The facilitatory effect of negative feedback on the emergence of analogical reasoning abilities

Linden J. Ball*, Alison M. Hoyle and Andrea S. T owse
Lancaster University, UK

This paper focuses on the development of analogical reasoning abilities in 5- and 6-year-old children. Our particular interest relates to the way in which analogizing is influenced by the provision of task-based feedback coupled with a self-explanation requirement. Both feedback and self-explanation provide children with opportunities to engage in self-reflective thinking about the process of analogical reasoning. To examine the role of such metacognitive factors in analogical strategy development the reported study combined a proportional analogy paradigm with a small-scale microgenetic approach involving multiple testing sessions over a restricted time period. The key manipulation involved exposing participants either to the correct or incorrect analogy completions of another reasoner that they were then asked to explain. The data revealed that the development of an effective analogizing strategy embodying a ‘relational shift’ from superficial to relational responding was modulated by the feedback condition that the child was placed in, with a negative feedback intervention providing the greatest developmental benefit. We suggest that the value of negative feedback for the acquisition of analogical reasoning abilities derives from the way in which a self-reflective analysis of the reasons for erroneous responses sensitizes the child to a deeper understanding of how to make effective relational mappings.

Analogical reasoning involves accessing and transferring previously acquired knowledge of objects, attributes, and relations to support current problem solving. Such reasoning has long been viewed as central to intelligent thought and creative cognition (e.g., Gentner & Stevens, 1983; Holyoak & Thagard, 1995), and recent studies have confirmed the importance of analogizing in a multitude of real-world domains, including those associated with scientific discovery (e.g., Dunbar & Blanchette, 2001), organizational management (e.g., Bearman, Ball, & Ormerod, 2007; Thompson, Gentner, & Loewenstein, 2000), legal argumentation (e.g., Marchant, Robinson, Anderson, & Schadewald, 1993), and innovative product design (e.g., Ball & Christensen, 2009; Christensen & Schunn, 2007).

* Correspondence should be addressed to Dr Linden J. Ball, Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK (e-mail: L.Ball@lancaster.ac.uk).

DOI:10.1348/026151009X461744
Not only is analogizing important in adult life, but it also features as a key cognitive ability in children. Older children have been found to use analogies to enhance their understanding of concepts in biology (Brown & Kane, 1988) and physics (Pauen & Wilkening, 1997). Very young children, including preverbal infants, have likewise been shown to reason spontaneously by analogy to solve problems (Chen, Sanchez, & Campbell, 1997; Goswami & Brown, 1989, 1990; Siegler & Svetina, 2002; Tunteler & Resing, 2002). However, despite this evidence supporting the capacity for children to engage in analogical reasoning, our understanding of the key factors that impact on the development of this ability remains piecemeal. Our aim in this paper is to examine one factor that has been given only limited research attention, that is, the child’s exposure to task-related feedback, especially feedback concerning erroneous performance.

Before reporting our study and its theoretical rationale we initially review the relevant literature. First, we examine the current state of research concerning the developmental determinants of analogical reasoning abilities. Second, we summarize key theoretical issues arising from this research that suggest avenues for further developmental studies. Third, we explore how the acquisition of analogical reasoning skills may be enhanced by exposure to task-based errors that afford the individual opportunities for re-conceptualizing their understanding of solution strategies. A key outcome of our literature review is the proposal that the development of analogizing skills can be facilitated through self-reflective, metacognitive processes cued by the provision of task feedback.

Factors influencing the development of analogical reasoning abilities

There is a general consensus that analogical reasoning skills develop significantly over the first few years of school (Hilford, 1993; Siegler & Svetina, 2002). Three main factors have been forwarded to explain the emergence of increasingly sophisticated analogical reasoning skills. One factor concerns the accretion of domain knowledge (e.g., Goswami, 1992, 2002; Goswami & Brown, 1989), and suggests that as children learn more about the world they are able to use this increasing knowledge to reason about the relations between items. This knowledge-based account therefore emphasizes the role of domain familiarity in the development of analogical reasoning. Children who are reasoning in familiar domains can make correct analogies, whereas children who lack domain knowledge will not understand the structural relations between items, which will lead them to focus on surface features (Goswami, 1992, 2002).

A second factor that has been implicated in the development of analogical reasoning relates to children’s abilities to ‘re-represent’ relations (e.g., Gentner, Rattermann, Markman, & Kotovsky, 1995). This capacity for re-representation is claimed to allow children to move from first-order predicate–argument representations such as larger(elephant, mouse) as a representation of the notion that ‘an elephant is larger than a mouse’, to higher-order predicate structures that support more complex analogical mappings such as greater[size(elephant), size(mouse)]. A key upshot of this ability to represent higher-order predicates is that children undergo a ‘relational shift’ (Gentner & Toupin, 1986; Rattermann & Gentner, 1998), whereby their analogical reasoning changes over time from being based initially on the surface features of stimuli (e.g., objects and attributes) to the inclusion of information about the relations between objects, eventually incorporating whole systems of relations including hierarchically nested relational structures.

A third factor claimed to play a pivotal role in the acquisition of analogical reasoning skill relates to domain-general maturational changes in working memory during
development, including changes in its capacity as well as attentional and inhibitory mechanisms associated with executive functioning (e.g., Andrews & Halford, 2002; Halford, Wilson, & Phillips, 1998). Halford and colleagues claim that one of the core constraints arising in cognitive development concerns the maximum relational complexity that can be concurrently processed in working memory (see Hummel & Holyoak, 1997, for related proposals). As such, the emergence of age-related changes in working memory facilitate children's ability to represent multiple dimensions, which, in turn, gives rise to improved performance on tasks involving more complex analogical mappings.

