated the influence of instructions to memorize source information in order to assess whether the effects of instructions to evaluate and to memorize might be additive or interactive.

Method

Design A two-factor, between-participants design was used, with Source Processing (evaluative vs. non-evaluative) as one factor, and Encoding Instruction (memory encoding vs. no memory encoding) as the other factor. The two dependent variables were the frequency of correct solutions assessed pre-hint and post-hint.

Participants Students and staff (37 female; 35 male) from Lancaster University received £3 for participating.

Materials The materials were similar to those of Gick and Holyoak (1980, 1983), with the target task being Dunker's radiation problem (see Appendix). Since Gick and Holyoak's source problem tends to produce ceiling effects on post-hint performance, a new source was developed to convey the critical 'dispersion-convergence' solution principle. This source involved a story about a nuclear power station that required a constant supply of water (see Appendix). This source analogue could be used by participants to generate solutions to the radiation problem at either a conceptual level (replacing a large force with multiple smaller forces) or at a superficial level (invoking surface-level entities such as pipes). Pilot studies indicated that the nuclear power station story produced intermediate levels of transfer (50% post-hint solution rates compared to 92% in Gick & Holyoak, 1980).

The nuclear power station story presents a source that is superficially different but conceptually similar to the radiation problem. Two distracter stories were also developed. The first was superficially similar but conceptually different to the radiation problem, and referred to an osteopath who wished to cure back pain using a novel machine (see Appendix). The second distracter was a *non-analogous* story that was neither conceptually similar nor superficially similar to the radiation problem, and described a general overthrowing a military headquarters using a bridge.

Procedure Although the basic experimental procedure closely followed the one employed by Gick and Holyoak (1980, Exp. 4), a subtle procedural change was instigated to encourage participants in the 'evaluation' conditions to believe that each presented solution was not necessarily the most effective one for solving a particular source problem, thereby promoting enhanced evaluative processing of that solution. This procedural innovation ran as follows for each source: (1) Following presentation of the source problem—which participants were asked to read through—participants were requested to select randomly a single solution from a set of five that, unbeknownst to them, were all identical; (2) Participants were then given instructions that required them either to evaluate or to summarize the problem and suggested solution. Participants in the memory conditions were instructed to attempt to memorize source problems and solutions for a later recall test. If participants lapsed into silence they were asked to keep talking. The order of the source stories was counterbalanced across participants. After processing the source stories participants worked on an interpolated task (Wason's, 1960, 2-4-6 task) for up to 10 minutes. Finally participants were given five minutes to tackle the radiation problem. If no solution had been generated in this time a hint was given that "one or more of the stories seen at the beginning of the experiment could be used to form a solution to the [target] problem". Participants were then given a further three minutes to try to solve the radiation problem.

Results and Discussion

Overall, only six participants (8%) generated a dispersion-convergence solution to the radiation problem without a hint to use the previous stories (one in the evaluation plus memory instruction present condition, two in the summary plus memory instruction absent condition, and three in the summary plus memory instruction present condition). This 8% pre-hint solution rate is comparable to the rates observed by Gick and Holyoak (1980, Exp. 4) for an analogous story and two distracters. There were no differences between conditions in the incorrect solutions produced pre-hint.

Table 1: Frequency of production of the target solution after a hint, by condition (n = 18).

	Source processing	
Encoding instruction	Evaluative	Non-evaluative
Memory encoding	5 (28%)	11 (61%)
No memory encoding	4 (22%)	9 (50%)

The post-hint solution frequencies by condition are presented in Table 1. A logistic regression using Source Processing and Encoding Instruction as predictors, and post-hint success at a dependent measure, yielded a significant model, $\chi 2 = 7.7$, $p = .02$, with Source Processing being the only reliable predictor, Wald = 6.7, $p = .009$. The data indicate that instructions to evaluate the source have a *detrimental* effect on later analogical transfer rather than a beneficial influence, as might be expected.

In relation to actual recall of the source stories during target processing, many participants (27 out of 72) were observed to have forgotten one or more of these sources. The non-analogous source was forgotten significantly *less* (just once) than either the superficially similar source (16 occasions) or the conceptually similar source (14 occasions), $\chi 2 = 12.8$, $p = .002$. However, when the number of source stories forgotten in each category (conceptually similar, superficially similar, and non-analogous) was analysed by Source Processing and Encoding Instruction no significant differences emerged. Participants were equally likely to forget either the superficially similar or the conceptually similar source story regardless of the condition that they were in.

In summary, Experiment 1 indicates that instructions to evaluate source information have a negative effect on participants' ability to invoke a conceptually similar but superficially different source analogue to create an effective solution to a target problem, despite participants having the post-hint knowledge that one of the three sources contains information that would be helpful for solving the target. These findings are clearly contrary to the expectation that evaluation should have a beneficial effect on the identification and application of useful source information during problem solving.

