Examining how difficulty affects the goal specificity effect in interactive and non-interactive tasks

Nicola Marie Crane

Lancaster University

Examining how Problem Difficulty Affects the Goal Specificity Effect in Interactive and Non-Interactive Tasks

Nicola Marie Crane

2010

The work submitted in this report is my own and has not been submitted in substantially the same form towards the award of another degree or other qualificatory work by myself or any other person. I confirm that acknowledgement has been made to assistance given and that all major sources have been appropriately referenced.

Nicola Marie Crane

8th September 2010

Contents

Abstract

The type of goal an individual is given when completing a problem solving task can affect the strategy that they use. Specific goals lead individuals to only learn about specific cases, whereas non-specific goals facilitate hypothesis testing. Previous research indicates that although non-specific goals are usually beneficial, in certain circumstances, they can actually be detrimental to task performance, such as when the difficulty of a task means that the learner will not be able to work out the rule. Little research has examined the combined effects of interactivity and goal specificity, with the literature from each area giving contradictory accounts of task proficiency. The present study aimed to explore the combined effects of goal specificity, difficulty and interactivity on task performance. Forty participants completed problem solving tasks. Goal specificity and interactivity were both manipulated between-participants, with difficulty as a within-participants factor. There was no effect of difficulty, but there was an interaction between goal specificity and interactivity. Participants who were given a specific goal performed best if they could interact with the task, whereas for those who were given a non-specific goal, interactivity had little effect. For participants who could interact with the task, goal specificity had no effect, whereas for those who could not interact with the task, a non-specific goal led to superior performance. It was concluded that non-specific goals and interactive tasks both encourage hypothesis testing strategies, highlighting the importance of hypothesis testing for effective problem solving.

Examining how Problem Difficulty Affects the Goal Specificity Effect in Interactive and Non-Interactive Tasks

Problem solving is a fairly wide-ranging area of research within cognitive psychology, but can generally be described as the study of the way in which individuals acquire skills and use different strategies in order to complete a task or achieve a particular goal. Much research within this area focuses on the factors within a task which can affect an individual's performance, such as whether the problem can be solved by analogy (e.g. Gick & Holyoak, 1980) or if the problem is well or ill-defined (e.g. Hayes, 1978). However, recent research has shown that even the instructions given to an individual prior to beginning a problem solving task can affect their ability to complete the task successfully. When there is an underlying rule to be discovered to successfully complete a task, whether or not the individual is told to directly search for this rule or not can affect task performance; this phenomena is known as the goal specificity effect (Sweller & Levine, 1982).

Theoretical positions on the goal specificity effect

Before examining the goal specificity effect in depth, it is necessary to briefly discuss the methodology commonly used in this area. Complex Dynamic Control Systems (CDCSs) are a type of task widely used when investigating the goal specificity effect. This type of task uses a scenario such as a sugar production plant (e.g. Berry, 1991), and participants are asked to manipulate inputs, in this case the number of employees and rate of pay, in order to achieve a specific output, in this case, the productivity of the factory. Although there is variation between the exact scenarios used by different researchers, CDCSs usually share a similar structure involving manipulating one or many inputs in order to alter one or many outputs. Input and output are both usually numerical. Alternatives to CDCSs do exists, for example, Miller, Lehman, and Koedinger's (1999) Electric Field Hockey methodology, and Pretz and Zimmerman's (2009) Balance Scale task, but all of these tasks share a common feature of requiring participants to gain an understanding of the relationship between input and output.

Research into the goal specificity effect examines how the type of goal an individual is given - specific or non-specific - affects task performance. A specific goal involves the individual being told to repeatedly make predictions related to individual instances of inputs and outputs. They generally take the structure of predicting an output from a given input, or manipulating an input in order to achieve a target output. When an individual is given a non-specific goal, the task may take the same structure, but instead of being told to make predictions about a series of instances, the individual is instructed to try to work out the general relationship between input and output.

It has been consistently demonstrated that when individuals are given a non-specific goal, they tend to show better performance on tasks than individuals who are given a specific goal (Burns & Vollmeyer, 2002; Geddes & Stevenson, 1997; Miller et al., 1999; Osman, 2008; Sweller & Levine, 1982; Trumpower, Goldsmith, & Guynn, 2004; Vollmeyer, Burns, & Holyoak, 1996), with the former group having a higher chance of rule discovery (Vollmeyer et al., 1996), and showing a greater depth of schematic knowledge about the relationship between input and output (Trumpower et al., 2004). Early research into the goal specificity effect posited that this was because the different types of goals encourage individuals to use different strategies when solving a problem. Sweller and Levine (1982) argue that when an individual is given a specific goal, they are more likely to engage in means-end analysis, a strategy which involves examining the difference between the current state and the goal state and trying to work out a series of steps to reduce this distance, until the goal is reached. However, non-specific goals encourage a different strategy - hypothesis testing. This involves trying to work out the relationship between the variables in the task using a sort of trial and error approach, with decisions guided by feedback acquired from previous decisions. Direct evidence of the link between non-specific goal and hypothesis testing comes from Burns and Vollmeyer (2002), who used a verbal protocol analysis technique to examine strategies used by individual during problem solving tasks.

Although it is generally agreed that non-specific goals encourage hypothesis testing, opinions differ on why this approach proves beneficial. One of the first theoretical interpretations of this phenomena is known as cognitive load theory (Sweller, 1988). Proponents of this theory argue that when an individual is learning how to complete a task, they must both engage in cognitive processing of information and building schemas about the knowledge which they have acquired. These two activities draw from the same pool of cognitive resources, and so if a problem solving strategy which utilises a higher number of resources is used, less resources are available for the generation of schemas, and so learning is less efficient. Sweller (1988) argues that means-end analysis invokes a higher cognitive load than hypothesis testing, which is why the latter approach proves more effective than the former. Means-end analysis, according to Sweller (1988), requires the individual to store the current state, goal state, a series of subgoals, and the relationships between different states in working memory, whereas hypothesis testing is much less complex, simply requiring the individual to work out a single rule which can be generalised to different instances, and thus has a lower cognitive load.