**Issues in developmental research on analogy**

All three of these aforementioned factors have a plausible role to play in the emergence of analogical reasoning abilities and there is much empirical support for their involvement (e.g., Gordon & Moser, 2007; Goswami, 2002; Halford, 2005; Holyoak & Thagard, 1995; Richland, Morrison, & Holyoak, 2006). At the same time, a fair degree of debate surrounds the area of analogy development, and conclusions favouring one perspective are often contested by those of a different theoretical persuasion. As a case in point, the evidence forwarded by Goswami (1998) and Rattermann and Gentner (1998) indicating that children of less than 5 years can form effective analogies involving multiple relations appears to be at odds with a maturational view of analogy development. Yet Halford et al. (1998) argue that this apparent contradiction can be explained in terms of the child dealing with ternary relations by decomposing them into more manageable binary ones. Moreover, Halford and colleagues acknowledge the criticality of knowledge accretion for effective analogizing, such that tensions between theorists really hinge more on matters of emphasis than on fundamental disputes as to the credibility of alternative perspectives.

Our view of the developmental literature on analogizing is coloured by previous research that has used the ‘microgenetic’ method to explore the acquisition of analogizing skills (e.g., Siegler & Svetina, 2002). The microgenetic method involves intensive testing over a period of developmental change, and provides a rich dataset that can afford a detailed understanding of how children's skills develop (Siegler, 2002, 2006; Siegler & Crowley, 1991). For example, Siegler and Svetina (2002) deployed the microgenetic method to examine analogizing skills with ‘matrix completion’ tasks. These embody proportional analogies of the form ‘a:b::c: . . . ?’, with solutions requiring relational mappings from the source to the target domain. Task presentation entails a participant completing a 2 × 2 item grid with the bottom-right item left blank. If the correct item is selected from a presented response-set then the top items in the grid should be related in the same way as the bottom items and the left-hand items should be related in the same way as the right-hand items.

In their study, Siegler and Svetina (2002) provided children with feedback on their performance on each trial (giving the correct solution if an error had been made), and also elicited explanations as to why an answer was correct. They observed that during 3 weeks of intensive testing children stopped using a dominant incorrect strategy (based on perceptual similarity) and increased their use of a correct strategy (based on relational mapping), a finding that supports Gentner's relational shift hypothesis (e.g., Gentner & Toupin 1986; Rattermann & Gentner, 1998). Siegler and Svetina could not, however, pinpoint the casual factors underpinning the shift from superficial to relational responding, and suggest that effective strategy development could have been due to explanation or feedback or both.
Two aspects of Siegler and Svetina’s (2002) microgenetic research on analogy development are especially appealing, and provide a counterpoint to domain-general and domain-specific accounts. First, their findings suggest that analogizing skills may develop through factors associated with self-explanation and/or feedback that have little to do with maturational changes in working memory (since microgenetic time scales are too short to permit maturation of core cognitive capacities) or knowledge accretion (since matrix completion trials are designed to embody relations that are easily understood by children). Second, their emphasis on analogizing as a strategy is conceptually very rich, given its attendant links to the notion that analogizing entails explicit, high-level thinking based around structure mapping and inferential reasoning tied to symbolic representations and attentional control mechanisms (Hummel & Holyoak, 1997; Spellman, Holyoak, & Morrison, 2001). Although the claim that analogizing involves explicit high-level cognition is not uncontested (see Leech, Mareschal, & Cooper, 2008, for an alternative relational-priming account that entails subsymbolic, paired-associate learning within connectionist networks), the strength of the strategy view is that it readily lends itself to an explanation of the sophisticated analogical mapping and inference processes that are a hallmark of adult-level competence (Holyoak & Hummel, 2008; Schwering & Kühnberger, 2008).

Recent microgenetic research on the development of analogical skills (Cheshire, Ball, & Lewis, 2005, 2008; Cheshire, Muldoon, Francis, Lewis, & Ball, 2007) has been inspired by Siegler and Svetina’s (2002) view of analogizing as a strategy, whose acquisition can be bootstrapped through the provision of opportunities for deliberative, analytic thinking. For example, Cheshire et al. (2005) demonstrated that self-explanation and feedback are both positively associated with analogy development, with the greatest benefits arising when feedback and self-explanation are combined. Moreover, the benefits arising from feedback were seen to be highly enduring, whereas the benefits of self-explanation were transitory, with children reverting to preferences for superficial object similarity when no longer asked to explain. To account for the differential effects of self-explanation and feedback on analogy development, Cheshire et al. (2005) suggest that self-explanation supports on-line metacognitive processing by guiding participants to maintain an attentional focus on task goals (cf. Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995); whilst feedback supports the completion of task-based mental models and an understanding of what is relevant to success.

The value of feedback as a factor that can enhance the acquisition of an effective and resilient analogical-reasoning strategy strikes us as especially interesting since it suggests that feedback may induce potent self-reflective learning processes than lead to deep-levels of task-based understanding. Indeed, in Cheshire et al.’s (2005) study not only did feedback give rise to enduring benefits in analogical reasoning but it also led to the transfer of analogy strategies to a final testing session that involved problem contents that were very different to those that children had been exposed to in earlier sessions.

**Aim of the study**

Notwithstanding the value of self-explanation and feedback for analogy development, there are some important issues surrounding the influence of these factors that require further examination. One such issue forms the focus of the present paper, and relates to whether benefits for the development of effective analogizing may arise from systematic exposure to another person’s correct or incorrect responding, as distinct from previous research that has only explored the influence of feedback relating to the child’s own
correct or incorrect responding (e.g., Cheshire et al., 2005; Gholson, Smither, Buhrman, Duncan, & Pierce, 1996; Siegler & Svetina, 2002). Our motivation for examining this feedback issue is twofold, and draws on both methodological and theoretical concerns.