Experiment 2

In light of the findings of Experiment 1, Experiment 2 aimed to examine whether the apparently negative effect of source evaluation on analogical problem solving was reliable. In addition to our interest in replicating this effect, however, we also set out to assess whether instructions for encouraging participants to evaluate their putative solutions to the *target* problem might reduce the detrimental effect of source evaluation. The rationale behind this latter 'target evaluation' manipulation derived from the notion of *transfer appropriate processing* as espoused by Roediger (e.g., 1990), which captures the observation that recall of information is reliably enhanced when the processes engaged in at retrieval match those that arose during encoding. In the field of analogy research, Dunbar, Blanchette and Chung (2001) have demonstrated that if the target problem is processed in the same manner as the source problem then increased analogical transfer is found. In particular, Dunbar et al. showed increased transfer from a source problem to a 'relational' probe but not to an 'instrumental' probe when participants had been instructed that they were going to have to produce a metaphor for the source problem (an instruction claimed by the authors to lead to relational processing). Likewise, Needham and Begg (1991) found facilitation for memory tasks with memory encoding of the source story, and facilitation for problem-solving tasks with problem-solving based encoding of the source.

The findings of Dunbar et al. and Needham and Begg lend support to the idea that the detrimental effect of evaluation might be overturned if participants are requested to engage in evaluative processing of target material. In Experiment 2, therefore, we assessed such transfer appropriate processing effects by crossing the instructions that participants received during source processing (evaluative vs. non-evaluative, summary instructions) with those that they received during target processing (evaluative vs. non-evaluative).

Method

Design There were two between-participants factors: Source Processing (evaluative vs. non-evaluative) and Target Processing (evaluative vs. non-evaluative). The two dependent variables were the frequency of correct solutions assessed pre-hint and post-hint.

Participants Students and staff (40 female; 32 male) from Lancaster University received £4 for participating.

Materials and Procedure The materials and procedure were the same as those employed in the Experiment 1 conditions that did not involve a source memorization requirement. The only exception to this was that prior to attempting to solve the target (radiation) problem participants were instructed either that: (1) "It does not matter if you do not think the solutions will work, as the aim is to brainstorm the problem" (i.e., *non-evaluative target processing*); or (2) "Please try not to provide solutions that will not work, as the aim is to produce the best possible solution" (i.e., *evaluative target processing*). In addition, after participants had attempted (whether successfully or unsuccessfully) to solve the radiation problem, they were finally requested to recreate the 'salient aspects' of the story about the nuclear power station (the conceptually similar and superficially different source).

Results and Discussion

Overall, only a single participant produced the dispersion-convergence solution to the radiation problem before a hint (1.3%). This participant was in

the condition that involved neither evaluative source processing nor evaluative target processing. This 1.3% pre-hint solution rate represents a significant drop from the standard 8% found by Gick and Holyoak (1980). There were no differences between conditions in the incorrect solutions produced pre-hint.

Post-hint solutions frequencies are presented in Table 2. A logistic regression using Source Processing and Target Processing as predictors of success yielded a significant model, $\chi 2 = 13, p = .01$, with Source Processing being the only significant predictor (Wald = 7, p = .008).

Table 2: Frequency of production of the target solution after a hint, by condition (n = 18).

	Source processing	
Target processing	Evaluative	Non-evaluative
Evaluative	4 (22%)	9 (50%)
Non-evaluative	3 (16%)	9 (50%)

In terms of memory for the source stories during target processing, participants were observed to have forgotten the superficially similar source on 18 occasions, the conceptually similar story on five occasions, and the non-analogous source on one occasion. These differences were reliable, $\chi 2 = 19.8$ (2), $p < .001$. As in Experiment 1, however, there were no significant differences between the conditions in terms of forgetting the source stories.

In order to investigate further what may be causing the detrimental effect of source-evaluation instructions, participants' recall of the nuclear power station story was investigated after the attempt (whether successful or unsuccessful) at the radiation task. The nuclear power station story can be decomposed into several critical features according to what is relevant for later transfer to the radiation problem. These features include *lower* forces in the pipes (corresponding to lower intensity X-rays in the radiation problem), *multiple* pipes (analogous to multiple X-rays), and *converging* forces (the same in both solutions). The presence or absence of these features in the participants' recall of the nuclear power station story thus formed a three-point coding scheme for investigating recall of the critical components of the source story. Participants in the evaluative source-processing conditions produced a significantly lower average score (1.8, SD = 1.6) than participants in the non-evaluative source-processing conditions (2.6, SD = 1.6), $F(1,65) = 4.4$, $p = .04$. This seemed to be attributable to participants in the evaluative source-processing conditions having a reduced tendency to mention 'lower forces' (18 in the evaluative conditions vs. 24 in the non-evaluative conditions), and a reduced tendency to mention 'converging forces' (nine in the evaluative conditions vs. 14 in the non-evaluative conditions).