This theory has been disputed by Miller et al. (1999), who demonstrated that specific goals are not always detrimental to performance. Miller et al. (1999) gave participants a task in which they had to learn about electrical charges and how they interact with each other. The task was in the form of a computer game which resembled an ice hockey pitch and participants were asked to place electrical charges on the pitch in order to guide a puck into the goal net. Participants were given one of three types of goals: a 'no goal' condition, similar to a non-specific goal in which participants could simply learn about the relationships between variables; a 'standard goal' condition, where participants were given the specific goal of getting the puck into the net; or a 'specific path' condition in which participants were given a specific goal of directing the puck into the net via a given trajectory. Miller et al. (1999) argue that, according to Sweller (1988), this third condition should lead to a higher cognitive load and therefore poorer performance. However, the findings of this research were that individuals in the 'no goal' condition performed best, followed by participants in the 'specific path' condition, and then those in the 'standard goal' condition. These results are interpreted by Miller et al. (1999) as demonstrating that it is not goal specificity which matters, but the appropriateness of the goal. Out of the two groups who were given specific goals, the group who were given a set trajectory performed better as this goal was more relevant to learning than the general goal of getting the puck into the net. The group who were given a non-specific goal were able to make their own appropriate goal. Burns and Vollmeyer (2002) note that this interpretation focuses on tasks which mainly involve the use of means-end analysis and although mentions the use of hypothesis testing, does not cite it as an important influence on task performance.

Taking a different approach, dual-space theory examines the goal specificity effect and considers the use of both means-end analysis and hypothesis testing. Vollmeyer et al. (1996)

argue that when an individual is trying to complete a problem solving task, they examine two different types of problem space; rule space and instance space. When an individual searches instance space, they compare the current state to the goal state, whereas when the individual searches rule space they formulate hypotheses which cannot be tested in rule space, but in instance space. Vollmeyer et al. (1996) argue that when an individual is given a specific goal, they will have a tendency to mainly or exclusively search instance space in order to reach the goal. However, when there is a non-specific goal, there is no goal present in instance space. Therefore, the individual searches rule space, using this to guide the way in which they search instance space. This approach leads to a deeper understanding of the relationship between input and output.

Dual-space theory seems to provide a feasible explanation for the findings of research into the goal specificity effect, and is currently a widely accepted theoretical position on the topic. However, it has quite a limited scope, and it seems important to examine how it can be integrated with more general theories of learning and skill acquisition. This gap is successfully bridged by Geddes and Stevenson (1997), who discuss how dual space theory relates to dual process theories of learning. It has been argued that knowledge can be acquired using two distinct systems: explicit learning and implicit learning. The difference between these two systems is quite straightforward; explicit knowledge is that which the individual is aware of and has made a conscious effort to acquire, whereas implicit knowledge is not consciously acquired, and although the individual may be able to use this knowledge, he or she will be unable to report it verbally. There is debate over whether different types of learning can be discussed in terms of the dichotomy of explicit and implicit learning, and some argue that it is more likely to be a continuum (Cleeremans, 1994), and that we cannot say for sure that knowledge has been acquired non-consciously (Shanks & St. John, 1994). However, Geddes and Stevenson (1997) argue that regardless of whether knowledge is truly explicit or implicit, evidence points to two different types of learning, and that despite differing terminology, theoretical and empirical evidence suggests that implicit knowledge is the same type that is gained through instance learning (Dienes & Fahey, 1995) and explicit knowledge is the same as knowledge acquired through hypothesis testing (Shanks & St. John, 1994). They argue that the literature on implicit and explicit knowledge supports this assumption, and so from here on, unless stated otherwise, any reference to explicit and implicit learning can be taken to mean the type of learning frequently described in this way, without making any further inference as to its nature.

An empirically demonstrated dissociation between explicit and implicit learning supports view that they are separate systems. Commonly cited evidence for this dissociation comes from Berry and Broadbent (1988), who examined the effects of manipulating task salience and task approach on performance. Participants were given a task in which they had to choose an appropriate input in order to achieve a given output from a computer program. Salience was manipulated by the output either having a direct relationship with input (salient), or having a relationship with the input on the previous trial (non-salient). Participants were either encouraged to use an implicit approach, and just had to achieve a specified output, or an explicit approach, being told to “crack the pattern”. It was found that when the task was salient, participants benefitted from an explicit approach, whereas on less salient tasks, an explicit approach resulted in a detriment in performance. After the experiment, participants were asked about the nature of the relationship between input and output, a test of explicit knowledge, and it was found that those who were given an implicit approach on the less salient task were less likely to be able to verbally describe the relationship, despite showing better performance, than those who took an explicit approach, hence showing a dissociation between explicit and implicit knowledge. Similar results have also been found by Hayes and Broadbent (1988) and Berry (1991).

These 'explicit' and 'implicit' instructions used by Berry and Broadbent (1988) closely resemble those used in research concerning goal specificity, and more recent research by Pretz and Zimmerman (2009) supports these findings. Although using different methodology, the factors examined by Pretz and Zimmerman (2009) were almost identical to those discussed by Berry and Broadbent (1988). Pretz and Zimmerman (2009) gave participants a task in which they were shown a scale with objects of equal weight balanced on equally spaced pegs, at varying distances from the centre. Participants had to predict whether the scale would tip left, right or stay balanced in the centre. The correct way to solve this problem is to multiply the number of weights on each peg by the distance from the centre, then add up the result for each peg to give a total for each side. The scale will tip to the side with the highest total. There were three difficulty levels; the most difficult trials could only be solved by using the correct calculation, the medium difficulty trials could also be solved by simply adding the number of weights on each side, and the simplest trials could be solved using the multiplicative way, additive way, or simply by a quick visual examination of the scale. Participants were either told to work out the rule that could be used to work out which way the scale would tip (i.e. a non-specific goal) or to simply make predictions about which way it would tip (a specific goal). Pretz and Zimmerman (2009) found an interaction between task difficulty and goal specificity, with individuals who were given a non-specific goal performing better on easier trials than those who were given a specific goal, but showing worse performance on more difficult tasks.