In terms of methodology, we note that studies where the researcher corrects a child’s erroneous analogy responses by way of feedback suffer from limitations in experimental control. This is because there will be differences between children in the number of correct and incorrect trials that arise in a test session, as well as in the types of error that each child makes. What this means is that each child ends up being subjected to a unique experimental procedure that is different to that experienced by another child, making it difficult to derive a conclusive interpretation of findings. For example, although it is apparent from previous studies that feedback gives rise to better learning than no feedback, it is uncertain whether positive and negative feedback are equally advantageous, or instead whether one type of feedback is particularly beneficial for skill development. In the present study, we therefore systematically manipulated the valence of the feedback assigned to participants in a manner that is distinct from feedback approaches adopted in previous studies of analogical reasoning with children (e.g., Cheshire et al., 2005; Gholson et al., 1996; Siegler & Svetina, 2002) and adults (e.g., Cummins, 1992; Price & Driscoll, 1997), where feedback related to a reasoner’s own responses. Thus, in one condition of our experiment, all participants were exposed to another reasoner’s erroneous responses (negative feedback condition) whilst in the other condition all participants were exposed to another reasoner’s correct responses (positive feedback condition). This novel manipulation permitted an unambiguous assessment of the relative benefits of positive versus negative feedback on analogical skill development.

In terms of theory, there are important a priori reasons for systematically examining the valence of the feedback that participants are exposed to during skill acquisition. In particular, there is evidence from research with both adults and children suggesting that feedback concerning failed task performance may be highly beneficial for the acquisition and application of conceptual knowledge and procedural skills (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Freeman, Antonucci, & Lewis, 1999; Gopnik & Astington, 1988; Muldoon, Lewis, & Berridge, 2007; Muldoon, Lewis, & Freeman, 2003, 2009; Rittle-Johnson, 2006). For example, Muldoon et al. (2003) showed that the opportunity for children to reflect on why counting responses made by another individual were incorrect had a striking effect on their emerging conceptualization of what is achieved by counting. Children who were able to reason about miscounts were more likely to count when asked to compare two sets of items – and to succeed in creating numerically equivalent sets – than children who were unable to reason about another’s miscounts.

Rittle-Johnson (2006) also examined how a feedback manipulation that included opportunities to reason about another child’s errors influenced the acquisition of skills, in this case in relation to solving mathematical equivalence problems (e.g., $4 + 9 + 6 = 4 +$), which tap the idea that the two sides of an equation represent the same quantity. It was found that feedback accompanied with a self-explanation request had a significant benefit on the learning and transfer of procedural skills relevant to solving equivalence problems, with this benefit persisting on a delayed post-test. Rittle-Johnson (2006) suggests that a range of mechanisms underpin the facilitated skill acquisition that arises from self-explanation combined with feedback, including:

1. enhanced invention of new problem solving approaches that assist transitions to more efficient procedures (cf. Siegler & Jenkins, 1989);
2. broadening of the range of
problems to which children accurately apply correct procedures (cf. Anderson, 1993); and (3) flexible adaptation of procedures to solve novel problems that do not allow rote application of the procedure.

In summary, there are sound methodological and theoretical reasons to predict that advantages may arise for the acquisition of analogical reasoning skills from the provision of feedback on another individual's failed task performance. The present study set out to test this prediction. We employed a proportional analogy paradigm based on picture analogies (cf. Goswami, 2002) and a small-scale microgenetic approach that involved three test sessions analysing changes in reasoning ability coupled with a feedback intervention prior to the second test session, which exposed participants either to another reasoner's correct analogy responses or to another reasoner's erroneous analogy responses. In addition, given the evident benefits for analogizing that arise from combining feedback with an explanation requirement (e.g., Cheshire et al., 2005; Rittle-Johnson, 2006; Siegler & Svetina, 2002), we likewise asked all participants to provide an explanatory rationale for the correct or incorrect feedback that they were subjected to. In terms of specific hypotheses, we expected that negative feedback (i.e., opportunities to reflect on incorrect responses) would facilitate enhanced strategy development (i.e., more correct analogical mappings and less superficial responses) relative to positive feedback (i.e., opportunities to reflect on correct responses).

Method

Participants

Twenty-seven 5- to 6-year-old children (13 female; 14 male) participated in this experiment (mean age: 5 years and 6 months; age range: 5 years and 1 month to 6 years and 9 months). All came from a local primary school and were in the same class.

Design

The participants took part in three test sessions that each involved them tackling a series of picture-based proportional analogies. Testing took place over a 1-week period, with 1 day between each test session. Test Session 1 involved an initial set of 10 analogy trials that children had to complete, which provided a baseline measure of analogy performance. Three children were excluded from the experiment after Test Session 1 since they scored over 80%, indicating a level of understanding that was too advanced to reveal further developmental progression (cf. Cheshire et al., 2005).

Immediately prior to Test Session 2 the remaining 24 children in the study were matched into pairs on the basis of age, gender, and ability (using a teacher-based assessment that did not involve their Session 1 performance). Members of these matched pairs were then randomly assigned either to the positive and negative feedback conditions and viewed a series of 10 new, completed analogies. The children in the positive feedback condition were provided with correct answers to these analogy problems whilst the children in the negative feedback condition were provided with incorrect answers to the same analogy problems. Directly following this feedback manipulation all participants then entered Test Session 2 and tackled a series of 10 new analogy problems in order to provide a measure of post-feedback analogy performance. Test Session 3 involved another 10 new analogy problems and provided a measure of the durability of any performance effects that emerged in Test Session 2.
The experiment thus embodied a two-factor mixed design, with the within-participants factor being test session (Test Sessions 1, 2, and 3) and the between-participants factor being feedback condition (positive vs. negative). All three test sessions provided performance data in terms of the proportions of correct and erroneous responses.