As a verification of this latter result, the actual solutions that participants generated from the analogous source story post-hint were investigated. In the same way as the story recall was analysed for the presence of critical components, so too can the solution generated from the analogous source be analysed for the presence of any mention of *lower* forces, *multiple* forces, and *converging* forces. In addition, solution generation based on the analogous source was scored for the mention of *X-rays*, making this measure a four-point score. This score is inferior to the one based purely on story recall since a measure based on solution generation will be correlated with task success, which is already known to differ between the groups (i.e, a person producing a correct solution will have generated all four critical solution components). Notwithstanding this caveat, however, the measure of story recall showed that participants in the evaluative source-processing conditions had a significantly lower mean score for the presence of the critical components of the source information (1.7, SD = 0.8) than participants in the non-evaluative source-processing conditions (2.1, SD = 0.7), $F(1,65) = 4.5$, $p = .03$.

In addition, it was found that the number of words of the conceptually similar source story recalled by the participants in the evaluative source-processing conditions was significantly lower (56 words, SD = 22.5) than the number of words recalled by participants in the non-evaluative source-processing conditions (70.5 words, SD = 22.6), $F(1,62) = 6.6$, $p = .01$. This further supports the idea that participants in the evaluation conditions have developed a generally impoverished representation of the source material.

In summary, the detrimental effect of evaluative processing of source stories on post-hint analogizing found in Experiment 1 was replicated in Experiment 2. It was also evident that engaging participants in transfer appropriate processing (i.e., requiring evaluation of the target problem as well as the source information) could not overturn this effect. In addition, it was observed that participants who evaluated the source information did not retain as much of that information as did people who merely summarized the material. It thus seems that focusing processing effort on the quality of a solution leads to important information being lost from the representation of the source analogue.

General Discussion

Experiments 1 and 2 have demonstrated that evaluative source processing can lead to reliably worse performance than non-evaluative source processing in an analogical problem-solving context requiring the application of prior solution ideas to a conceptually similar but superficially different problem encountered

just 10 minutes after initially examining source information. In addition, it has been shown that a superficially similar but conceptually different source is recalled significantly less often than either a conceptually similar but superficially different source or a non-analogous source.

These striking failures of analogical transfer with source-evaluation instructions appear, on initial consideration, to fly in the face of previous research that has demonstrated the benefits of active source processing on analogical problem solving. What, then, may be the reason for these seemingly discrepant findings? In order for analogical transfer to occur it is commonly claimed (e.g., Gick & Holyoak, 1983) that it is necessary for individuals to represent source information at a relatively *deep* level of abstraction whereby the surface features of the problem and solution are less salient than the underlying solution principle itself. The process of 'solution generation' appears to encourage participants to represent the deep structure of source information due to its requirement that participants engage fully with the problem and its potential solutions. Thus, the difference between the present experiments and previous studies that have revealed the benefits of source processing may well hinge on the fact that solution evaluation carries no requirement to generate information. As such, the representation of source information associated with evaluative processing may not be sufficiently enriched as would be the case if the representation derived from a generative process. Instead, source representations appear to be restricted in nature—perhaps because participants focus in *depth* upon a few specific aspects of the problem and solution at the expense of others.

Rather than arguing that evaluation produces a *negative* effect on transfer of an analogous solution, however, another possible interpretation of the present dataset is that the process of summarizing source stories actually has a *positive* effect on transfer. Summarizing source information might well lead to the creation of an abstract representation of such information, since the details in the story will be reduced and replaced with higher-level ideas and concepts. Such reduction and replacement of detail might not occur with evaluative processing, where the focus is on considering whether a solution is going to be effective or not. This would be consistent with previous research findings showing that explicitly instructing participants to produce an abstract representation of source material has beneficial effects on pre-hint solution-transfer rates (Gick & Holyoak, 1983; Mandler & Orlich, 1993).

A further possible interpretation of our findings is that participants in the evaluative-processing conditions may have been less effective at producing the analogous dispersion-convergence solution because they may have believed that the dispersion-convergence solution encountered during the first phase of the study was in some sense a sub-optimal way to tackle the nuclear power station problem. However, most participants were actually observed to consider that dispersion-convergence was a very good solution to the power station problem. In addition, a qualitative assessment of the data indicated that dismissing or accepting the dispersion-convergence solution to the nuclear power station source problem was not related to task success. Finally, Experiment 1 was explicitly designed to encourage the production of a wide range of solution ideas even if these were felt to be sub-optimal. It should be noted, too, that participants were scored as having produced the dispersion-convergence solution even if they didn't think that it would work successfully.

