

Positive transfer and Negative transfer/Anti-Learning of Problem Solving Skills

Magda Osman

University College London

Department of Psychology

University College London

Gower Street

London WC1E 6BT

England

Phone: +4420 7679 7572

Fax: +4420 7436 4276

Email: m.osman@ucl.ac.uk

Abstract

In problem solving research insights into the relationship between monitoring and control in the transfer of complex skills remain impoverished. To address this, in four experiments participants solved two complex control tasks that were identical in structure but varied in presentation format. Participants learnt either to solve the second task, based on their original learning phase from the first task, or learnt to solve the second task, based on another participant's learning phase. Experiment 1 showed that, under conditions in which participants' learning phase was experienced twice, performance deteriorated in the second task. In contrast, when the learning phases in the first and second tasks differed, performance improved in the second task. Experiment 2 introduced instructional manipulations that induced the same response patterns as Experiment 1. In Experiment 3 further manipulations were introduced that biased the way participants evaluated the learning phase in the second task. In Experiment 4, judgments of self-efficacy were shown to track control performance. The implications of these findings for theories of complex skill acquisition are discussed.

Keywords

Induction, self-regulation, monitoring and control, observation versus action, skill learning

Positive transfer and Negative transfer/Anti-Learning of Problem Solving Skills

Central to skill development are two interrelated behaviors: Control and Monitoring. These behaviors generate and track processes involved in pursuing and fulfilling goals (e.g., Bandura & Locke, 2003; Burns & Vollmeyer, 2002; Lerch & Harter, 2001; Locke & Latham, 2002; Rossano, 2003; Schraw, 1998; Sweller, 1988; VanLehn, 1996). Monitoring refers to online awareness and self-evaluation of one's goal-directed actions. Control refers to the generation and selection of goal-directed actions. However, studies of skill learning in complex dynamic problem solving tasks have focused almost exclusively on understanding control behaviors, while neglecting monitoring behaviors. Without understanding how individuals monitor their behavior, little can be said about how evaluative processes are employed when transferring learnt skills to achieve unpracticed goals.

For example, Pilot A is training to fly the Boeing plane. In a flight simulation, they fly the plane on a two-hour night flight. The schedule includes the tutor replaying Pilot A their flight profile, to help them assess his performance. Pilot B experiences the same initial training routine as Pilot A, except that, after their flight, they are played Pilot A's flight profile, not their own. A final briefing session reviews both pilots' competence, and assesses how to transfer their training successfully to new flight patterns. Such training procedures are commonly used in educational (e.g., Pintrich & De Groot, 1990), clinical (e.g., Giesler, Josephs, & Swann, 1996), and military domains (e.g., Hill, Gordon, & Kim, 2004), to enable individuals to identify, correct, and improve their behaviors. In the example, both pilots share a precise goal that involves accurately and reliably controlling a complex dynamic control task (CDC-task: i.e., the aircraft). The critical difference is that Pilot A's training and assessment are based on self-generated behavior; whereas Pilot B's assessment is based on comparing self- and other-generated training behavior. The critical question that is raised by this example is: How will the two pilots' different learning

experiences impact on their later ability to transfer their knowledge to similar and different goals? In a series of analogous CDC-tasks, this study addresses a related and, as yet, unexplored question: How does monitoring affect the transfer of control behaviors in a complex skill learning task? More specifically, how does self-evaluation of one's goal-directed actions (task knowledge and performance) influence what is successfully transferred from one task to an analogous task? To answer these questions, this study introduces a theoretical framework, developed from Burns and Vollmeyer (2002) Dual-space hypothesis and Bandura's (1986; 1991) Social Cognitive theory, that relates monitoring to control processes. It proposes that people track and assess the effectiveness of their skill learning in complex dynamic learning environments. Negative evaluations will prevent relevant skilled knowledge from being applied to practiced and unpracticed goals, while positive assessments will enable the transfer of relevant skilled knowledge to different goals.

Monitoring: Self-Regulatory Mechanisms

Studies of skill acquisition show that monitoring is critical in the acquisition of complex behaviors, from athletic and musical performance to managerial decision making and stockbroking (Bandura, 1991; Bandura & Locke, 2003; Ericsson & Lehman, 1996; Karoly, 1993; Rossano, 2003; Stanovich, 2004). Why? Essentially, skilled behaviors are goal-directed pursuits, and monitoring thus serves a regulatory function, tracking and selecting out relevant information bearing on a desired outcome. One way in which this is demonstrated is by tracking ongoing performance through error detection (Bandura, 1991; Bandura & Locke, 2003; Karoly, 1993; Lehmann & Ericsson, 1997; Rossano, 2003). Error detection, or *reactive control*, is one of two self-regulatory mechanisms (*reactive control*, *proactive discrepancy*) that Bandura's (1986; 1991) Social Cognitive theory proposes people use. The reactive control mechanism is used to evaluate and then adjust peoples' behavior in order to reach a goal (Bandura & Locke, 2003; Karoly, 1993). The second type of regulatory

mechanism, known as *proactive discrepancy*, involves people tracking the current status of their performance and then incrementally setting more and more difficult challenges. Through this people can reach and even exceed their initial targets. In essence, the theory proposes that monitoring involves making online judgments about one's behavior and its relationship to a goal, and that this process is necessary in the acquisition and execution of skilled behaviors. This study examines whether it also follows that the self-regulatory mechanisms proposed by Social Cognitive theory will influence the transference of control skills to different tasks.

Regulatory Mechanisms through Self-Observation

In the example, the training regime that the pilots follow involves error correction and detection through observation. One pilot observes another's flight simulation behavior; the other observes his own behavior. The latter is known as the self-observation technique, and is used extensively in educational (e.g., Covington, 2000; Pintrich & De Groot, 1990) and clinical domains (e.g., Bailey & Sowder, 1970; Dowrick, 1983; Giesler et al., 1996), to identify and improve on maladaptive behaviors. For example, developmental studies (Fireman & Kose, 1991, 2002; Fireman, Kose, & Solomon, 2003) report that children improve their problem solving ability by examining videotaped presentations of their previous attempts. In Fireman et al.'s (2003) study, children completed the Tower of Hanoi (TOH) task and were then shown their own moves, or another child's previous inefficient moves, or another child's correct completion of the task. Presented with a new TOH task, the children that had observed their own previous behaviors performed best.

Similarly, the self-observation technique has been found to improve a range of skills (e.g., meta-perception; motor learning, dart throwing) in adults (e.g., Albright & Malloy, 1999; Carroll & Bandura, 1982; Fireman & Kose, 1991, 2002; Knoblich & Flach, 2001). These studies indicate that the technique encourages people to use monitoring behaviors of the kind described by Bandura, in which detection of inefficient behaviors can

be corrected and efficient behaviors exploited. The limitation of studies that have used the technique thus far is that they have focused on people's detection of, and improvement to their behaviors whilst observing themselves in action, which provides no insight into how people monitor and correct internally represented behaviors, by which is meant decision-making, reasoning, and hypothesis testing behaviors. This study examines monitoring and its effects on transfer of skilled behaviors, by re-exposing problem solvers to products of their own strategic thinking, rather than to visual (i.e., video) presentation of themselves performing a task. It is thus possible to empirically control the information that their self-regulatory mechanisms operate on, and examine the impact on the transfer of control behaviors.

Complex Dynamic Control Tasks (CDC-Tasks)

CDC-tasks, like the one referred to in the example, have been a popular task environment (Brehmer, 1992; Cañas, Quesada, Antoli, & Fajardo, 2003; Funke, 2001; Kerstholt, 1996; Lipshitz, Klein, Orasanu, & Salas, 2001) for examining the acquisition and transfer of control skills in dynamic goal directed environments. The simulated environments used (e.g., air-traffic control, subway systems) often relate closely to genuine control systems, and thus provide strong ecological validity (Buchner & Funke, 1993). Typically, a CDC-task (e.g., water purification system) includes several inputs (salt, carbon, lime) that are connected via a complex structure or rule to several outputs (chlorine concentration, temperature, oxygenation) (Figure 1). Common to studies using CDC-tasks is the inclusion of a learning phase, in which learners familiarize themselves with the system. Here learners interact with a CDC-task by changing the inputs. They are able to learn about the input-output relations by using the continuous feedback received on the output variables that change as a result of the changes to the inputs. In the test phase, the participants operate the system and demonstrate their ability to control it, by achieving a specific goal.

As a problem solving skill, controlling a dynamic system necessarily involves reaching and maintaining goals. Thus, one approach to understanding control behaviors in CDC-tasks compares different types of goal instructions during learning (e.g., Burns & Vollmeyer, 2002; Osman, in press; Sweller, 1988; Vollmeyer et al., 1996). For instance, instructions like “explore the system,” a non-specific goal, are contrasted with “learn about the system while trying to reach and maintain specific output values,” a specific goal. In the test phase, specific goal learners perform more poorly than non-specific goal learners (e.g., Burns & Vollmeyer, 2002; Geddes & Stevenson, 1997; Sweller & Levine, 1982; Trumpower, Goldsmith, & Guynn, 2004; Vollmeyer et al., 1996).