**Task and materials**
The experiment involved 42 picture-based proportional analogies in total that were modelled on stimuli devised by Goswami (2002). These analogies included relatively complex relational mappings such as: ‘Dog:Kennel::Bird:Nest’ (relation = inhabits); ‘Toothpaste:Teeth::Shampoo:Hair’ (relation = cleans); and ‘Hat:Head::Glove:Hand’ (relation = insulates). No analogies were repeated; 10 were used in each of the three test sessions, and a further 10 were used for the feedback trials that immediately preceded Test Session 2. The remaining 2 analogies were used as examples prior to the commencement of Test Session 1. Counterbalancing was employed to ensure that each set of 10 analogies appeared an equal number of times in each of the three test sessions. However, the 10 analogies used in the feedback session were always the same. The actual order of the 10 analogies appearing in the test and feedback sessions was independently randomized for each participant.

The analogies used in the test sessions were displayed as individual problems, mounted on black card and presented as two pictures being equivalent to another two pictures, the second of which was always a question-mark. The analogies used in the feedback session were presented in a similar manner except that the question-mark was replaced by either a correct or an incorrect item, depending upon the feedback condition. In the negative feedback condition the incorrect picture used to complete the analogy always had some semantic association with the ‘c’ term (e.g., rather than the correct response of a ‘Sliced Apple’ for the analogy ‘Bread:Sliced Bread::Apple:...?’ an ‘Orange’ was presented). All analogies were depicted as pictures extracted from Microsoft clip-art.

In the feedback session, a fictional character, Mr Splodge, was invoked as the protagonist who had made correct or incorrect responses on all trials. Mr Splodge was displayed pictorially on positive feedback trials (correct answers) as having a happy expression, whilst on negative feedback trials (incorrect answers) he had a sad expression.

**Procedure**
Test and feedback sessions took place one-to-one with the experimenter in a quiet corner of the school library. Each participant’s answer sheet was filled in by the experimenter during testing. No indication was provided to the child of their success or failure on any trial. Prior to Test Session 1, all participants were introduced to the nature of the task using two example proportional analogies that included correct solutions. For both analogies participants were simply asked if they could see a pattern between the first two pictures and the last two and to explain this pattern. Children unable to do this were provided with a standardized explanation of the pattern by the experimenter. At this point all participants were also given an ethical briefing which stressed that they were not to worry if they did not have an answer to any subsequent problems that they would be presented with. In Test Session 1, the participants were presented with 10
analogy and were required to generate their own answer for the missing picture that would fill in the space denoted by the question-mark. On each trial the experimenter spoke aloud the verbal label for the depicted images. For example, the experimenter would state: ‘Dog is to Kennel as Bird is to . . . ?’. The child simply needed to respond with a word-based answer.

After a day’s respite all participants received the feedback manipulation, experiencing either positive or negative feedback, and were required to provide explanations as to why Mr Splodge had got the analogy correct (happy expression) or incorrect (sad expression). Children were carefully introduced to the idea that Mr Splodge’s happy expression denoted a correct response whereas his sad expression denoted an incorrect response. Their understanding of this was checked via questioning and no children demonstrated any difficulty with the expression-correctness mapping. Test Session 2 followed directly on from the feedback session and was a performance test that used a different set of 10 analogies to those previously seen. After another day’s rest, Test Session 3 consisted of another performance test, again with a different set of 10 analogies.

Results and discussion

Correct answers

Figure 1 displays the mean proportion of all responses that were coded as correct answers within each feedback condition at each test session. These data suggest that whilst children started off with similar performance levels, there was a subsequent divergence in correct responding after the feedback manipulation, with those in the negative feedback condition showing more pronounced facilitation of analogical reasoning than those in the positive feedback condition. The data also suggest that the performance differences emerged directly post-feedback (i.e., at Test Session 2), and that subsequent performance improvements over time were equivalent for the two feedback conditions (i.e., the separation between conditions remained stable from Test Sessions 2–3).

![Figure 1](image_url)

**Figure 1.** Mean proportion of all responses that were coded as correct answers within each feedback condition at each test session (error bars depict standard error of the mean).
The application of a mixed between–within analysis of variance (ANOVA) to examine the proportion of responses that were coded as correct answers revealed a significant main effect of test session, $F(2, 44) = 114.45$, MSE = 0.006, $p < .001$, $\eta_p^2 = .84$, but no main effect of feedback condition, $F(1, 22) = 2.76$, MSE = 0.048, $p = .11$, $\eta_p^2 = .11$. However, a reliable interaction was evident between feedback condition and test session, $F(2, 44) = 11.86$, MSE = 0.006, $p < .001$, $\eta_p^2 = .35$.

Simple main effects analyses indicated no reliable difference in the proportion of correct answers between feedback conditions at Test Session 1, highlighting the similar level of initial task performance, $F(1, 33.87) = 0.49$, MSE = 0.02, $p = .49$. Significant differences between feedback conditions were, however, evident at Test Session 2, $F(1, 33.87) = 5.25$, MSE = 0.02, $p = .03$, and at Test Session 3, $F(1, 33.87) = 8.20$, MSE = 0.02, $p = .007$, supporting the emergence of a performance separation after the critical feedback session.

Further simple main effects analyses were pursued examining correct performance at the two feedback levels. A reliable effect of test session was evident for the negative feedback condition, $F(2, 22) = 72.61$, MSE = 0.009, $p < .001$, with Bonferroni-corrected pairwise comparisons revealing that performance at Test Session 2 was better than at Test Session 1 ($p < .001$) and that performance at Test Session 3 was better than at both Test Session 2 ($p = .003$) and Test Session 1 ($p < .001$). Thus, for the negative feedback condition there was evidently a steady increase in correct performance following the feedback intervention, with participants doubling their initial level of correct performance by the end of the study and getting close to maximum scores (i.e., 87% correct).