The key finding of this research was that the goal specificity effect is not universal to all situations, and it can be moderated by other factors, such as difficulty. Pretz and Zimmerman (2009) argue that this is because non-specific goals encourage the use of a hypothesis testing strategy. On simple tasks, this is useful, as hypothesis testing leads to a deeper understanding of the relationship between the relevant variables. However, when the task is more complex, individuals tend to spend too much time focusing on incorrect hypotheses, and specific goals, which encourage a more passive examination of the input-output relationship, are therefore beneficial.

This claim requires further examination, however, due to the methodology used by Pretz and Zimmerman (2009). The balance-scale methodology is radically different to the methodology commonly used in research exploring goal specificity. Although this is not necessarily a problem in itself, it differs from other research in that participants could potentially receive large amounts of misleading feedback, due to the incorrect strategies which could be used to guess the correct answer on a large number of trials. This differs greatly from more widely-used methodology, such as CDCSs, in which the output consists of nominal values, and so participants can see not only if they were accurate or not, but also how close they were to a correct answer if they fail to choose the correct values. Although participants may choose strategies which are not optimal for achieving the given goal, or working out the relationship between input and output, the intrinsic nature of CDCSs make them less ambiguous than the balance scale task. Further evidence for this assertion that the impact of incorrect hypotheses was methodologically based comes from Vollmeyer and Burns (1996, as cited in Burns & Vollmeyer, 2002), who found that participants who were told an incorrect hypothesis about the relationship between input and output, and given a non-specific goal still performed better than those who were told only accurate information but given a specific goal to pursue. However, this incorrect information was only given at the start of the experiment, and subsequent feedback would have disconfirmed this misleading information.

Another potential problem with Pretz and Zimmerman's (2009) research, which is acknowledged by the authors, is that unlike many previous studies concerning goal specificity, participants were not able to interact with the task. Pretz and Zimmerman (2009) concede that this may have led to participants who were given a specific goal undertaking more passive learning than in comparable studies, which may have been advantageous to their performance.

Interaction and observation

The effects of task interactivity had been given little consideration in the goal specificity literature, until recent research by Osman (2008). Research examining interactive tasks similar to those commonly used in goal specificity research has found that observation-based tasks seem to lead to the acquisition of explicit knowledge, whereas interactive tasks provide the individual with implicit knowledge (Berry, 1991). Additionally, interactive tasks have been found to be beneficial for skill acquisition and transfer to structurally different tasks (Sun, Merrill, & Peterson, 2001). Osman (2008) points out that these findings are at odds with those found in the goal specificity literature, which argues that it is explicit knowledge which is key for skill acquisition. Therefore, Osman (2008) examined the effects of both interactivity and goal specificity in order to assess which of these factors contributed most to skill acquisition. Interactivity was manipulated by creating two groups, one interactive and one observational. A CDCS task was used and the interactive group were able to choose input values and then clicked a button to reveal the corresponding outputs. The observational group were unable to select the input value themselves, and instead were yoked to a participants from the interactive group whose input choices were displayed to them on each trial. Other than this, the procedure was identical for the two groups. It was found that goal specificity had a significant effect on skill acquisition, and that there was little difference between individuals who could or could not interact with the task, and the main differences in task performance were the result of the specificity of the goal that they had been given, with non-specific goals proving beneficial. Osman (2008) argues that this discrepancy between these findings and those that would be predicted by research on interactivity can be explained by the instructions that individuals were given in earlier research, such as that of Berry (1991) and Lee (1995). Participants who were not able to interact with the task were expressly told not to use hypothesis testing strategies, and so this may have led to interactivity being confounded with goal specificity.

The findings of Osman (2008) and conclusions drawn contradict those of other research involving interactivity, with no difference found between groups on the basis of interactivity, and it seems important to consider potential explanations for these differences. Although Osman (2008) argues that goal specificity has frequently been confounded with interactivity, one would still expect interactivity to have some effect on individuals' task performance. The methodology employed by Osman (2008) required participants in both interactive and observational conditions to complete tests part way through the learning phase of the experiment which asked participants to state what they thought the relationship between input and output was. The inclusion of these tests may have diminished the differences between the two conditions, by encouraging participants to develop a more explicit understanding of relationship between input and output than would have been experienced otherwise, or encouraging a hypothesis-testing strategy. Therefore, the 'observational' and 'interactive' conditions used by Osman (2008) can be seen to be materially different to those used in other research which does not involve these interventions (e.g. Evans & Gibbons, 2009).

There is an abundance of evidence which does not risk confounding interactivity and goal specificity, which suggests that interactivity may play an important role in learning and skill acquisition. To give one such example, Evans and Gibbons (2009) presented individuals with either interactive or non-interactive stimuli which involved learning about the physics of how bicycle pumps works, which was followed by a transfer task requiring participants to use the knowledge acquired in the learning phase to answer questions about novel scenarios. Individuals who were given the interactive stimuli performed better on this task, and Evans and Gibbons (2009) concluded that this was because interactivity promotes deeper, more active learning. Interestingly, although those who could interact with the stimuli spent longer examining it during the learning phase of the experiment, no correlation was found between time spent in the learning phase and later performance. This also indicates that the advantage of interactivity was not simply due to more exposure to the information to be learned. Other research has also shown that interactive tasks are linked with hypothesis testing strategies (Lee, 1995; Taatgen & Wallach, 2002).