Control Behaviors in CDC-Tasks

Burns and Vollmeyer's (2002) extension of Dual-Space theory (Klahr & Dunbar, 1988; Simon & Lea, 1974) has been used to explain the goal-specificity effect and other problem solving behaviors in CDC-tasks. Burns and Vollmeyer propose that skilled control behaviors are acquired by using the principles underlying scientific discovery. The CDC-task is described as analogous to a hypothesis testing environment with two spaces: the rule space, which determines the relevant relationship between inputs and outputs; and the instance space, which includes examples of the rule being applied. Successful control skills develop because exploration encourages both hypothesis generation and testing, whereas under goal-specific conditions learners simply generate instances that fulfill goals, with no opportunity to formulate hypotheses. Crucially, Burns and Vollmeyer leave open the possibility that monitoring has a mediating role in the acquisition of control behaviors. They posit that self-evaluative processes are recruited during hypothesis testing, to track the hypotheses being tested, and to update them accurately from the results of these tests.

In contrast, the dissociationist approach (Berry, 1991; Berry & Broadbent, 1984, 1987, 1988; Dienes & Berry, 1997; Lee, 1995; Stanley, Mathews, Buss, & Kotler-Cope, 1989) proposes that the knowledge acquired in CDC-tasks is procedural, and represents

“knowing how” to perform actions tied to specific goals. This is independent of declarative knowledge, which is “knowing that” of particular facts about the underlying actions, and structural knowledge of the environment being operated. These forms of knowledge are not only independent of each other: It is also claimed that functionally separate cognitive mechanisms support them (see Osman, 2004, for a review). One method used to demonstrate this involves training people on a procedural task by observing another perform it first: The observers are described as generating declarative knowledge because they are explicitly monitoring the action of another (e.g., Kelly & Burton, 2001; Kelly, Burton, Riedel, & Lynch, 2003). Berry (1991) and Lee (1995) used this method to compare the effects of procedural-based and observation-based learning. They showed that, when participants later came to problem solve, the observers’ ability to perform the procedural task was poorer than that of procedural-based learners. They claimed monitoring has a detrimental affect on control behaviors in CDC-tasks, and that acquisition of control behaviors is dependent on active interaction with the CDC-task.

Present Study

Social Cognitive theory and Dual-Space theory assume that that monitoring behaviors are necessary in order to track and modulate control performance. Therefore, monitoring should have a mediating affect on the transfer of control skills to new goals. In contrast, dissociationists claim that procedural—not declarative—knowledge is necessary in the acquisition of control behaviors. Thus monitoring should have a detrimental affect on the transferability of control behaviors in CDC-tasks. To understand how monitoring influences the transfer of control skills to analogous CDC-tasks, the present study asks (1) Does control performance improve if monitoring is based on one’s prior self-generated behavior, rather than the behavior of another individual? (2) Can people discriminate between their own self-generated behavior and that of another individual? (3) Is control performance improved if monitoring of self-generated or other-generated behaviors occurs

online rather than via observation? (4) Can indices of monitoring behavior accurately predict the transferability of control behaviors in a complex skill learning task?

General Method

In the following four experiments, participants performed two problem solving tasks, each consisting of a learning phase and a test phase. All participants solved the first problem in the same way by completing the learning and test phase, and in each experiment the critical manipulation concerned the contents of the learning phase in the second problem (i.e. Self conditions, Other conditions). In ‘self-labeled conditions, participants in the second problem were exposed to their own learning phase from the first problem. In ‘other’-labeled conditions, participants were yoked to a participant in the corresponding ‘self’ condition, and in the second problem were exposed to that individual’s learning phase. In addition, the presentation format of the learning phase in the second problem was varied: i.e., it was either action-based (Experiments 1, 2, 3) or observation-based (Experiments 1, 4), and the cover story was manipulated so that either the second problem was different to the first (Experiment 1, 4) or identical (Experiment 2, 3). A further manipulation concerned the instructions presented prior to the presentation of the second problem (Experiment 2, 3).

Experiment 1

Experiment 1 included four conditions. In each, participants solved two CDC-task problems. All participants solved the first problem in the same way, by generating their own learning experience in the learning phase. However, in the second problem, half the participants re-experienced their original learning phase from the first problem, through either observation-based (Observe-self) or action-based (Act-on-self) learning. The remainder experienced a different learning phase from their own, through either observation-based (Observe-other) or action-based (Act-on-other) learning.

Dissociationists (e.g., Berry, 1991; Berry & Broadbent, 1988; Lee, 1995; Sun et al., 2001) propose that only procedural processes are necessary in the acquisition and transfer of knowledge in CDC-tasks. Therefore, in Experiment 1, transfer of control performance should be facilitated if the learning phase of the first and second problems is procedural-based (Act-on-self, Act-on-other), and performance should increase across problems. Additionally, decrements in control performance should be found in conditions in which the learning formats of the first and second problems are different (Observe-self, Observe-other), because declarative knowledge is brought to bear during observation-based learning and invokes monitoring behaviors, which interfere with procedural processes (Berry, 1991; Berry & Broadbent, 1987), and thus prevent transfer of control skills.

If, however, consistent with Social Cognitive theory and Dual-Space theory, monitoring mediates control behaviors, transfer of control behaviors should be facilitated, whatever the presentation format of the learning phases. If monitoring is involved, then, during the learning phase, people will be sensitive to the kind of information presented (i.e., the source of the second learning phase), not its presentation format (observation-based, action-based). In this case, participants will demonstrate knowledge of the difference in the source of the second learning phase.

Method

Seventy-two graduate and undergraduate students from University College London volunteered to participate in the experiment and were paid £6. Participants were aged between 19 and 35, and 48 were women. Participants were randomly allocated to one of four conditions (observe-self, act-on-self, observe-other, act-on-other), with eighteen in each. Participants were tested individually.

Design and Materials

Experiment 1 was a mixed design that included two between subject variables comparing re-exposure to self-generated learning instances and exposure to other-generated learning instances (i.e., Self vs. Other), and the effects of learning format on transfer of control performance (Observation, Action). Two within subject variables examined transfer of skill across two CDC-tasks, one measuring control performance in two tests (Tests 1-2), the other measuring structural knowledge in four tests (Structure Tests 1-4). The order of presentation of the two CDC-tasks was randomized for each participant. The critical manipulation was the contents of the second learning phase. In the first problem, all participants generated their own learning experiences. In the second, half the participants re-experienced their original learning phase (observe-self, act-on-self), and the other half experienced the learning phase generated by another participant (observe-other, act-on-other). Full details are provided in the procedure section.

CDC-tasks

The design and underlying structure of the two CDC-tasks used (Water-Tank control system, Ghost Hunting control system) were based on the Water Tank system (see Figure 1). The only differences between the two problems were the visual layout of each system on screen, and the cover story (see Appendix). In the Water-Tank control system, participants were told that, as workers of the plant, their job was to inspect the water quality of the system. The system was operated by varying the different levels of salt, carbon, and lime (inputs), which then changed the three water quality indicators: oxygenation, temperature, and chlorine concentration (outputs). Participants controlling the system had to reach specific values of the water quality indicators. In the Ghost Hunting control system, participants were told that they were newly recruited ghost hunters, and had returned from a field experiment. Their job was to examine three pieces of equipment used in the field: GGH Meter, Anemometer, Trifield Meter (inputs), and the

readouts of the three phenomena that these detect: Electro Magnetic Waves, Radio Waves, Air Pressure (outputs). Controlling the system involved modifying the levels of the readouts of the phenomena, by manipulating the dials on each machine.

Procedure

First problem: Learning phase. In the learning phase of the first problem, participants were presented with a computer display with three input and three output variables. Each trial consisted of participants interacting with the system by changing any input by any value they chose, using a slider corresponding to each.¹ Each slider had a scale from -100 to 100 units. When participants were satisfied with their changes to the inputs, they clicked a button labeled “output readings,” which revealed the values of all three outputs. When they were ready to start the next trial, they clicked a button “next trial,” which hid the output values from view. On the next trial, the newly changed inputs affected the output values from the previous trial: thus, the effects on the outputs were cumulative from one trial to the next.² After the first block of 6 trials, participants were presented with Structure Test 1. A diagram of the system was shown on screen, and participants were asked to indicate which input was connected to which output. After this, participants began the next set of 6 trials.³ On completion of the second block, Structure Test 2 was presented. The inputs that changed on each trial, the values they were changed by, and the corresponding effects on the outputs comprised the trial history of each participant.

Test Phase of both problems (Test 1 and Test 2). After the learning phase, participants’ ability to control the system was tested (Tests 1-2). In this phase, all participants had to change the input values to achieve and maintain set output values. In the first and second problems, for the course of 6 trials, the criterion values participants had to reach in Test 1 were the same, and only the labels of the outputs were different: Output 1 (Water Tank = Oxygenation, Ghost Hunt = Radio Waves) = 50; Output 2 (Water Tank = Chlorine Concentration, Ghost Hunt = Electro Magnetic Waves) = 700; Output 3 (Water Tank =

Temperature, Ghost Hunt = Air Pressure) = 900, for the course of 6 trials. On completing Test 1, participants were presented with Structure Test 3 and the second test. In Test 2, the criterion values they had to achieve were Output 1 = 250; Output 2 = 350; Output 3 = 1100, for the course of 6 trials. Participants were then presented with Structure Test 4.