In the positive feedback condition, a reliable effect of test session on the proportion of correct answers was also in evidence, $F(2, 22) = 43.11$, MSE = 0.004, $p < .001$, with Bonferroni-corrected pairwise comparisons again revealing that correct performance at Test Session 2 was higher than at Test Session 1 ($p = .007$), and that correct performance at Test Session 3 was higher than at both Test Session 2 ($p = .003$) and Test Session 1 ($p < .001$). Overall, the positive feedback condition revealed a similar steady increase in the proportion of correct answers following the feedback intervention as the negative feedback condition, but with participants failing to exhibit as great an advantage at Test Sessions 2 and 3 compared with their counterparts in the negative feedback condition.

**Incorrect responses**
We also examined the responses that were deemed to be incorrect within each condition at each test session. A coding scheme was devised using three categories to denote different types of incorrect response: (1) superficial answers (i.e., answers that simply repeated the name of an object or an attribute in the presented pictures or that indicated that some kind of generalization had occurred); (2) structural errors (i.e., incorrect ‘relational’ responses that involved transferring an inappropriate relation from the ‘a:b’ picture pair to the ‘c: . . . ’ situation, but that were still suggestive of a higher level of inferential processing taking place than superficial responses); and (3) uncodable responses (i.e., answers that appeared to have no connection to the pictures presented, as well as null responses, of which there were very few, i.e., 7%). All incorrect responses were allocated to their appropriate category by the primary coder (the second author of this paper). We checked the reliability of the primary coder’s categorization by asking a second coder who had no association with the research to
categorize the full set of incorrect responses (note that determining correct responses was error-free since only a single, predefined target answer was available for each analogy). The second coder conducted their categorization on a randomized participant-by-participant basis whilst being blind to the experimental condition that each participant had been in. This procedure confirmed the reliability of the primary coder’s categorization of incorrect responses, with an inter-rater coding agreement of 94%. Coding discrepancies showed no systematic association with any particular category. Given the good degree of reliability evident in the coding of incorrect responses it was decided to use the primary coder’s categorization in all subsequent statistical analyses.

Superficial answers

The first data concerning incorrect responding that we examined were the proportion of all responses that were coded as superficial answers within each feedback condition at each test session (Figure 2). Superficial answers were the dominant form of incorrect response at Test Session 1, with broadly equivalent proportions arising in both of the feedback conditions, again confirming the similarity between these groups’ initial performance. The proportion of superficial answers decreased more rapidly after the negative feedback intervention than after the positive feedback intervention. By Test Session 3 there was still a performance gap between the two feedback conditions in terms of the proportion of superficial answers, but the performance difference was less pronounced than at Test Session 2.

A mixed between–within ANOVA was undertaken on the proportion of all responses that were coded as superficial answers. This revealed a significant main effect of test session, $F(2, 44) = 35.37$, $MSE = 0.01$, $p < .001$, $\eta^2_p = 0.62$, but no main effect of feedback condition, $F(1, 22) = 2.25$, $MSE = 0.02$, $p = .15$, $\eta^2_p = .09$. The interaction between test session and feedback condition was significant, $F(2, 44) = 3.18$, $MSE = 0.01$, $p = .049$, $\eta^2_p = .13$.

![Figure 2. Mean proportion of all responses that were coded as superficial answers (errors) within each feedback condition at each test session (error bars depict standard error of the mean).](image-url)
Simple main effects analyses revealed no reliable difference in superficial answers between feedback conditions at Test Session 1, \(F(1, 55.54) = 0.29, \text{MSE} = 0.01, p = .59\), supporting the claim that levels of task performance started off as equivalent for both groups. At Test Session 2, however, a significant difference in superficial answers was observed between feedback conditions, \(F(1, 55.54) = 5.94, \text{MSE} = 0.01, p = .018\), but this difference was no longer reliable at Test Session 3, \(F(1, 55.54) = 1.96, \text{MSE} = 0.01, p = .17\).

Further simple main effects analyses were pursued examining the proportion of superficial answers at the two feedback levels. A significant effect of test session was evident for the negative feedback condition, \(F(2, 22) = 24.53, \text{MSE} = 0.01, p < .001\), with Bonferroni-corrected pairwise comparisons indicating that lower proportions of superficial answers were made at Test Sessions 2 and 3 than at Test Session 1 (\(p = .003\) and \(p < .001\), respectively), but there was no reliable difference in the proportion of superficial answers between Test Sessions 2 and 3 (\(p = .23\)). Thus, for the negative feedback condition there was evidently a notable reduction in superficial answers after the feedback intervention that persisted through to the final test session.

In the positive feedback condition, a significant effect of test session was likewise apparent for superficial answers, \(F(2, 22) = 13.31, \text{MSE} = 0.009, p < .001\). Bonferroni-corrected pairwise comparisons revealed a rather different pattern of effects relative to the negative feedback condition, with no significant difference in the proportion of superficial answers between Test Sessions 1 and 2 (\(p = .77\)), a finding which is indicative of no immediate reduction in superficial answers post-feedback intervention. However, significant differences in the proportion of superficial answers did emerge between Test Sessions 2 and 3 (\(p = .007\)) and between Test Sessions 1 and 3 (\(p = .001\)), indicating a late-arising benefit in terms of avoidance of superficial answers.