There is evidence that the effects of task interactivity can be affected by task difficulty, or the salience of the relationship between input and output. In research using a CDCS, Berry (1991) found that when the relationship between input and output was non-salient, interacting with the task led to far better performance than simply observing someone else interacting with it. However, when the relationship between input and output was salient, observation actually led to better task performance than interactivity. Task performance was compared with how well participants could explicitly state the relationship between input and output on a questionnaire, and although there was a positive correlation between task performance and questionnaire score for salient relationships, when the relationship between input and output was non-salient, task performance and questionnaire score were negatively correlated, providing further evidence for two distinct systems of knowledge acquisition.

In summary, it has been consistently demonstrated that giving an individual a non-specific goal often has a positive effect on task performance, due to the increase in the use of hypothesis testing strategies (Burns & Vollmeyer, 2002). This effect is not, however, universal, and when the task is sufficiently difficult for hypothesis testing to have a detrimental effect on performance, then it is a specific goal that will enhance task performance (Pretz & Zimmerman, 2009). Until recently, the combined effects of goal specificity and task interactivity had been given little attention. Although Osman (2008) claims that goal specificity has a greater effect than interactivity, issues with the methodology used means that the evidence supporting this claim may be flawed, and so requires further investigation. Interestingly, the literatures on goal specificity and interactivity disagree on what type of learning is necessary for skill acquisition, with goal specificity research emphasising the importance of explicit knowledge (Evans & Gibbons, 2009) and interactivity research arguing that it is implicit knowledge which is necessary for problem solving proficiency (Berry, 1991). Interactivity and goal specificity both interact with task salience (Berry, 1991; Berry & Broadbent, 1988; Pretz & Zimmerman, 2009). When a task is salient, non-specific goals have been shown to lead to better task performance, as have observational tasks. However, for non-salient tasks, a higher degree of interactivity and a specific goal both have been demonstrated to lead to better performance. Although the interactions between task salience and each of these factors have been studied in isolation, they have not yet been considered together in the same study.

The present study

The present study aims to replicate the findings of Pretz and Zimmerman (2009), that on simple tasks, non-specific goals lead to better performance than specific goals, but on more complex tasks, specific goals lead to improved performance. Additionally, the effects of interactivity will also be examined, with the intention of challenging Osman's (2008) claims that the effects of interactivity are neutralised when used in conjunction with goal specificity.

Much of the goal specificity literature claims that observational tasks lead to explicit knowledge, whereas interactive tasks lead to implicit knowledge and it appears that task performance is optimal when difficulty, goal specificity and the degree of interactivity in a task all engage congruent learning processes. When examined in conjunction with salient tasks, which have been shown to encourage explicit learning, observational tasks and non-specific goals, both of which have also been linked with explicit learning, lead to a greater degree of task proficiency than interactive tasks and specific goals. In contrast, on non-salient tasks, individuals seem to benefit from specific goals and interactivity, all of which have been linked with implicit learning. Interestingly, claims that non-specific goals are beneficial when a task is salient because they lead to hypothesis testing seem to be contradicted by the fact that interactive tasks are believed to also be linked to hypothesis testing, but yet are only advantageous on non-salient tasks, according to much of the related literature. This seems to indicate that it is congruency of learning processes and not simply encouraging hypothesis testing, which leads to superior task performance.

It is predicted that a three way interaction between goal specificity, interactivity and difficulty will be found. For the easiest problems, it is expected that when participants are able to interact with the task, specific goals will lead to better performance than non-specific goals. However, when they are not able to interact with the task, a non-specific goal will lead to better performance. Overall, on the simple problems, participants in the observational condition who are given a non-specific goal are expected to perform best during the testing phase.

On the difficult problems, the interaction between goal specificity is expected to operate in the same way as predicted for the easy problems. However, here, participants who have been given a specific goal and are able to interact with the experiment are expected to show the best testing phase performance.

Due to the aforementioned issues within the methodology used by Pretz and Zimmerman (2009), a different approach will be taken. Although there is a well-established range of methodologies already in use in goal specificity research, the novel combination of manipulating difficulty, goal specificity and interactivity makes it problematic to adapt these methodologies for use in the present study. Furthermore, all of the previous studies involve a single rule governing the relationship between input and output which participants must attempt to deduce across a number of trials. The present study uses multiple different rules, which change after each testing phase, to see if the effects in question can be generalised to different methodologies.

Method

Participants

The participants were 40 students from Lancaster University, selected using opportunity sampling.

Design

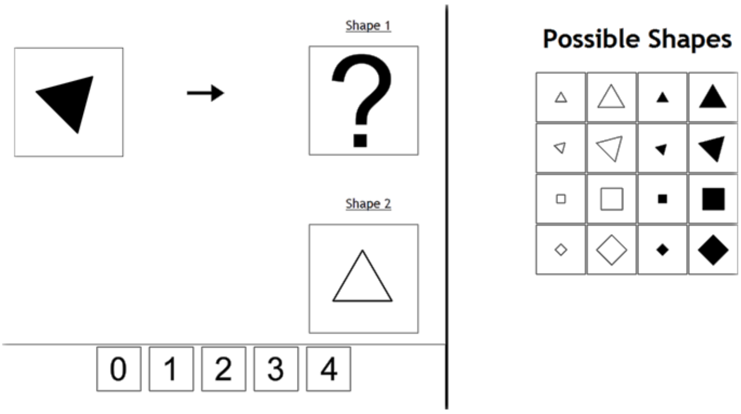
A 2x2x3 design was used. Difficulty was a within-participants factors and had three levels; easy, medium and hard. Goal specificity was a between-participants factor and had two levels; specific or non-specific goal. Interactivity was also a between-participants factor and had two levels; interactive or observational.

Materials and Procedure

The experiment was conducted using a computer program which had been written using PHP, HTML, Javascript, and CSS.