Second problem: Observation-based learning phase. In the second problem, the learning phase was observation-based for half the participants. Instead of changing the inputs, on each trial participants pressed a button “reveal inputs,” then observed the sliders of the inputs changing automatically according to pre-specified values. Then they pressed a button “reveal outputs,” which displayed the corresponding effects on the output values. After studying them, participants clicked a button “ready for next trial,” which cleared the input and output values ready for the next trial. As in the first problem, after Trials 6 and 12 participants were presented with a Structure Test. The Observe-self condition watched their own trial history, which they had generated from the first problem; the Observe-other condition observed the trial history of a participant from the Observe-self condition. For example, in the first learning phase, Participant A, from the Observe-self condition, changed Input 1 on Trial 1 by 50 units. In the second learning phase, the Observe-self condition now watches Input 1 change by 50 units on Trial 1. In the first learning phase, Participant B, from the Observe-other condition, changed Input 2 on Trial 1 by 70 units. The Observe-other condition is randomly allocated the trial history of Participant A, and so in the second learning phase, they simply observe Input 1 change on Trial 1 by 50 units.⁴

Second problem: Action-based learning phase. For the remaining participants, the second learning phase was action-based. At the start of the learning phase, the Act-on-self condition was presented with a trial history sheet listing the inputs changed and the values they were changed by, for each of 12 trials. The Act-on-self condition was instructed to interact with the system on each trial by making the changes listed on the record sheet.

They were thus mimicking the learning behaviors from the first learning phase. The procedure was the same for the Act-on-other condition, except that they were randomly allocated the trial history of a participant from the Act-on-self condition.

Post-test question. After completing the experiment, participants were informed of the manipulation to the second learning phase and were asked which of the two (i.e., Self or Other) trial histories they were exposed to. This question served as an index of self-insight.

Scoring

Structure scores. The method used to score performance on Structure Tests 1-4 computed the proportion of input-output links correctly identified for each test. A correction for guessing was incorporated, based on Vollmeyer et al.'s (1996) procedure, which was $\frac{\text{correct responses (i.e., the number of correct links included, and incorrect links avoided)} - \text{incorrect responses (i.e., the number of incorrect links included, and correct links avoided)}}{N}$ (the total number of links that can be made). The maximum value for each structure score was 1. This scoring scheme was applied to score performance on all structure tests in Experiments 1-4. Successful performance is indicated by an increase in structure scores.

Tests 1 and 2. The procedure used in Experiments 1-4 was based on Burns and Vollmeyer's scoring system. Control performance was measured as error scores in Tests 1-2. Error scores were based on calculating the difference between each target's output value (i.e., the criterion according to the test) and the actual output value produced by the participant for each trial of the transfer test. A log transformation (base 10) was applied to the error scores of each individual participant for each trial, to minimize the skewedness of the distribution of scores. All analyses of error scores for Test 1 were based on participants' mean error, averaged over all 6 trials, across all three output variables. Error scores for Test 2 were calculated in the same way. Success in control performance on

transfer tasks is indexed by the difference between the achieved and target output values, thus lower error scores indicate better performance.

Results

This section first analyzes initial differences between conditions, then control performance, then structural knowledge in each CDC-task. Correlation analyses examine the potential association between control performance and structural knowledge. Finally, responses to the post-test question are analyzed. In all analyses reported in this article, a significance criterion of $\alpha = .05$ was used. The results of non-significant findings are not reported.

Performance measures in the 1st problem. The control performance of both conditions in the first problem was initially compared, to rule out any possibility of initial group differences influencing any later main effects. A 2x4 ANOVA with test (Test 1, Test 2) as a within subject variable, and condition (Observe-self, Observe-other, Act-on-self, Act-on-other) as a between subject variable, was conducted on mean error scores. The analysis revealed a significant main effect of test: $F(1, 68) = 14.82$, $MSE = 0.37$, $p < 0.0005$, $\eta^2 = 0.18$. This indicates that the tests may have differed in difficulty, which is consistent with the findings reported by Burn and Vollmeyer. As with the analyses of control performance, analyses of structure tests score averaged across the four tests revealed no significant differences between conditions.

Comparison of test scores in the 1st and 2nd Problems. The mean error score of all four conditions presented in Figure 2 suggests that, for the Observe-self and Act-on-self conditions, control error scores increased (indicating worse performance) in both tests in the second problem. The reverse trend is indicated for the Observe-other and Act-on-other conditions. To analyze this, a 2x2x2x2 ANOVA was carried out using test (Test 1, Test 2) and problem (1st Problem, 2nd Problem) as within subject variables, and condition (self, other) and learning format of the second problem (observation, action) as the

between subject variables. The analysis showed a significant main effect of test: $F(1, 68) = 29.52$, $MSE = 0.82$, $p < 0.005$, $\eta^2 = 0.30$. There was also a significant main effect of condition, $F(1, 68) = 11.59$, $MSE = 0.87$, $p < 0.001$, $\eta^2 = 0.15$, and a significant Condition x Problem interaction, $F(1, 68) = 53.27$, $MSE = 2.46$, $p < 0.0005$, $\eta^2 = 0.44$. Given that there was no Condition x Problem x Test interaction, the scores were collapsed across tests.

The significant increase in control error scores (worse performance) between the first and second problem of the Observe-self and Act-on-self conditions (Figure 2) was confirmed by further comparisons: $t(35) = -4.52$, $p < 0.001$, $d = -1.53$ and $t(35) = -6.25$, $p < 0.0005$, $d = -2.11$, respectively. The significant decrease in control error scores (improved performance) between the first and second problem of the Observe-other and Act-on-other conditions was also confirmed by a planned comparison: $t(35) = 3.75$, $p < 0.001$, $d = 1.27$ and $t(35) = 3.65$, $p < 0.001$, $d = 1.23$, respectively. Thus, the evidence suggests that the difference in the patterns of transfer of control performance was the result of the content of the second learning phase, not its presentation format.

Comparison of structure test scores in 1st and 2nd Problems. For each participant, the scores from Structure Tests 1-4 were averaged across the first problem, and again for the second problem. The averages of these scores from each of the four conditions are presented in Figure 3, which indicates that, for the Observe-self and Act-on-self conditions, structure scores decreased (worse performance) in the second problem. The reverse trend is indicated for the Observe-other and Act-on-other conditions. This was analyzed using a 2x2x2 ANOVA over averaged structure test scores, using problem (1st Problem, 2nd Problem) as a within subject variable, and condition (self, other) and format (observation, action) as the between subject variables. There was a significant main effect of condition, $F(1, 68) = 4.37$, $MSE = 32.23$, $p < 0.05$, $\eta^2 = 0.06$, and a significant Condition x Problem interaction, $F(1, 68) = 35.95$, $MSE = 129.34$, $p < 0.001$, $\eta^2 = 0.35$. The significant decrease

in structure scores (worse performance) between the first and second problem of the Observe-self and Act-on-self conditions (Figure 3) was confirmed by a further comparison: $t(17) = 5.16, p < 0.005, d = 2.50$, and $t(17) = 2.32, p < 0.05, d = 1.13$, respectively. In addition, the significant increase in structure scores (improved performance), in the second problem, of the Observe-other and Act-on-other conditions, was confirmed: $t(17) = -3.38, p < 0.005, d = -1.64$ and $t(17) = -2.40, p < 0.05, d = -0.57$, respectively. Thus, the evidence suggests a negative transfer of declarative knowledge in the Observe-self and Act-on-self conditions, and a positive transfer in the Observe-other and Act-on-other conditions.

Correlation between control performance and structural knowledge. A correlation analysis was carried out on control error scores (averaged across Tests 1-2), and structure test scores (averaged across Structure Tests 1-4) from the first and second problems. A significant negative relationship was found between structure test scores and test error scores in the first problem, $r(72) = -0.29, p < 0.05$, and in the second problem, $r(72) = -0.38, p < 0.001$. These findings strongly indicate that, for both types of learning phase (observation-based, procedural-based), there is a relationship between control performance and structural knowledge.

Post-test question. Eighty-three percent of participants in the Observe-self condition and 67% in the Act-on-self condition reported accurately which of the two conditions they were in. Seventy-eight percent of participants in the Observe-other condition and 78% in the Act-on-other condition answered correctly. Pearson's chi-squared analysis revealed no significant difference in correct and incorrect response by condition.

Discussion

The evidence from Experiment 1 is summarized, as follows: First, successful transfer of control performance was found to be independent of the format of the learning phases of each problem. Second, structural knowledge and control performance were associated in both problems. Third, participants' accurate self-insight enabled them to

correctly identify the source of the second learning phase. Fourth, there was positive transfer of structural knowledge and control performance in Observe-other and Act-on-other conditions, and negative transfer in Observe-self and Act-on-self conditions.