**Structural errors**

Structural errors were a less prevalent form of incorrect responding compared with superficial answers, although the scores depicted in Figure 3 still suggest that the data relating to structural errors are sensitive enough to reveal some potentially informative differences across test sessions and feedback conditions. In particular, for the negative feedback condition it is apparent that the proportion of structural errors remained stable in Test Sessions 1 and 2 but declined in Test Session 3. In contrast, for the positive feedback condition the proportion of structural errors declined from Test Sessions 1 to 2, but increased again at Test Session 3.

A mixed between–within ANOVA was undertaken on the proportion of all responses that were coded as structural errors. This revealed a significant main effect of test session, \(F(2, 44) = 7.19, \text{MSE} = 0.006, p = .002, \eta_p^2 = .25\), but no main effect of feedback condition, \(F(1, 22) = 0.06, \text{MSE} = 0.009, p = .81, \eta_p^2 = .003\). The interaction between test session and feedback condition was, however, reliable, \(F(2, 44) = 4.20, \text{MSE} = 0.006, p = .022, \eta_p^2 = .16\).

Simple main effects analyses revealed no reliable difference in the proportion of structural errors between feedback conditions at Test Session 1, \(F(1, 64.18) = 0.95, \text{MSE} = 0.007, p = .33\), once again supporting the view that levels of task performance were equivalent for both groups at the start of the experiment. At Test Session 2, there was also no significant difference in the proportion of structural errors between feedback conditions \(F(1, 64.18) = 0.95, \text{MSE} = 0.007, p = .33\). However, at Test Session 3 a reliable difference between feedback conditions was apparent,
$F(1, 64.18) = 5.69, \text{ MSE} = 0.007, p = .02,$ confirming the existence of an increase in the proportion of structural errors for the positive feedback group relative to the negative feedback group at Test Session 3. We note that this increase in the proportion of structural errors in this group at Test Session 3 was not driven by a small subset of children. In fact, 8 out of 12 children showed between 10 and 30% structural errors on these reasoning trials, attesting to the consistency of these errors across individuals.

Next, we turn to the simple main effects analyses that examined the proportion of structural errors at the two feedback levels. A significant effect of test session was evident for the negative feedback condition, $F(2, 22) = 7.16, \text{ MSE} = 0.07, p = .004, \eta_p^2 = .39$, with Bonferroni-corrected pairwise comparisons indicating a lower proportion of structural errors between Test Sessions 1 and 3 ($p = .003$), with no other comparisons approaching significance. Thus, for the negative feedback condition there was evidently a small but continuous decline in the proportion of structural errors after the feedback intervention. In the positive feedback condition, a significant effect of test session was also apparent for the proportion of structural errors, $F(2, 22) = 4.09, \text{ MSE} = 0.006, p = .03, \eta_p^2 = .27$. However, Bonferroni-corrected pairwise comparisons revealed a different pattern of effects relative to the negative feedback condition, with no equivalent difference in the proportion of structural errors between Test Sessions 1 and 3. Indeed, the only comparison to approach significance was that between Test Sessions 1 and 2 ($p = .08$).

**Coding and analysis of explanations**

Many of the explanations provided by children during the feedback trials were written down by the experimenter. This record was incomplete owing to a degree of limited or piecemeal verbal reporting on the part of most children, combined with the procedural demands on the experimenter, who prioritized the efficient administration of trials so as to maintain each child's task engagement. Recording constraints had a particular impact in the negative feedback condition, where explanations tended to be
more elaborate than those in the positive feedback condition such that the written record was far harder to capture. The experimenter also omitted recording vague explanations and occasional ‘don’t know’ responses. With hindsight, a technological solution to recording explanations would have circumvented many of the factors limiting the quality of the explanation data. Whilst acknowledging these weaknesses, we nevertheless believe that the explanations that were successfully recorded do provide a further useful source of information about children’s reasoning strategies during the feedback manipulation. We therefore report a descriptive analysis of key aspects of these data below, whilst emphasizing the need for caution in interpreting the implications of these data for theory.

Explanations in the negative feedback condition
The negative feedback condition gave rise to 30 complete explanations that were coded using the scheme summarized in Table 1. This scheme categorized explanations into two basic types: those where the child showed an awareness that Mr Splodge had made an error that arose from him identifying a semantic association to some aspects of the presented analogy (i.e., Semantic Explanation) and those where the child provided an account of an assumed erroneous relational mapping that Mr Splodge had made when deriving his incorrect solution (i.e., Relational Explanation). Semantic explanations were further divided into two subcategories: those that contained a reference to the correct solution (i.e., Semantic Explanation + Correct Solution) and those that only contained a semantic explanation of Mr Splodge’s error in terms of an association mistake (i.e., Semantic Explanation Only). Relational explanations were likewise split into two subcategories: those that included an account or statement of the correct relational mapping (i.e., Correct Solution + Relational Explanation) and those that did not include an account or statement of the correct relational mapping (i.e., Relational Explanation Only).

<table>
<thead>
<tr>
<th>Explanation type</th>
<th>Example explanation for the incorrect analogy ‘Car:Road::Boat:Harbour’</th>
<th>Percentage of coded trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic explanation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Explanation + Correct Solution</td>
<td>It should be water but he thinks that all the boats are in the harbour</td>
<td>23</td>
</tr>
<tr>
<td>Semantic Explanation Only</td>
<td>Because boat and harbour go together</td>
<td>30</td>
</tr>
<tr>
<td>Relational explanation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relational Explanation + Correct Solution</td>
<td>A boat lives in a harbour, but a car lives in a garage . . . but a boat moves on water just like a car moves on roads</td>
<td>13</td>
</tr>
<tr>
<td>Relational Explanation Only</td>
<td>The boat is in the harbour but a car would be in a car park</td>
<td>33</td>
</tr>
</tbody>
</table>

This coding scheme successfully captured all complete explanations recorded in the negative feedback condition. Table 1 shows the percentage distribution of these codes across the pool of explanations. It is evident that semantic explanations (53%) and relational explanations (46%) were reasonably evenly distributed (we also note that all
children produced a fairly even mix of all explanation types). Over one-third of explanations contained either a statement of the correct solution or a relation-based account of the correct solution, suggesting that children were, at least some of the time, actively engaged in an attempt to identify what the correct response should be as part of their attempt to account for Mr Splodge’s erroneous answer. The relational explanations were typically sophisticated (as shown in the examples in Table 1), reflecting an apparent appreciation of the value of identifying the relational mappings associated with the analogies. We elaborate on possible implications of these data in the discussion section.