Participants were given instructions relevant to the condition which they had been assigned to. Instructions for the participants who were given a non-specific goal emphasised that they should look for the rule, whereas instructions for participants in the specific goal condition emphasised the importance of making predictions, but made no mention of looking for a rule. Pilot testing had shown that some participants had difficulty understanding the experiment, and so all participants were shown 3 example trials, although no hints were given as to how to solve them. Any participants who still did not understand the experiment were given an opportunity to ask the experimenter to clarify the instructions before the experiment began, however, no new information was given when explaining the procedure.

Figure 1. Display shown to participants



Participants were shown a computer screen, similar to the image shown in Figure 1. They were told that the experiment involved judging how many aspects of two shapes were the same. The aspects that they would be evaluating were size, number of edges, colour and rotation. Each of these were binary (i.e. so a shape could be big or small, have 3 edges or 4 edges, be black or white, and be rotated or unrotated). It was explained that all the possible shapes which could occur in the experiment were shown in the box labelled "Possible Shapes". Participants were told that the randomly chosen shape on the top left of the screen had undergone some kind of transformation. At the easiest level, the transformation was a change in just one of the aspects of the shape, but as the experiment progressed and got harder, the number of changes increased to two for the medium difficulty, and three for the most difficult trials. The result of this transformation was hidden behind the question mark labelled "Shape 1". Participants were told that they were required to guess what the shape hidden behind the question mark looked like and select the number of aspects that were the same as in the shape below, labelled "Shape 2". Participants in the interactive condition were able to choose the shape in the "Shape 2" box themselves, by selecting a shape from the "Possible Shapes" box on the right. Participants in the observational condition were yoked to a participant from the interactive condition, and so “Shape 2” was already filled in, based upon a previous participant’s selection.

As the act of being required to examine the “Possible Shapes” box to choose “Shape 2” may have led to an advantage for participants in the interactive condition, participants who were in the observational condition were asked to click on the shape in the “Possible Shapes” box which corresponded with “Shape 2” before they answered how many aspects of the two shapes matched.

After a selection was made, participants were shown a screen giving feedback on their answer. This was identical to the previous screen, except for that the number they had chosen was surrounded by a thicker border, and the correct answer had a green background. Below this, they were shown text which displayed the answer which they had chosen, and the correct answer.

Participants were able to study this screen for as long as they liked, and once they were ready to continue, they clicked a button marked “Next” to proceed to the next trial. In this trial, the transformation was the same, but the initial shape on the left was different. This learning phase repeated ten times until a message appeared on screen, informing participants that they had entered the testing phase. During this part of the experiment, which was the same for participants in the interactive and observational conditions, the task was almost the same as in the previous phase, except for that “Shape 2” was a randomly chosen shape, and participants just had to evaluate how many aspects of “Shape 1” and “Shape 2” matched. The transformation was the same in the testing phase as it had been in the preceding learning phase. After 5 trials within the testing phase, another box appeared on screen informing participants that they were entering a new learning phase and that the way in which the shapes were transformed had changed.

Each block of ten learning trials followed by 5 test trials was repeated 5 times for each of the three difficulties, giving a total of 225 trials for each participant. The first five blocks of trials were easy, next five blocks were medium difficulty, and the final five blocks were the most difficult. For each trial, the accuracy of response and response time was recorded.

Results

Scoring

Scores were calculated by examining the number of correct responses by each participant in each testing phase. Participants were tested 5 times in each testing phase, and so if they scored four or five out of five, they were deemed to have been successful. Participants were given a score for each difficulty, based on the number of testing phases in which they had successfully worked out the rule. Response times for the testing phases were also examined, measured from when the participants was first shown the problem until when they selected an answer.

Test performance

Table 1: Mean number of testing phases passed for each condition, with standard deviation in parentheses

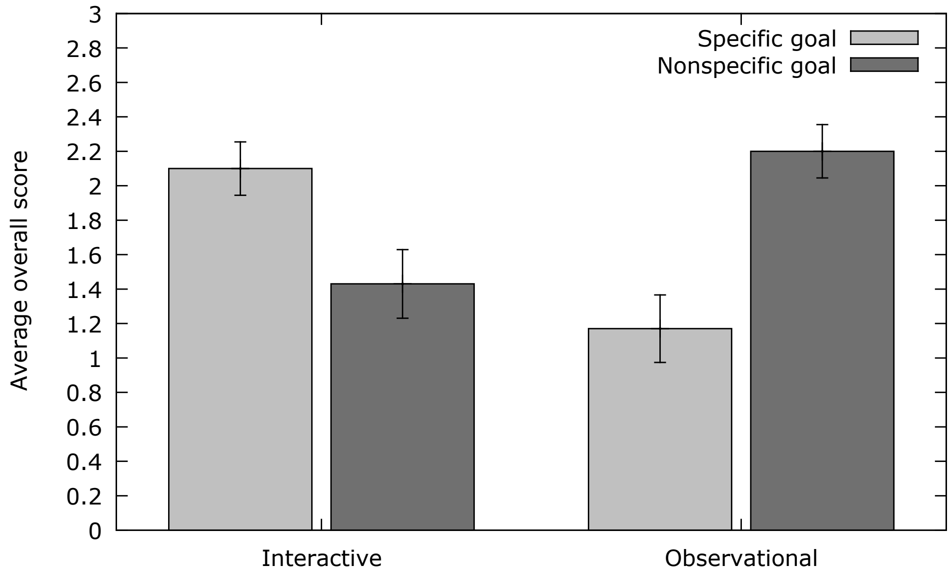
	Difficulty		
Condition	Easy	Medium	Difficult
Observation			
Specific goal	1(1.49)	1(1.33)	1.5(1.43)
Nonspecific goal	1.80(1.14)	2.30(1.34)	2.50(0.85)
Interactive			
Specific goal	1.20(1.23)	2.00(1.41)	3.10(0.99)
Nonspecific goal	1.00(1.33)	1.50(1.65)	1.80(1.48)

The data were analysed using a 2x2x3 mixed factorial ANOVA, with interactivity and goal specificity as between-participants factors, and difficulty as a within-participants factor. The results are shown in 1.