Taken together, the evidence indicates that procedural knowledge and declarative knowledge in CDC-tasks are associated. Although inconsistent with dissociationists' claims, the findings are consistent with Social Cognitive theory and Dual-Space theory, and indicate that monitoring mediates the transfer of control behaviors. For both theories, monitoring serves a regulatory function, because it tracks and selects out relevant information bearing on a desired outcome. This is through evaluation of either skilled behaviors (Bandura, 1986; Carroll & Bandura, 1987; Cervone et al., 1991), or of the hypothesis testing strategies developed during learning (Burns & Vollmeyer, 2002). It is hypothesized that in Experiment 1 monitoring mediated the transferability of control behaviors, and this was based on the content of the second learning phase. Furthermore, it is postulated that the usefulness of this was retrospectively evaluated from participants' control performance in the test phase of the first problem. Both *self* conditions appear to have judged negatively their own learning phase, and so, assuming it to be less effective, failed to transfer relevant knowledge that would have enabled them to successfully control the system in the second problem. Both *other* conditions, in contrast, appeared to have judged the learning phase of the second problem positively. These evaluations may have been the result of having identified the learning phase as not their own, and thus assuming that it provided a new opportunity to learn. Consequently, they transferred relevant knowledge gained from the first problem to the second, thus facilitating positive transfer of control skills. Before fully exploring the basis for the negative and positive transfer effects found in Experiment 1, a further experiment was devised to investigate the reliability of these effects.

Experiment 2

Experiment 2 examined the reliability of the transfer effect reported in Experiment 1 under practice rather than transfer conditions, and whether disguising the origin of the second learning phase interfered with the transfer effects found in Experiment 1.

Experiment 2 included four conditions: Self, Other, Self-as-instructed-other, Other-as-instructed-self. Unlike Experiment 1, Experiment 2 examined the development of skilled performance through practice rather than transfer of knowledge. Participants were presented with perceptually and structurally identical CDC-tasks, and the learning phase of both problems was procedural-based. This was more likely to produce general practice effects because, in the second presentation of the problem, participants would be highly familiar with the learning phase. If this is so, positive transfer should therefore be found in all four conditions. If, however, the negative and positive transfer effects found in Experiment 1 were robust, then, on measures of performance, the Self condition should show decrements across problems, whereas the Other condition should show improvements across problems.

To complement this, two further conditions were added, in which the origin of the second learning phase was disguised. The Self-as-instructed-other condition were presented with their trial history of the first problem, but told another participant had generated it. The Other-as-instructed-self were presented with the trial history of another participant, but told it was based on theirs from the first problem. This manipulation was intended to examine whether participants negatively evaluate their own learning phase, thus impairing later control performance; and, conversely, whether participants positively evaluate another participant's learning phase. It was hypothesized that, if monitoring influences control behavior, then manipulating belief in the origin of the learning phase would also affect control behavior. If so, then control performance should increase in the Self-as-instructed-other condition, because they now believed the origin of the learning

phase was another participant. Decreases in performance across problems in the Other-as-instructed-self condition should occur, because they now believed they were re-exposed to their own learning phase.

Method

Seventy-two graduate and undergraduate students from University College London volunteered to take part in the experiment and were paid £6. Participants were aged between 19 and 28, and 54 were women. Participants were randomly allocated to one of four conditions (Self, Other, Self-as-instructed-other, Other-as-instructed-self), with eighteen in each. Participants were tested individually.

Design

Experiment 2 was a mixed design. The between subject variable examined the effects of manipulating belief on control performance (Self, Other, Self-as-instructed-other, Other-as-instructed-self), and two within subject variables measured control performance and structural knowledge. In each condition, half the participants were presented with the Water-Tank system problem twice, and the remainder with the Ghost Hunting problem twice. With the exception of the instructional manipulation introduced prior to the second problem, the design of Experiments 1-2 was identical.

Procedure

In all four conditions, the critical manipulation occurred in the second problem. The Self and Other conditions followed the same procedure as in Experiment 1. The Self was presented with the trial history of their own learning phase from the first problem. The Other condition was randomly assigned the trial history of a participant from the Self condition. The Self-as-instructed-other and Other-as-instructed-self differed from the other two conditions in the following way: Before receiving the trial history during the second learning phase, the Self-as-instructed-other condition was told that it was generated

by a participant who had just completed the problem. In fact, the trial history was their own learning phase from the first problem. Before presentation of the second problem, the Other-as-instructed-self condition was randomly assigned a trial history from the Self-as-instructed-other condition. However, they were told that it was based on their learning phase in the first problem. Evidence from self-observation studies show that people reliably detect self-generated behaviors (e.g., Knoblich & Flach, 2001; Knoblich & Prinz, 2001). Therefore the phrasing “based on” rather than “identical to” was used, to appear maximally plausible.

Results

Comparison of test scores in 1st and 2nd Problems. Preliminary analyses revealed no significant difference between conditions based on performance on test and structure test scores in the first problem. The mean error score presented in Figure 4 suggests that, for the Self and Other-as-instructed-self conditions, control error scores increased (worse performance) in both tests in the second problem, whereas control error scores decreased (improved performance) in the Other and Self-as-instructed-other conditions. To analyze this, a 2x2x2x2 ANOVA was carried out using test (Test 1, Test 2) and problem (1st Problem, 2nd Problem) as within subject variables, and source of second learning phase (Self-generated, Other-generated) and belief in the origin of the second learning phase (undisclosed, disguised) as the between subject variables. The analysis showed a significant main effect of Test, $F(1, 68) = 11.47$, $MSE = 0.30$, $p < 0.001$, $\eta^2 = 0.14$, and Belief, $F(1, 68) = 6.09$, $MSE = 0.20$, $p < 0.05$, $\eta^2 = 0.08$. The following interactions were also significant: Source x Problem, $F(1, 68) = 4.86$, $MSE = 0.14$, $p < 0.05$, $\eta^2 = 0.07$; Belief x Problem, $F(1, 68) = 6.84$, $MSE = 0.20$, $p < 0.01$, $\eta^2 = 0.09$; and Belief x Source, $F(1, 68) = 9.47$, $MSE = 0.31$, $p < 0.005$, $\eta^2 = 0.12$. There was also a three-way interaction between Source x Belief x Problem: $F(1, 68) = 23.32$, $MSE = 0.68$, $p < 0.005$, $\eta^2 = 0.26$. Further comparisons confirmed the trends in Figure 4. The increase in control error scores (worse

performance) across problems in the Self condition was confirmed: $t(35) = -5.72, p < 0.0005, d = -1.93$. The decrease in control error scores (improved performance) across problems of the Other and Self-as-instructed-other conditions was also confirmed: $t(35) = 2.49, p < 0.05, d = 0.84$, and $t(35) = 3.32, p < 0.005, d = 1.12$, respectively.

Comparison of structure test scores in 1st and 2nd Problems. As with control performance, Figure 5 indicates that, for the Self and Other-as-instructed-self conditions, structure scores decreased (worse performance) across problems, whereas, for the Other and Self-as-instructed-other conditions, structure scores increased (improved performance) across problems. To analyze this, a 2x2x2 ANOVA was carried out on structure scores averaged across the four tests of each problem, using problem (1st Problem, 2nd Problem) as within subject variables, and source of second learning phase (Self-generated, Other-generated) and belief in the origin of the second learning phase (undisclosed, disguised) as the between subject variables. The following interactions were significant: Source x Belief, $F(1, 68) = 5.05, \text{MSE} = 38.23, p < 0.05, \eta^2 = 0.07$, and Belief x Problem x Source, $F(1, 68) = 26.04, \text{MSE} = 97.62, p < 0.0005, \eta^2 = 0.27$. Further comparisons revealed that the decrease in structure scores (worse performance) between the first and second problems of the Self condition was significant: $t(17) = 2.37, p < 0.05, d = 1.15$. The significant increase in structure scores (improved performance) across problems of the Other and Self-as-instructed-other conditions was also confirmed: $t(17) = -3.08, p < 0.01, d = -0.73$, and $t(17) = -4.57, p < 0.0005, d = -2.22$, respectively.

Correlation between control performance and structural knowledge. A correlation analysis was carried out on mean control error scores and mean structure test scores from the first and second problems, and revealed a significant negative relationship between structure test score and test error scores in the second problem: $r(72) = -0.45, p < 0.001$.

Post-test question. Sixty-seven percent of participants in the Self and 47% in the Self-as-instructed-other conditions reported accurately which condition they were in. Seventy-

eight percent of participants in the Other condition and 61% in the Other-as-instructed-self answered correctly. Pearson's chi-squared analysis revealed no significant difference in the accuracy of responses between conditions.

Discussion

The first objective of Experiment 2 was to examine the reliability of the negative transfer effect found in *Self* conditions and the positive transfer effect found in *Other* conditions. Experiment 2 presented participants with two perceptually and structurally identical problems and, in the Self and Other conditions, replicated the effects reported in Experiment 1.