Explanations in the positive feedback condition
The positive feedback condition gave rise to 63 complete explanations. These were coded using the scheme depicted in Table 2, which dichotomized explanations in terms of whether they were attribute-based explanations (i.e., explanations that referred to the name of an object or an attribute in the presented pictures without any clear-cut attempt at relational reasoning), or relational explanations (i.e., explanations that referred to, or mapped out, the relational structure of the components of the presented analogy). This coding scheme captured all of the explanations recorded in the positive feedback condition. Table 2 shows the percentage distribution of these codes across this pool of explanations.

Table 2. Coding of children’s explanations for why Mr Splodge’s answers were correct in the positive feedback condition (see text for further details)

<table>
<thead>
<tr>
<th>Explanation type</th>
<th>Example explanation for the correct analogy</th>
<th>Percentage of coded trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute-based</td>
<td>A boat doesn’t have wheels so needs the sea</td>
<td>64</td>
</tr>
<tr>
<td>Relational</td>
<td>Cars drive on roads and boats float on the</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>water</td>
<td></td>
</tr>
</tbody>
</table>

It is evident from Table 2 that attribute-based explanations dominated children’s reasoning in this condition for the complete explanations that were successfully recorded. These attribute-based explanations appear to reflect a relatively superficial level of processing of analogies in these positive feedback trials. Nevertheless, slightly more than one-third of explanations were relational in nature, showing at least some attempt at a more sophisticated analysis of the presented problems. Two children’s responses were dominated by relational explanations and four children’s responses were dominated by attribute-based explanations. The remaining children showed a fairly even mix of both response types.

General Discussion
The present research set out to examine the role of positive and negative task feedback on children’s development of analogical reasoning skills using standard, picture-based proportional-analogy problems (e.g., Goswami, 2002). The study employed a small-scale microgenetic approach involving multiple testing sessions over a relatively short time-frame (cf. Siegler & Svetina, 2002). In this way, it was possible to control for factors
Feedback and analogical reasoning

associated with maturational changes in working memory (e.g., Halford et al., 1998) and the accretion of domain-specific knowledge (e.g., Goswami & Brown, 1989) that have been implicated as being important for developmental progression in analogical reasoning. The overriding prediction was that the provision of negative feedback in the form of opportunities for children to explain another person’s incorrect analogizing would promote enhanced strategy development in relation to analogical reasoning when compared with the provision of positive feedback based on opportunities to explain another’s correct analogizing. This prediction derived from previous research emphasizing the benefits for skill learning that can arise from encouraging reflection on erroneous responses and principle violations (e.g., Anderson et al., 1995; Freeman et al., 1999; Gopnik & Astington, 1988; Muldoon et al., 2007, 2003, 2009; Rittle-Johnson, 2006). To date, however, the valance of presented feedback appears not to have been examined systematically in studies examining the development of analogical reasoning strategies.

In terms of findings, our data for both feedback conditions revealed clear evidence for the development of an effective analogizing strategy over testing sessions in the form of a ‘relational shift’ from superficial to relational responding (cf. Rattermann & Gentner, 1998). This enhancement in analogical reasoning arose despite maturational and knowledge-accretion factors being controlled for through our use of a microgenetic methodology, where testing was conducted intensively over a single week. This general evidence for a relational shift appears to support previous evidence (Cheshire et al., 2005, 2007; Siegler & Svetina, 2002) that some form of task-related feedback is an effective method for enhancing the development of analogizing skills. We recognize, however, that our study did not include a practice-based control condition (i.e., where participants simply engaged in an analogical reasoning task during the intervention period that equated for the time spent engaging with the analogies in the feedback trials). As such, we cannot tell from the present study whether participants would have also shown practice-based learning effects over test sessions even without any feedback-based intervention. This remains an important issue that future studies could profitably address.

As for the prediction relating to the feedback manipulation, our findings suggest that the observed strategic switch from superficial to relational responding was indeed modulated by the feedback condition that the child had been placed in. In particular, whilst it was evident that significant improvements in correct analogical responding took place between Test Sessions 1 and 3 for both feedback interventions, it was the negative feedback condition that improved to the greatest extent over the time-course of the study. Not only did negative feedback participants show improved correct responding post-intervention compared to positive feedback participants, but they also showed a steady decline in incorrect strategies over time, whether these strategies were based around superficial responding driven by surface features of the tasks, or by attempts to transfer inappropriate relations that resulted in incorrect analogy completions (i.e., structural errors).

Whilst the response patterns for the negative feedback condition over the time-course of the experiment were clear-cut, the response patterns for the positive feedback condition showed a more complex developmental trajectory, though one that also led to performance improvements over test sessions. In particular, the positive feedback intervention appeared to induce a slower rate of developmental progression over time compared with the negative feedback intervention, with participants only gradually moving away from superficial responding towards more effective relational mapping. Particularly intriguing was the way in which structurally incorrect relational reasoning
increased significantly for the positive feedback group in Test Session 3 relative to the negative feedback group, perhaps indicating that participants who received the positive-feedback intervention had a more fragile grasp of the process of effective relational mapping (we note that 36% of complete explanations arising during the positive feedback intervention were coded as being relational in nature, indicating that at least some relational reasoning was taking place at this early point in the study). To put it another way, it seems that by Test Session 3 many participants in the positive feedback condition had gained an appreciation of the poverty of superficial responding and the importance of structure-based relational responding, but had not perfected their relational processing so as to achieve consistent levels of correct responding.