There was a main effect of difficulty, $F(2,72) = 13.12, p < 0.01, \eta_p^2 = 0.27$, with individuals performing best on the most difficult problems ($M = 2.23$), worse on the medium problems ($M = 1.70$) and least well on the easiest problems ($M = 1.25$). Bonferroni pairwise comparisons showed that there was a significant difference between the difficult and medium problems, $p = 0.02$, and between the difficult and easy problems, $p < 0.01$, but not between the easy and medium problems, $p = 0.10$. There were no main effects of interactivity, $F(1,36) = 0.06, p = 0.82, \eta_p^2 = 0.01$, or goal specificity, $F(1,36) = 0.27, p = 0.61, \eta_p^2 = 0.01$.

There was a significant interaction between interactivity and goal specificity, $F(1,36) = 5.69, p = 0.02, \eta_p^2 = 0.14$, but no significant interactions between difficulty and interactivity ($F(2,72) = 1.94, p = 0.15, \eta_p^2 = 0.05$) or difficulty and goal specificity ($F(2,72) = 1.18, p = 0.31, \eta_p^2 = 0.03$). The interaction between interactivity and goal specificity was further analysed, and it was found that goal specificity had little effect on score for individuals who could interact with the task, $F(1,18) = 1.75, p = 0.20$. However, for those in the observational condition, goal specificity was almost significant, with those who were given a non-specific goal performing better ($M = 2.2$) than those who were given a specific goal($M = 1.17$), $F(1,18) = 4.22, p = 0.055$. The simple main effects analysis of goal specificity showed that for those who were given a non-specific goal, there was little difference whether or not the task was interactive, $F(1,18) = 2.283, p = 0.15$. However, when a specific goal was given to participants, there was almost a significant difference, with individuals who could interact with the task performing better ($M = 2.10$) than those who could not ($M = 1.17$),

Figure 2. Mean score for each condition with standard error



$F(1, 18) = 3.48, p = 0.07$.

No significant interaction was found between interactivity, goal specificity and difficulty, $F(2, 72) = 1.481, p = 0.23$.

Response times

Response times for each condition can be seen in 2. A significant main effect of difficulty was found, $F(2, 72) = 15.62, p < 0.01, \eta_p^2 = 0.30$. Examining the Bonferroni pairwise comparisons indicated that individuals were significantly quicker on the medium difficulty

Table 2: Mean time in seconds taken to respond during testing phases passed for each condition, with standard deviation in parentheses

	Difficulty		
Condition	Easy	Medium	Difficult
Observation			
Specific goal	10.03(3.63)	8.10(5.18)	8.19(5.15)
Nonspecific goal	12.69(2.70)	9.64(4.06)	10.57(3.78)
Interactive			
Specific goal	18.61(7.13)	12.13(3.86)	12.67(3.41)
Nonspecific goal	11.46(2.23)	10.33(2.20)	8.61(2.88)

problems ($M = 10.05s$) than on the easy questions ($M = 13.20s$), $p = 0.001$. They were also significantly faster on the difficult problems ($M = 10.01s$) than on the easy problems, $p < 0.001$. However, there was little differences in response time between the medium and difficult problems $p = 1.00$. Interactivity also had a significant effect, $F(1, 36) = 5.38, p = 0.03, \eta_p^2 = 0.13$, with participants responding faster in the interactive condition ($M = 9.87s$) than in the observational condition ($M = 12.30s$). However, goal specificity did not have a significant effect on response times, $F(1, 36) = 1.05, p = 0.31, \eta_p^2 = 0.03$.

There were no significant interactions between difficulty and interactivity ($F(2, 72) = 1.71, p = 0.19, \eta_p^2 = 0.05$) or between difficulty and goal specificity ($F(2, 72) = 1.35, p = 0.27$), $\eta_p^2 = 0.04$. There was, however, a significant interaction between interactivity and goal specificity, $F(1, 36) = 9.71, p < 0.01, \eta_p^2 = 0.21$. Further analysis of this interaction showed that for those in the observational condition, goal specificity made little difference to their response times, $F(1, 18) = 1.85, p = 0.19$. However, for those in the interactive condition, a non-specific goal led to a faster response time ($M = 10.13s$) than a specific goal did ($M = 14.47s$), $F(1, 18) = 10.48, p < 0.01$. Additionally, when individuals were given a non-specific goal, there was little difference in response time whether they were in the interactive or observational condition, $F(1, 18) = 0.56, p = 0.46$. However, when they were given a specific goal, they were quicker to respond if they were in the observational condition than in the interactive condition, $F(1, 18) = 10.30, p < 0.01$.

There was a near significant three-way interaction between goal specificity, interactivity and difficulty, $F(2, 72) = 3.07, p = 0.053, \eta_p^2 = 0.08$.

Comparing test performance and response times

A Pearson's correlation coefficient was used to assess if there was a relationship between the average amount of time each participant took studying the problem during the learning phase, and their average score during the testing phase. There was no significant relationship between these two variables, $r(38) = 0.26, p = 0.10$.

Discussion

The results did not support the hypothesis that there would be an interaction between goal specificity and task difficulty, moderated by interactivity. Although pilot testing had indicated reliable differences between the three difficulty levels, the results did not reflect this, with performance improving as the experiment became more difficult. A feasible explanation for this is that the differences between the difficulty levels were simply due to practice effects. This lack of effect of difficulty perhaps also reflects the erroneous way in which difficulty or task complexity and task salience are frequently used interchangeably, both here and in other research. Although these terms often describe similar things, they are not identical. A task can be made more difficult or complex by increasing the number of inputs or outputs, but if the relationship between input and output is still apparent, the salience of the task may not be affected much or at all. To illustrate this point, a manipulation of salience could be considered to be one similar to that used by Berry and Broadbent (1988), with a salient task being one in which the output depended on the immediately related input, and a non-salient task being one in which output on a trial was related to input from the previous trial. There is no increase in the number of features of the task that participants must consider, but the relationship between input and output is still less apparent.