The negative and positive transfer effects reported in Experiment 1 revealed that monitoring mediated the transferability of control behaviors, based on the content of the second learning phase. Accuracy in responding to post-test questions suggests that participants accurately tracked their learning behaviors, to identify correctly the source of the second learning phase. Moreover, by monitoring this learning phase, participants made biased assessments of its effectiveness. To examine this, Experiment 2 included two further conditions, in which the origin of the second learning phase was disguised. The findings confirmed the prediction that participants would make biased assessments of the second learning phase. The evidence showed that, although participants in the Self-as-instructed-other condition were re-exposed to their own learning phase, the belief that it belonged to another participant generated patterns of performance consistent with the Other condition. Control performance and accuracy of structural knowledge increased in the second problem, and only 47% of participants responded accurately to the post-test question. This suggests that the instruction convincingly persuaded them to attribute the learning phase to another individual. Moreover, the obvious similarity between the second and first learning phases must have been rationalized in terms of others generating similar hypothesis testing behaviors. In contrast, over 70% of participants in the Other-as-

instructed-self condition responded accurately to the post-test question, suggesting that they may have been less convinced by the instructional manipulation. This may also explain why control performance and accuracy of structural knowledge in the second problem were equivalent in both problems. To explore further the effects of monitoring on the transferability of control behaviors, a third experiment was designed. Whereas Experiment 2 manipulated belief, Experiment 3 introduced biases that led to negative and positive evaluations of self-generated learning instances, which in turn affect later control behavior.

Experiment 3

Consistent with Social Cognitive theory and the Dual-Space hypothesis, the evidence from Experiments 1 and 2 indicates that monitoring influences the transferability of control skills across CDC-tasks. The findings suggest that, when presented with the second CDC-task during the learning phase, participants judge the relevance of the hypothesis testing behavior they are experiencing, to gain structural knowledge and experience of the system. In turn, these judgments have consequences for the transferability of skills gained in the initial CDC-task. To examine this, Experiment 3 included two conditions. Both were *self* conditions: Act-on-self-high, Act-on-self-low. Before the second problem, in both conditions, participants were given advance knowledge that they would be presented with their own trial history from the first problem. They were also given bogus information about the average performance of other participants that had solved the first problem. Participants in the Act-on-self-high condition were told that average performance was extremely good, whereas participants in the Act-on-self-low condition were told that it was extremely bad. If monitoring mediates the transferability of control skills, then the information presented to both conditions prior to the second problem should influence their judgments of their learning and control behaviors. The Act-on-self-high condition should judge negatively their control performance in the first problem, and the learning behaviors that contributed to their understanding of it, thus

impeding transfer of control skills. Conversely, the Act-on-self-low condition should positively judge their control performance and their learning behaviors in the first problem, thus facilitating transfer of control skills. Alternatively, as all participants are told that they will be experiencing their first trial history again, they may simply fail to attend to it, in which case, with no basis to learn about the system again, both conditions should show equivalent levels of performance in both the first and second CDT-task.

Method

Thirty-six graduate and undergraduate students from University College London volunteered to take part in the experiment and were paid £6. Participants were aged between 18 and 32, and 18 were women. Participants were randomly allocated to one of two conditions (Act-on-self-high, Act-on-self-low), with eighteen in each. Participants were tested individually.

Design

Experiment 3 was a mixed design that included a between subject variable examining the effects of biases on transferability of skill, by comparing two conditions (Act-on-self-high, Act-on-self-low), and two within subject variables measuring control performance and structural knowledge. In each condition, half the participants were presented with the Water-Tank system problem twice, and the remainder with the Ghost Hunting problem twice. Both conditions were presented with a trial history of their learning phase from the first problem. The critical difference between the conditions was the information provided before the presentation of the second problem.

Procedure

Both conditions performed the first problem following the same procedure as in Experiments 1-2. Before presentation of the second problem, participants were informed that they would be presented with a trial history of their learning phase from the first

problem. The Act-on-self-high condition was told that the average performance of other participants that had solved the first problem was extremely good: within $\pm 20^5$ of each target value for each output on both tests. The Act-on-self-low condition, however, was told it was extremely poor: within ± 200 of each target value for each output on both tests.

In addition to the two standard measures of performance used in this study, a memory test was presented directly after the second learning phase, to examine the effects of biases on the accuracy of participants' recall of the learning phase. The memory test consisted of a blank trial history sheet, which participants required, on average, three minutes to complete. In it, for each of the 12 trials, they recalled which inputs they had changed, and the values they were changed by. Participants were not told in advance that they would receive this test, but were warned that they would receive a test based on their knowledge of the second learning phase. Knowledge of an impending test was designed to motivate participants to attend closely to the second learning phase, in particular given that they were aware that they would be experiencing their original learning phase again.

Results

Comparison of test scores in 1st and 2nd Problems. Preliminary analyses revealed no significant difference between conditions based on performance on test and structure test scores in the first problem. Figure 6 suggests that, for the Act-on-self-high condition, control error scores increased (worse performance), whereas for the Act-on-self-low control error scores decreased (improved performance) between the first and second problem. A 2x2x2 ANOVA was carried out, using test (Test 1, Test 2) and problem (1st Problem, 2nd Problem) as within subject variables, and condition (Act-on-self-high, Act-on-self-low) as the between subject variable. The analysis showed a significant main effect of test: $F(1, 34) = 15.54$, $MSE = 0.42$, $p < 0.0005$, $\eta^2 = 0.34$. There was a significant main effect of condition, $F(1, 34) = 12.13$, $MSE = 0.26$, $p < 0.001$, $\eta^2 = 0.38$, and a significant

Condition x Problem interaction, $F(1, 34) = 21.14$, $MSE = 1.15$, $p < 0.001$, $\eta^2 = 0.26$. The significant increase in control error scores (worse performance) across problems of the Act-on-self-high condition (Figure 6) was confirmed by a further comparison: $t(35) = -4.70$, $p < 0.0005$, $d = -1.59$. The significant decrease in control error scores (improved performance) across problems of the Act-on-other-low condition was also confirmed: $t(35) = 2.54$, $p < 0.05$, $d = 0.86$.

Comparison of structure test scores in 1st and 2nd Problems. Figure 7 indicates that, for the Act-on-self-high condition, structure scores decreased (worse performance) across problems, whereas, for the Act-on-self-low condition, structure scores (improved performance) across problems. To analyze the pattern of behavior indicated in Figure 7, a 2x2 ANOVA was carried out, using problem (1st Problem, 2nd Problem) as the within subject variable, and condition (Act-on-self-high, Act-on-other-high) as the between subject variable. There was a significant Condition x Problem interaction: $F(1, 34) = 8.46$, $MSE = 210.98$, $p < 0.01$, $\eta^2 = 0.19$. The decrease in structure scores (worse performance) between the first and second problems of the Act-on-self-high condition (Figure 8) was confirmed: $t(71) = 3.21$, $p < 0.005$, $d = 0.73$. In addition, the increase in structure scores (improved performance) across problems of the Act-on-other-low condition was confirmed: $t(71) = -3.83$, $p < 0.0005$, $d = -0.90$.

Memory scores. Responses to the memory test presented at the end of the learning phase in the second problem were scored in two ways. The recalled inputs changed on each trial were scored similarly to structure scores: i.e., (the number of correctly recalled input changes, and incorrect input changes avoided) / N (the total number of input changes that can be made). The final score for each participant was converted to a percentage that was used in the analysis. The recalled input values of those changed on each trial were scores based on subtracting the absolute recalled value for an input that was correctly recalled from the actual value of that input. This procedure was carried out for each trial of

the learning phase, and the average difference between recalled and actual input values represented the input value score.⁶ Given that the data was not normally distributed, non-parametric tests were conducted on the data. However, in a more stringent analysis, parametric tests were also conducted. Both analyses found no difference between conditions based on their recall of which inputs were changed, or by how much, in the learning phase.

Correlation between control performance and structural knowledge. The analysis revealed a significant negative relationship between structure test scores and test error scores in the first problem, $r(36) = -0.46, p < 0.005$, and second problem, $r(72) = -0.41, p < 0.05$.

Discussion

Consistent with Social Cognitive theory and Dual-Space hypothesis, the evidence from Experiment 3 confirmed the hypothesis that monitoring mediates the transferability of control skill across problems. The prediction was confirmed that, if sensitive to the instructional manipulations, the Act-on-self-high condition would show negative transfer, and the Act-on-self-low condition would show positive transfer. Similar findings have been reported in studies showing that erroneous feedback influences performance on tests of stamina (Litt, 1988), physical strength (Weinberg et al., 1981), strategic thinking (Bouffard-Bouchard, 1990), and complex decision making (Hogarth, Gibbs, McKenzie, & Marquis, 1991). In these examples, participants were presented with bogus normative standards that either suggested they had performed higher than the mean—which later elevated their performance—or lower than the normative standards—which then impaired performance.