In sum, the differential rates of strategy acquisition associated with the positive and negative feedback sessions provide good grounds for the view that the valence of presented feedback is a critical catalyst for enhanced strategy development. What, though, of a theoretical account of the analogical reasoning differences arising in the negative and positive feedback conditions? Muldoon et al.’s (2003, 2007) proposals concerning the importance of children’s understanding of task-based errors and principle violations strikes us as an important contender as part of an account of our findings. The idea here is that through the experience of conceptualizing the existence of, and reasons for, an incorrect response the child may gain insights into the solution principles associated with successful completion of the task at hand (see also Freeman et al., 1999; Muldoon et al., 2009). Thus, the child who is confronted with negative feedback trials is prompted to reflect on why a response is erroneous in order to be able to provide an explanation of the error to the experimenter; this reflection may also encourage an active search for a correct solution, as was indeed seen in the negative feedback trials where over one-third of coded explanations involved a statement of the correct response.

One potential strength of such self-reflective reasoning when confronted with negative feedback is that it can alert the child to core aspects of the task that are relevant for success (e.g., relational mapping) as well as core aspects of the task that are superficial distractions that are unrelated to success (e.g., surface-level objects, attributes, and associations). The child may additionally be sensitized to the possibility that there are multiple relations that may play out in the source domain of a particular problem, but that many of these will be inappropriate for an effective relational mapping to the target domain. Armed with this deep understanding of the need to identify and map an appropriate relation, the child may not only avoid being lured by superficial aspects of the presented task, but may also be more cautious during their relation identification and mapping processes, with consequent benefits for task success.

In the positive feedback conditions it is possible that the child may often assume that simply because an answer has already been provided then there is no challenge to reason out an explanation as to the effectiveness of the presented solution, despite the experimenter’s request for such an explanation. However, our experimental observations indicated that children were generally keen to accept the task and to strive towards the generation of viable accounts of correct responses in this condition. More plausible, perhaps, is the suggestion that explanation-based reasoning for positive responses may be the subject of more superficial reasoning, perhaps involving simplistic rationalizations of correct responses. Again, the coded explanations for the positive feedback condition provide some tentative support for this proposal in that a majority of complete explanations (64%) focused on relatively superficial attribute-based aspects of presented analogies that failed to get to the heart of relevant relational solutions.
The examples of children’s explanations presented in Tables 1 and 2 reveal in more detail the important contrast arising during children’s self-reflective reasoning within the different feedback trials. The example explanations in Tables 1 and 2 concerned a relatively complex analogy: ‘Car is to Road, as Boat is to Water’ (note that ‘Water’ = correct response, positive feedback condition; ‘Harbour’ = incorrect response, negative feedback condition). As illustrated, a child in the positive feedback condition commented, ‘A boat doesn’t have wheels so needs the sea’. It would appear that this child was paying close attention to the surface (superficial) aspects of the problem in terms of available objects and attributes, which promoted a similarly superficial justification of the already correct response ‘Water’. This kind of rationalization process could arise because the individual who is subjected to positive feedback simply does not need to modify the presented response in any way (since it is correct), and therefore may not need to engage fully or deeply in the explanation process.

In contrast, a child in the negative feedback condition noted that, ‘A boat lives in a harbour, but a car lives in a garage . . . but a boat moves on water just like a car moves on roads’. This child appeared to be using structural (relational) information to work backwards from ‘Harbour’ to explicate a possible rationale for Mr Splodge’s incorrect solution and to modify the incorrect analogy to a correct relational mapping. A child who explains another person’s incorrect response in this way is clearly engaging in a deep level analytic processing that is evidence for good conceptual understanding of what analogical reasoning entails. Moreover, generating a response whilst modifying the answer would provide practice for subsequent task attempts, thereby facilitating more rapid improvement in the application of procedural skills (cf. Berardi-Coletta et al., 1995; Rittle-Johnson, 2006).

In conclusion, the present study indicates that providing conditions that enable children to reflect on and explain another person’s incorrect responses can enrich the development of strategic skills that relate to the effective identification and mapping of analogical relations. We contend that the acquisition of such analogical reasoning abilities is not simply due to knowledge accretion (e.g., Goswami & Brown, 1989; Leech et al., 2008). Nor can maturational changes in working memory and executive functioning easily account for performance benefits at analogizing that arise in a microgenetic training context. Instead, we suggest that it is opportunities for the child to undertake self-reflective, metacognitive processing that plays a core role in the enhanced development of analogizing abilities (cf. Cheshire et al., 2005, 2007). We finally note that our findings, whilst clearly needing replication and generalization across alternative analogical reasoning paradigms, have potentially important implications for education as a whole, and particularly for methods aimed at providing optimal learning benefits for children. We suggest that approaches to learning that encourage children to explain why an incorrect response has arisen may help promote the rapid and efficient development of core cognitive skills such as analogical reasoning.

Acknowledgements

We are grateful for the cooperation of the teachers and pupils at St Bernadette’s Catholic Primary School, Lancaster, UK. We also thank Andrew Bremner and two anonymous reviewers for their valuable comments on previous versions of this paper, and Charlie Lewis for insightful discussions concerning children’s understanding of errors during the acquisition of complex cognitive skills.
References


Received 10 November 2008; revised version received 27 May 2009