Despite this lack of expected effect of difficulty, however, there was an interaction between goal specificity and interactivity, with non-specific goals leading to better performance than specific goals for participants who could not interact with the task, but specific goals being advantageous when individuals were able to interact with the task. Although the interaction was statistically significant, with a reasonable effect size, simple main effect analyses did not indicate any of the factors to be statistically significant, possible due to the relatively low sample size. Therefore, the implications of this interaction will be interpreted by examining the components of the interaction which were closest to being significant, although the following discussion should be understood with the proviso that any conclusions drawn require further testing to see if the findings still stand when a larger sample or different methodology is used.

There was no relationship between the amount of time each participant spent examining each problem and their score, indicating that factors other than length of time of exposure to stimuli had affected the scores. Individuals in the interactive condition showed little difference in performance whether they were given a specific or non-specific goal. However, for those in the observational condition, a non-specific goal led to better performance than a specific goal. Also, for those who were given a non-specific goal, there was little difference in performance whether individuals were in the interactive or observational condition. However, when a specific goal was given, individuals who were able to interact with the experiment performed better than those in the observational condition. One explanation for these results is that a hypothesis testing strategy was key to task success. Previous research indicates that interactivity leads to hypothesis testing, whereas observation does not (Taatgen & Wallach, 2002). The lack of goal specificity effect for those in the interactive condition could be explained by the fact that non-specific goal are thought to encourage a hypothesis-testing strategy (Burns & Vollmeyer, 2002). If the interactivity has already encouraged this type of strategy, then goal specificity is redundant here. This also explains the lack of difference between interaction and observation in the non-specific goal condition; the non-specific goal has led to a hypothesis-testing strategy being used, and so the benefits conferred by interactivity have been made redundant. These findings are consistent with dual-space theory; a greater understanding of the underlying relationships within the task will be attained by searching rule space, testing different hypotheses.

Is learning on an interactive task implicit or explicit?

These findings are inconsistent with other research, which found an effect of goal specificity when individuals could interact with the task (e.g. Berry & Broadbent, 1988; Geddes & Stevenson, 1997; Miller et al., 1999; Trumpower et al., 2004), . These differences, however, may be due to the nature of the task that participants had to complete. The present study required participants to make estimates about the number of aspects of two shapes which matched, whereas methodology used in other research examining the goal specificity effect required participants to control a system and achieve a given output based on their input. When participants are required to control a system, interactivity may not encourage hypothesis testing, with the demands of a specific goal leading participants to use a means-end analysis approach and simply alter inputs until they achieve the required output. However, in the present study, the specific goal of making predictions about the number of similarities of two shapes still required a deeper knowledge of the underlying relationship, and so the hypothesis testing approach facilitated by interactivity led to similar success rates for the interactive and observational conditions. In other words, interactivity operates differently depending on whether the task is concerned with learning about *how* (procedural knowledge) or *what* (declarative knowledge). Another way of looking at this is to further examine the literature on salience in interactive tasks. Hayes and Broadbent (1988) argue that when an individual completes an interactive task, they engage in one of two types of learning; s-mode and u-mode. S-mode is described as conscious knowledge, and u-mode is described as unconscious knowledge, hence, we can equate these terms to explicit and implicit knowledge respectively. Hayes and Broadbent (1988) tested participants on tasks which required interaction and were identical except for that they differed in salience. It was found that when the relevant variables had a salient relationship, explicit learning occurred, whereas when the relationship was less salient, implicit learning took place. Hayes and Broadbent (1988) concluded that task salience determined what kind of learning took place in interactive tasks. These results are consistent with those of Berry and Broadbent (1988), who also used an interactive task and found that salience was key to the kind of knowledge acquired. These findings contradict the claims of those such as Dienes and Fahey (1995) that interactive tasks lead to implicit knowledge. However, these claims only concern CDCSs, which are usually designed to be inherently complex tasks, which, to some degree validates these arguments. Other findings such as those of Berry (1991), who examined both interactive and non-interactive tasks show that task salience determines what kind of learning takes place, and that different levels of interactivity may facilitate different types of learning, but not cause them. This explanation would also explain why there is inconsistency in how learning through interactivity has been described in different research. Learning during interactive tasks has been described as implicit by some authors (i.e. Berry & Broadbent, 1988), and as encouraging hypothesis testing by others (i.e. Taatgen & Wallach, 2002). These arguments are not compatible with the idea that hypothesis testing is an explicit strategy of learning (Geddes & Stevenson, 1997), and hence there is discrepancy between these different accounts, which could be explained by different task demands, and the present study constituting a salient task, in contrast to non-salient CDCSs.

The results also contradict those of Osman (2008), who found little difference between individuals who could or could not interact with the experiment, and no goal specificity by interactivity interaction. However, unlike Osman (2008), participants in the present study were not given a test of explicit knowledge, requiring them to categorically state what they believed the relationship between input and output to be, and this supports the earlier assertion that the inclusion of such tests modify the strategy that participants use, and possibly encourages hypothesis testing when this approach had not previously been used.