The results of the memory test indicated that, for both conditions, the instructional manipulations in the second learning phase did not differentially affect recall of it. This result speaks to an important issue concerning the negative transfer effect reported in *self-experience* conditions (i.e., Experiments 1-2). It could be argued that participants' familiarity with the second learning phase leads to failure to attend to it, thus disadvantaging them

because they do not use it as an opportunity to learn. The evidence from the memory tests rules out this possible explanation: Although both conditions demonstrated equivalent memory of it by paying attention to it, their control performance still differed.

Experiment 4

Thus far, the learning phase in each problem has been exploratory. However, during learning participants gain no experience of controlling the system to criterion, which is necessary for the test phase. Moreover, the number of trials used in each learning phase has so far been only 12, as in the original CDC-task by Burns and Vollmeyer (2002). Therefore, Experiment 4 included two manipulations. First, there were 40 learning trials, not 12. Second, instead of being exploratory, in each learning phase participants pursued a specific goal, in which they were required to control the system to criterion—the same criterion of the first test in each problem. Two conditions were used (Observe-self, Observe-other), in which the original and transfer CDC-tasks differed. It was thus possible to examine the generality of the transfer effects found in Experiments 1-3, under conditions that provided greater opportunity to learn about two perceptually different systems, and how to control them.

Experiment 4 further examined the relationship between monitoring and control, by including judgments of self-efficacy: This refers to people's belief in their ability to exercise control over environmental events. This was achieved, after each learning phase, by asking participants to estimate how well they could control the system. In addition, after the second learning phase, participants were asked to estimate how much they based their understanding of the system on their structural knowledge of the first system. This question was included to examine whether the way prior experience was used in the second problem discriminated between the conditions. If monitoring influences the transfer of control behavior, then judgments of self-efficacy taken prior to the second test phase should be lower in the Observe-self condition than in the Observe-other condition, and in

each condition they should correspond to control performance. Moreover, if participants in the Observe-self conditions negatively evaluate the effectiveness of their learning phase from the first problem, then, in the second problem, they should report relying less on previous structural knowledge than the Observe-other condition.

Method

Thirty-two graduate and undergraduate students from University College London volunteered to take part in the experiment and were paid £10. Participants were aged between 22 and 31, and 16 were women. Participants were randomly allocated to one of two conditions (Observe-self, Observe-other), with sixteen in each. Participants were tested individually.

Design

Experiment 4 examined the effects of monitoring on transferability of control skill, by increasing the trials in the learning phase. It was a mixed design that included a between subject variable (Observe-self, Observe-other), and four within subject variables that measured control performance, structural knowledge, self-efficacy, and use of prior knowledge. Each participant solved the Water-Tank and Ghost Hunting CDC-task problems, and the order of presentation of the two problems was randomized for each participant. Each CDC-task now comprised 40 trials in the learning phase, divided into four blocks of 10, after which a structure test was presented. In the learning phase of each CDC-task, participants were instructed to learn about the system whilst trying to control it to specific criteria, which were the output criteria, exactly as presented in Test 1. In all other respects, Experiment 4 was identical to Experiment 1.

Procedure

During the learning phase of the first problem, participants were told that they would have practice at controlling the system to specific criteria whilst trying to learn about

how it operated (see Appendix for instructions). For each block, participants had to reach and maintain the same output criteria.³ Participants were unaware that the criteria they had to follow were the same criteria used in the first test. A structure test was presented after each block. After the learning phase, the test phase was presented.

The learning phase of the second problem was observation-based, comprising 40 trials, blocked into 4. Both conditions were given specific goal instructions (see Appendix). Their job was to observe carefully the changes to the inputs and outputs on each trial, and to assess how successfully the output values met the criteria output values, which were identical to those of the first test of the test phase. As in the procedure in Experiment 1, the Observe-other condition were randomly allocated to the learning phase generated by a participant from the Observe-self condition, whereas the Observe-self condition watched the learning phase they had generated from the first problem.

For each problem, after the learning phase was completed, participants were presented with the following question: *Based on what you have learnt, how well do you think you can now control the system?* Participants were told to imagine that they would be tested on their ability to control the system to the same criteria that they had just practiced on. They were told to estimate how close, on average, they could get to the criteria values, by choosing from the following ranges: 1) +/- 25; 2) +/- 50; 3) +/- 75; 4) +/- 100; 5) +/- 125; 6) +/- 150; 7) +/- 175; 8) +/- 200. After the test phase, participants were asked to make another similar self-efficacy judgment: *Now you have had a chance to control the system to different criteria, how well do you think you controlled the system in general?* In addition to the self-efficacy judgments, after the second learning phase, participants were also asked *To what extent did you base your current understanding of the relationship between the inputs ... [salt, carbon, lime/ GGH Meter, Anemometer, Trifield Meter] and the outputs ... [oxygenation, chlorine concentration, temperature/ Electro Magnetic Waves, Radio Waves, Air Pressure] on your understanding of the*

relationship between the inputs and outputs from the previous problem? Responses were made using a 9-point scale ranging from negative (Not at all) to positive (Mostly).

Results

Learning phase. Because participants were presented with a specific goal in each of the four blocks of the learning phase, the following analysis examines control performance in the learning phase. A 4x2 ANOVA was conducted on control error scores using block (Block 1-4) as the within subject variable, and condition (Observe-self, Observe-other) as the between subject variable. No effects were significant, suggesting that across blocks there was no difference in control performance during the first learning phase.

In addition, for each participant, a simple strategy analysis was conducted, based on the number of input variables changed for each trial recorded, and averaged for each block for each condition, as shown in Figure 8. Figure 8 indicates that participants changed fewer inputs in the second and third blocks of the learning phase than in the first and final blocks. To analyze this, a 4x2 ANOVA was conducted using block (Blocks 1-4) as the within subject variable, and condition (Observe-self, Observe-other) as the between subject variable. The analysis revealed a significant main effect of block: $F(3, 90) = 4.86$, $MSE = 2.17$, $p < 0.005$, $\eta^2 = 0.14$. Confirming Figure 8, t-tests showed significant differences in the number of inputs changed between Block 2 and Block 1, $t(31) = -2.20$, $p < 0.05$, $d = -0.79$, between Block 2 and Block 4, $t(31) = -3.51$, $p < 0.001$, $d = -1.26$, and between Block 3 and Block 4, $t(31) = -2.44$, $p < 0.05$, $d = -0.88$. Taken together, the findings suggest that between conditions there appeared to be no improvement in participants' ability to control the system across blocks, possibly because they were developing different strategies in the course of the learning phase, and thus varied, from block to block, the number of inputs that they changed. This is consistent with Burns & Vollmeyer's (2002) findings.

Comparison of test scores in 1st and 2nd Problems. Preliminary analyses revealed no significant difference between conditions based on performance in the test phase, and

structure test scores, in the first problem. Figure 9 shows that, for the Observe-self condition, control error scores increased (worse performance) across problems, whereas for the Observe-other condition control error scores decreased (improved performance) across problems. To analyze this, a 2x2x2 ANOVA was carried out, using test (Test 1, Test 2) and problem (1st Problem, 2nd Problem) as within subject variables, and condition (Observe-self, Observe-other) as the between subject variable. The analysis showed a significant main effect of test: $F(1, 30) = 7.72$, $MSE = 0.22$, $p < 0.01$, $\eta^2 = 0.21$. There was a significant main effect of condition, $F(1, 30) = 4.99$, $MSE = 0.16$, $p < 0.05$, $\eta^2 = 0.14$, and a significant Condition x Problem interaction, $F(1, 30) = 13.17$, $MSE = 0.28$, $p < 0.001$, $\eta^2 = 0.31$. The increase in error scores (worse performance) between problems of the Observe-self condition (Figure 9) was confirmed by a further comparison: $t(31) = -3.05$, $p < 0.005$, $d = -1.09$. The decrease in control error scores (improved performance) across problems of the Observe-other condition was also confirmed: $t(31) = 2.31$, $p < 0.05$, $d = 0.83$. These trends are consistent with those reported in Experiment 1.

Comparison of structure test scores in 1st and 2nd Problems. Figure 10 shows that, for the Observe-self condition, structure scores decreased (worse performance) across problems, but structure scores increased (improved performance) for the Observe-other condition. A 2x2 ANOVA was carried out, using problem (1st Problem, 2nd Problem) as the within subject variable, and condition (Observe-self, Observe-other) as the between subject variable. There was a significant main effect of Condition, $F(1, 30) = 6.20$, $MSE = 54.89$, $p < 0.05$, $\eta^2 = 0.17$, and Condition x Problem interaction, $F(1, 30) = 5.38$, $MSE = 33.45$, $p < 0.05$, $\eta^2 = 0.15$. The decrease in structure scores (worse performance) between the first and second problem for the Observe-self condition (Figure 10) was confirmed: $t(15) = 2.27$, $p < 0.05$, $d = 1.17$. The increase in structure scores (improved performance) across problems of the Observe-other condition approached significance: $t(15) = -2.05$, $p = 0.058$, $d = -1.06$.