Overall, then, it seems that contrary to initial expectations, evaluative processing is reliably less effective at promoting analogical transfer than non-evaluative summarization of source analogues. This effect appears to be caused by evaluation instructions encouraging participants to focus on only some aspects of the source information at the expense of others. Summarizing the source information does not promote such focusing, and may lead to a broader and more flexible representation of source information that can be invoked and applied in subsequent problem solving. The finding that evaluative processing has an apparently detrimental effect on post-hint transfer is not only theoretically interesting, but also crucially important when possible applications of analogy research to real-world problem solving are considered. For example, since evaluative processing of solution ideas is a vital aspect of domain-based problem solving, then it would seem important to explore mechanisms that might eliminate the negative effect of evaluative source processing on subsequent analogising.

References

Anderson, J. R. (1989). The analogical origins of errors in problem solving. In D. Klahr, & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon.* Hillsdale, NJ: LEA.

Ashton, R H., & Kennedy, J. (2002). Eliminating recency with self-review: The case of auditors 'going concern' judgements. *Journal of Behavioural Decision Making*, 15, 221-231.

Bearman, C. R., Ball, L. J., & Ormerod, T. C. (2002). An exploration of real-world analogical problem solving in novices. In W.D. Gray, & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 101-106). Mahwah, NJ: LEA.

Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory and Cognition, 28,* 108-124.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Darses, F. (2002). A framework for continuous design of production systems and its application in collective re-design of production line equipment. *Human Factors & Ergonomics in Manufacturing, 12*, 55-74.

Dunbar, K. (2001). The analogy paradox: Why analogy is so easy in naturalistic settings, yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *Analogy: Perspectives from cognitive science*. Cambridge, MA: MIT Press.

Dunbar, K., Blanchette, I., & Chung, T. (2001). *Goals, encoding, and analog retrieval.* Unpublished manuscript, Department of Psychology, McGill University.

Easton, G. (1992). *Learning from case studies (2nd Edn.)*. New York: Prentice Hall.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition, 15,* 332-340.

Keane, M. (1987). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology, 39A,* 29-41.

Klein, G. (1999). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Lewis, M. W., & Anderson, J. R. (1985).Discrimination of operator schemata in problem solving: Learning from examples. *Cognitive Psychology, 17,* 26-65.

Mandler, J. E., & Orlich, F. (1993). Analogical transfer: The roles of schema abstraction and awareness. *Bulletin of the Psychonomic Society, 31,* 485-487.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the structured of behaviour. New York: Holt, Reinhhart and Winston.

Nezu, C. M., & Nezu, A. M. (1995). Clinical decision making in everyday practice: The science in art. *Cognitive and Behavioural Practice, 2*, 5-25.

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer, memory-oriented training promoted memory for training. *Memory and Cognition, 19*, 5453-557.

Ormerod, T. C., & Ridgeway, J. (1999). Developing task design guides through cognitive studies of expertise. *Proceedings of the Third European Conference on Cognitive Science*. Sienna: Consiglio Nazionale Della Richerche.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45,* 1043-1056.

Schoenfeld, (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129-140.

Appendix

The Radiation Problem

A doctor is faced with a patient who has a malignant tumour in his stomach. It is impossible to operate on the patient, but unless the tumour is destroyed the patient will die. X-rays can be used to destroy the tumour. If an X-ray reaches it at sufficiently high intensity the tumour will be destroyed. Unfortunately, at this intensity the healthy tissue that the ray passes through on the way to the tumour will also be destroyed. At lower intensities the ray is harmless to healthy tissue, but it will not affect the tumour either. It looked like the patient was going to die.

The Nuclear Power Station Story

A water pumping station in the hills above Peshawar has recently been adapted to feed a nuclear power facility, where the water acts as a coolant. Unfortunately the main pipe that leads from the water pumping station to the nuclear facility cannot stand the increased pressure and will eventually rupture. The nuclear power facility must have a continuous supply at a certain volume or it will overheat and go into meltdown. Experts have found that it is impossible to build a single large pipe to carry the water with the materials available that will not also rupture due to the high pressure of the water flowing through it. The suggestion is: to build a number of pipes that have less water pressure.

The Osteopath Story

After many years of research, a Philadelphia doctor of Osteopathy thinks that he has found a way to alleviate chronic back pain. The doctor theorizes that if a high velocity thrust can be applied to a fairly wide section of the lumber region of a person's back they will be cured. Unfortunately, the doctor has found it impossible to test his theory because he is unable to deliver a high velocity thrust of sufficient force because such a thrust requires more strength than the doctor possesses. The suggestion is: to have the doctor develop a machine that is capable of delivering the high velocity thrust.