One problem with the experiment was that that due to the complex nature of the task, many participants initially had difficulty understanding what the task actually involved. Although pilot testing indicated that it was important to stress that participants' estimates of features the shapes had in common should be based on a comparison of 'Shape 1' and 'Shape 2' and not the initial shape and 'Shape 2', many participants reported only acquiring a good understanding of the task after they had completed a large number of trials. Additionally, participants who were given a non-specific goal found the task easier to understand, due to the emphasis on being told that the task involved trying to work out a rule. This confusion is reflected in the fact that as the task became more complex, task performance actually improved, as participants gained an understanding of the task. However, it is unlikely that this caused any real differences in performance between the non-specific and specific-goal conditions, due to the lack of difference in performance when completing an interactive task. Nevertheless, this may have had some effect and one could speculate that the specific goal groups may have performed better on a more straightforward task.

Another potential issue with these results is that the testing phase was the same for participants in both the interactive and observational conditions. For participants who had completed the observational condition, the testing phase was identical to what they had experienced in the learning phase, except for that ‘Shape 2’ was randomly generated, rather than yoked to the choice of a participant from the interactive condition. Less significantly, they also did not have to click on the corresponding shape before giving an answer. However, for the participants in the interactive condition, the difference was greater between the learning and testing phase, as previously they had been able to choose which shape went into the ‘Shape 2’ box themselves. This structural difference between learning and testing phase for participants in the interactive condition may have led to a decrement in their performance. Previous research has shown that that performance on a later task is much more likely to be successful if the task is structurally similar or identical to the task used in the learning phase of the experiment (Osman, 2008). The same potential confound may have also affected the way in which goal specificity had an effect on task performance, as it has been found that nonspecific goals facilitate greater transfer than specific goals (Burns & Vollmeyer, 2002) and so this may have led to an advantage to participants who were given a nonspecific goal.

As discussed earlier, it has been demonstrated that there is a dissociation between the implicit and explicit knowledge acquired during a problem solving task (Berry & Broadbent, 1988; Hayes & Broadbent, 1988; Berry, 1991). Due to the methodological design, and the previously discussed potentially confounding nature of including tests of explicit knowledge mid-way through the experiment, the testing phase of the experiment only examined what is considered to be implicit knowledge. Had it been possible to include a test of explicit knowledge, the results could have potentially been very different for this measure. This problem has also been acknowledged in previous research (i.e. Reber & Lewis, 1977; Berry, 1991).

Implications of these findings and suggestions for future research

The findings of the present study have important implications for the practical applications of different strategies in learning and teaching. As the popularity of electronic resources has increased over recent years, so has the potential for a higher degree of interactivity in learning. The findings presented here show that although when learning allows the individual to interact with the materials, the type of the goal they are given is not important, the reverse is true when learning is observational. In this case, it is important to encourage individuals to explore the underlying relationships within a system, rather than providing them with rigid, specific goals to achieve.

This claim, however, does not take into account the salience of relationship between the relevant variables and further research should explore if altering task salience affects what the most appropriate type of goal and learning environment is. Further research should also examine how goal specificity, interactivity and task salience affect learning over short and longer periods of time. The present research found that encouraging a hypothesis-testing approach led to improved task performance. However, hypothesis-testing has been found to be detrimental to task performance when the task is sufficiently difficult, or the individual is likely to pursue misleading hypotheses, so it would be useful to test this theory with both goal specificity and interactivity simultaneously.

Additionally, as proficiency on a task develops, it is the implicit knowledge acquired which is believed to be what distinguishes novices from experts (Berry, 1987). Therefore, it would be interesting to investigate whether it is more important that tasks encourage a hypothesis testing approach or simply facilitate the acquisition of implicit knowledge to have a beneficial effect on long term skill development. Additionally, it may be interesting to explore how tests giving during learning affect the type of knowledge gained and if such tests can have a beneficial effect on later task performance.

The present study indicates the important of hypothesis testing for successful problem solving, with the presence of a factor which leads to the use a hypothesis testing strategy being crucial for determining task success.

References

Berry, D. C. (1987). The problem of implicit knowledge. *Expert Systems*, *4*, 144–151.

Berry, D. C. (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology*, *43*, 881-906.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, *79*, 251–272.

Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *The Quarterly Journal of Experimental Psychology Section A*, *55A*, 241–261.

Cleeremans, A. (1994). Awareness and abstraction are graded dimensions. *Behavioral and Brain Sciences*, *17*, 402–403.

Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 848–862.

Evans, C., & Gibbons, N. J. (2009). The interactivity effect in multimedia learning. *Computers and Education*, *49*, 1147-1160.

Geddes, B. W., & Stevenson, R. J. (1997). Explicit learning of a dynamic system with a non-salient pattern. *The Quarterly Journal of Experimental Psychology Section A*, *50*, 742–765.

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

Hayes. (1978). *Cognitive psychology: Thinking and creating.* Homewood, IL: Dorsey Press.

Hayes, & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, *28*, 249276.

Lee, Y. (1995). Effects of learning contexts on implicit and explicit learning. *Memory and Cognition*, *23*, 723–734.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds.

Cognitive Science, *23*, 305–336.

Osman, M. (2008). Observation can be as effective as action in problem solving. *Cognitive Science*, *32*, 162–183.

Pretz, J. A., & Zimmerman, C. (2009). When the goal gets in the way: The interaction of goal specificity and task difficulty. *Thinking and Reasoning*, *15*, 405–430.

Reber, A. S., & Lewis, S. (1977). Implicit learning: an analysis of the form and structure of a body of tacit knowledge. *Cognition*, *5*, 333–361.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science*, *25*, 203-244.

Sweller, J. (1988). Cognitive load during problem solving:effects on learning. *Cognitive Science*, *12*, 257–285.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *8*, 463–474.

Taatgen, N. A., & Wallach, D. (2002). Whether skill acquisition is rule or instance based is determined by the structure of the task. *Cognitive Science Quarterly*, *2*, 1–42.

Trumpower, D. L., Goldsmith, T. E., & Guynn, M. J. (2004). Goal specificity and knowledge acquisition in statistics problem solving: Evidence for attentional focus. *Memory and Cognition*, *32*, 1379–1388.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, *20*, 75–100.