Judgments of self-efficacy and structural knowledge. The relationship between monitoring and control was analyzed by examining the association between judgments of self-efficacy and control performance (averaged across Tests 1-2 for each problem). Only efficacy judgments taken before the test phase were found to track control performance accurately, in both the first problem, $r(32) = 0.53, p < 0.005$, and the second, $r(32) = 0.60, p < 0.0005$. Prior to the second test phase, the mean self-efficacy judgment in the Observe-self condition was 151.56 (SD 30.91), but for the Observe-other conditions the range was 125 (SD 38.96). To analyze the pattern of judgments of self-efficacy between conditions, a 2x2x2 ANOVA was conducted on judgments of self-efficacy, using stage (before test phase, after test phase), and problem (1st Problem, 2nd Problem) as the within subject variables, and condition (Observe-self, Observe-other) as the between subject variable. The Stage x Condition interaction was significant: $F(1, 30) = 10.42, \text{MSE} = 7812.50, p < 0.005, \eta^2 = 0.26$. T-tests showed that there was a difference between conditions, based on self-efficacy judgments recorded prior to the control phase in the second problem: $t(15) = 2.29, p < 0.05, d = 1.18$. Thus, in the second problem, there were differences between conditions, showing that the Observe-self judgments of self-efficacy were lower than those of the Observe-other condition.

Participants were also asked to judge the extent to which they used their knowledge of the structure of the first system as a basis for understanding the second. The Observe-self condition's mean response was 2.5 (SD 1.93), lower than the Observe-other condition, 5.25 (SD 2.29). The responses between conditions were compared, and indicated that the Observe-other condition relied more on their prior knowledge to help them in the second problem than the Observe-self condition: $t(30) = -3.67, p < 0.001, d = -1.34$. These judgments were also correlated with mean structure scores in the second problem, and were found to be significant: $r(32) = 0.66, p < 0.005$. The judgments were also correlated with control performance, averaged across Tests 1-2 in the second problem, and revealed a

significant negative relationship: $r(32) = -0.45, p < 0.005$. This suggests that the extent to which prior knowledge was judged relevant was associated with measures of structural knowledge and control performance in the second problem.

Post-test question. The accuracy of responses to the post-test question was comparable to that in Experiment 1. Seventy-five percent of participants in the Observe-self and 81% in the Observe-other conditions accurately reported which condition they were in.

Discussion

The patterns of transfer across CDC-tasks found in Experiment 4 replicated those in Experiment 1. The Observe-self condition showed negative transfer and the Observe-other condition showed positive transfer. Consistent with Social Cognitive theory and the Dual-Space hypothesis, monitoring behaviors appear to mediate the transferability of control performance. Prospective judgments, not retrospective judgments, of self-efficacy of control performance were associated with actual control performance. Moreover, the prediction that estimations of control performance would be lower in Observe-self conditions than in Observe-other conditions was confirmed. To complement this, estimates of the relevancy of prior structural knowledge confirmed the prediction that the Observe-self condition would report relying less on their previous knowledge from the first problem than the Observe-other condition. The associations found between these judgments and control performance and structural knowledge in the second problem suggest the following: negative evaluations of the second learning phase prevented the Observe-self condition from transferring potentially relevant structural information from the first problem to the second. This disadvantaged them when controlling the system in the second test phase. For the Observe-other conditions, positive evaluations of the learning phase may have given them reason to draw comparisons with their own experiences, which they then viewed as relevant. By transferring relevant knowledge gained

from the first problem to the second, this condition's control performance improved in the second test phase.

General Discussion

The objective of this study was to uncover the effects of monitoring on the transferability of control behaviors across analogical skill learning tasks. To do this, the study asked four specific questions. (1) *Does control performance improve if monitoring is based on one's prior self-generated behavior, rather than the behavior of another individual?* Re-experiencing self-generated learning behaviors produced negative transfer of control skill, but experiencing another's learning behaviors produced positive transfer of control skill. These effects were found regardless of whether monitoring of prior learning behaviors was via observation, or direct interaction with the system (Experiment 1). Moreover, the effects were found both when the transfer task was identical to (Experiments 2-3), and perceptually dissimilar to, the original task (Experiments 1, 4). (2) *Can people discriminate between their own self-generated behavior and that of another individual?* Experiments 1, 2, and 4 showed that, in general, participants were highly accurate in judging whether the learning behaviors they were exposed to was self-generated or not. Importantly, accuracy of detection was found to be independent of the format in which learning took place (observation, action). (3) *Is control performance improved if monitoring of self-generated or other-generated behaviors takes place online rather than indirectly via observation?* Experiment 1 showed that facilitation of the transfer of control skills was found to be independent of the medium (observation, action) by which monitoring occurred. (4) *Can indices of monitoring behavior accurately predict the transferability of control behaviors in a complex skill learning task?* When forecasting their control performance, based only on their experiences during the learning phase, participants' judgments accurately predicted the transferability of their control skills across analogous CDC-tasks (Experiment 4).

The evidence from this study is consistent with Social Cognitive theory and the Dual-Space hypothesis, but it conflicts with Dissociationist claims that monitoring has detrimental effects on the transfer of skill. The following discussion examines this disparity, and the factors determining the transfer of control behaviors.

Disparity between Dissociationist claims and the present study

Dissociationists claim that control skills in CDC-tasks are procedural, and their transferability is limited because procedural knowledge is perceptually bound and inflexible (e.g., Berry, 1991; Berry & Broadbent, 1988; Dienes & Berry, 1997; Lee, 1995; Sun et al., 2001). This claim is supported by findings that control skills are transferred only if the transfer task itself is perceptually and structurally similar to the original (Berry & Broadbent, 1988), and that if learning in both is procedural-based (Berry, 1991; Berry & Broadbent, 1988). To explain the disparity between the dissociationist position and evidence from the present study, the following discussion considers the issues in terms of the Dual-Space hypothesis and Social Cognitive theory.

Common to studies that reveal dissociations in a CDC-task is that hypothesis testing behaviors are prevented during learning. This is either because learning takes place under specific goal conditions (e.g., Berry, 1991; Berry & Broadbent, 1984, 1987, 1988; Broadbent et al., 1986; Dienes & Fahey, 1995, 1998; Marescaux et al., 1989; Stanley et al., 1989), or because participants have been explicitly instructed to avoid hypothesis testing (Berry, 1991). Another common pattern is that dissociations are found when measures of declarative knowledge are taken after, rather than during, learning (e.g., Berry, 1991; Berry & Broadbent, 1984, 1987, 1988; Dienes & Fahey, 1995, 1998; Marescaux et al., 1989).

Common to studies of CDC-tasks that encourage hypothesis testing is evidence of associations between declarative knowledge and procedural knowledge (e.g., Burns & Vollmeyer, 2002; Gonzalez et al., 2003; Gonzales & Quesada, 2003; Jensen & Brehmer, 2003; Sweller, 1988). Moreover, performance measures of both types of knowledge exceed

those of conditions in which hypothesis testing is prevented (Burns & Vollmeyer, 2002; Osman, in press, Sweller, 1988; Vollmeyer et al., 1996). In addition, studies that take multiple measurements of declarative knowledge during learning reveal associations with procedural knowledge (Burns & Vollmeyer, 2002; Sanderson, 1989; Sanderson & Vicente, 1986; Voss, Wiley & Carretero, 1995).

Both, Social Cognitive theory and Dual-Space hypothesis posits that monitoring serves a regulatory function, because it tracks and selects out relevant information bearing on a desired outcome. More specifically, Dual-Space theory claims that hypothesis testing focuses the learners' attention on both relevant properties of the CDC-task: i.e., the rule and instance space. Additionally, taking multiple measurements of declarative knowledge during learning also prompts participants to keep track of their strategies, and continually update their knowledge of the input-output relations of the CDC-task. This provides a means of relating their understanding of the structure of the system to their experiences of how it operates. Hence why, instructions to hypothesis test and taking multiple measurements of declarative knowledge at the time of its acquisition included in this study revealed associations between declarative knowledge and procedural knowledge.

The Determinants of Transferable Complex Control Behaviors

Through monitoring, learners have online awareness of their behaviors during skill acquisition and application, and thus are able to appraise them. This study shows that monitoring and control behaviors are interrelated, and critical to the transfer of skilled behaviors. Experiment 1 indicated that the evaluation of the second learning phase was critical to the success of the transferability of control skills. How it was evaluated was then manipulated, by disguising its source (Experiment 2); providing external normative standards to judge it (Experiment 3); and introducing a specific goal by which to assess it validly (Experiment 4).

How does monitoring affect the transferability of control skills? This study posits that through hypothesis testing, learners search through rule and instance spaces. This search is also conducted in transfer tasks. By uncovering underlying similarities, the learner is alert to previous knowledge (i.e., input-output relations (rules), how the system operates (instances)) that can be brought to bear on the transfer task. The role of monitoring is to judge which knowledge is brought to bear. More specifically, monitoring is a means of diagnosing the effectiveness of the second and first learning phase which then determines the transferability of control skills acquired from the first task to the second. As an index of the quality of the learning phase, performance in the test phase of the first problem was used in a regression analysis as a function of the transferability of control performance in the second. Thus, performance in the second test phase of participants in *other* conditions $n=70$ (Exp. 1 Observe-other & Act-on-other, Exp.2 Act-on-other, Exp.4 Observe-other) was predicted by their own performance in the first test phase ($R^2=0.41$, $p<0.001$), and moderately predicted by the performance in the first test phase of the participants they were yoked to ($R^2=0.18$, $p<0.05$). However, in *self* conditions $n=70$ (Exp.1 Observe-self & Act-on-self, Exp. 2 Act-on-self, Exp.4 Observe-self), control performance in the second test phase was not predicted by the first ($R^2=0.06$).

To explain these differences this study proposes that in *Other* conditions, comparing one's own hypothesis testing behaviors during learning, with another's, was used to diagnose the relevance of prior experience. Prior knowledge was used to anchor assessments of the new information presented during the second learning phase. Similarity between the self's and another's (Experiment 1, 2, & 4) or supposed other's (Experiment 2, Self-as-instructed-other) strategies was judged positive, and consequently previous knowledge was judged relevant to the transfer task (Experiment 4, prior knowledge judgments). The reason for this is because individuals use performance standards to consolidate both the knowledge they have gained and beliefs in their self-efficacy, which

enhances performance in similar ways as proactive regulatory mechanism (Bandura & Cervone, 1986).

Without external normative standards, such as making a comparison with another's learning experiences, self-perceptions of the knowledge and control ability of *self* conditions (Experiments 1, 2 (Act-on-self), 4) lead to negative self-assessments. Moreover, along with poor self-efficacy judgments, individuals also undervalued the relevancy of previously gained knowledge in assisting them in the transfer task (Experiment 4). The reason for this is because individuals overcompensate in their error detection and correction, which leads them to ignore rather than transfer relevant prior information. As a consequence this has negative affects on their performance because they have failed to utilize prior relevant knowledge, much like an overactive reactive regulatory mechanisms (Bandura & Locke, 2002).

Conclusion

Previous research on CDCTs has provided an impoverished understanding of the types of knowledge that are transferable, and the modulating factors that lead to successful and unsuccessful transfer of skilled behavior. The present article was designed to address this and, by studying transferability of skills, was able to provide new insights into the learning process that takes place in CDC-tasks. The evidence revealed an association between declarative and procedural knowledge (Experiments 1-4), the acquisition of procedural skills through observation (Experiments 1, 4), and accurate monitoring of internally represented behaviors and self-insight (Experiments 1, 2, 4).

Both Social Cognitive theory and Dual-Space theory provide foundations for claiming that problem solvers are sensitive to, and influenced by, their assessment of the effectiveness of self-generated learning instances, and that this plays a significant role in facilitating and attenuating the transferability of knowledge in complex dynamic learning environments. The two most pivotal findings demonstrating this were the successful

transfer of control skills and structural knowledge across analogous CDCTs, and the atypical negative transfer (anti-learning) effect in which measures of control skills and structural knowledge in the transfer problem were impaired relative to the original.

Footnotes

1. In Burns and Vollmeyer's study, participants were shown the starting values of input and output values before they began the task. In the present experiment, participants were shown only the starting values of the input values, and not the output values, which were revealed only on the first trial, and not before. The rationale for this change was simply to encourage participants to pay special attention to the effects on the outputs resulting from the manipulations they made.
2. If a participant changed the input Salt by 50 units on Trial 1, this would in turn change the output value of Chlorine Concentration to 556 (i.e., Chlorine Concentration starting value = 500 units, + Salt value change = 50 units, + Constant added noise on input-output connection = 6 units). If on Trial 2 the input Salt was changed by 100 units, then the output value of Chlorine Concentration would be 662 (i.e., Chlorine Concentration starting value = 556 units, + Salt value change = 100 units, + Constant added noise on input-output connection = 6 units).
3. For each problem at the start of each block of the learning phase, and at the beginning of each test, the input values were set to 0, and the output levels were set as follows: Output 1 (Water Tank = Oxygenation, Ghost Hunting = Radio Waves) = 100; Output 2 (Water Tank = Chlorine Concentration, Ghost Hunting = Electro Magnetic Waves) = 500; Output 3 (Water Tank = Temperature, Ghost Hunting = Air Pressure) = 1000.
4. When participants received a learning trial history (whether to observe or act on) from the first problem, the labels were changed to reflect the new problem.
5. The mean discrepancy between achieved values and target values for each participant in each test, problem, condition, and experiment was ranked and used as a basis for generating the values +/-20 and +/- 200 used in Experiment 3. These values were the extreme ends of the range generated.
6. The mean Input change scores of the Act-on-self-high condition and the Act-on-self-low condition were fairly low (38% and 36%, respectively). However, when input change scores were based on the first 6 trials, the Act-on-self-high condition and the Act-on-self-low condition scored above 50% (68% and 66%, respectively); but, when input change scores were based on the remaining 6 trials, the figures

were lower (19% and 19%, respectively). The trend strongly suggests a primacy effect.

References

- Albright, L., & Malloy, T. E. (1999). Self-observation of social behavior and metaperception. *Journal of Personality and Social Psychology*, *77*, 726-734.
- Bailey, K. G., & Sowder, W. (1970). Audiotape and videotape self-confrontation in psychotherapy. *Psychological Bulletin*, *74*, 127-137.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*, 248-287.
- Bandura, A., & Cervone, D. (1986). Differential engagement of self-active influences in cognitive motivation. *Organizational Behavior and Human Decision Processes*, *38*, 92-113.
- Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, *88*, 87-99.
- Berry, D. (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology*, *43*, 881-906.
- Berry, D., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, *36*, 209-231.
- Berry, D., & Broadbent, D. E. (1987). The combination of implicit and explicit knowledge in task control. *Psychological Research*, *49*, 7-15.
- Berry, D. C. & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, *79*, 251-272.
- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211-241.

- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *Journal of Social Psychology, 130*, 353-363.
- Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *Quarterly Journal of Experimental Psychology, 55*, 241-261.
- Carroll, W. R., & Bandura, A. (1982). The role of visual monitoring in observation learning of action patterns: Making the unobservable observable. *Journal of Motor Behavior, 14*, 153-167.
- Cañas, J. J., Quesada, J. F., Antoli, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics, 46*, 482-501.
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative view. *Annual Review of Psychology, 51*, 171-200.
- Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*, 3-23.
- Dienes, Z., & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 848-862.
- Dienes, Z., & Fahey, R. (1998). The role of implicit memory in controlling a dynamic system. *Quarterly Journal of Experimental Psychology, 51*, 593-614.
- Dowrick, P. W. (1983). Self-modeling. In P. W. Dowrick & S. J. Briggs (Eds.), *Using video: Psychological and social applications* (pp. 105-124). New York: Wiley.
- Ericsson, K. A. (Ed.). (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Hillsdale, NJ: Erlbaum.
- Ericsson, K. A., & Lehman, A. (1996). Expert and exceptional performance: Evidence of maximal adaptation to constraints. *Annual Review of Psychology, 47*, 273-305.
- Fireman, G., & Kose, G. (1991). Video training as a means for enhancing self-awareness in problem solving among young children. *Resources in Education, 330*, 492.

- Fireman, G., & Kose, G. (2002). The effect of self-observation on children's problem solving. *Genetic Psychology, 163*, 410-423.
- Fireman, G., Kose, G., & Solomon, M. (2003). Self-observation and learning: The effect of watching oneself on problem solving performance. *Cognitive Development, 18*, 339-354.
- Funke, J. (2001). Dynamic systems as tools for analyzing human judgment. *Thinking and Reasoning, 7*, 69-89.
- Geddes, B. W., & Stevenson, R. J. (1997). Explicit learning of a dynamic system with a non-salient pattern. *Quarterly Journal of Experimental Psychology, 50A*, 742-765.
- Giesler, B. R., Josephs, R. A., & Swann, W. B. (1996). Self-verification in clinical depression: The desire for negative evaluation. *Journal of Abnormal Psychology, 105*, 358-368.
- Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40*, 631-666.
- Gonzales, C., & Quesada, J. (2003). Learning in dynamic decision making: The recognition process. *Computational & Mathematical Organization Theory, 9*, 287-304.
- Hill, R. J., Gordon, A., and Kim, J. (2004). Learning the lessons of leadership experience: Tools for interactive case method analysis. In *Proceedings of the Twenty-fourth Army Science Conference*.
- Hogarth, R. M., Gibbs, B. J., McKenzie, C. R. M., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 734-752.
- Jensen, E., & Brehmer, B. (2003). Understanding and control of a simple dynamic system. *Systems Dynamic Review, 19*, 119-137.
- Karoly, P. (1993). Mechanisms of self-regulation: A systems view. *Annual Review of Psychology, 44*, 23-52.

- Kelly, S., & Burton, A. M. (2001). Learning complex sequences: No role for observation. *Psychological Research, 65*, 15-23.
- Kelly, S., Burton, A. M., Riedel, B., & Lynch, E. (2003). Sequence learning by action and observation: Evidence for separate mechanisms. *British Journal of Psychology, 94*, 355-372.
- Kerstholt, J. H. (1996). The effect of information costs on strategy selection in dynamic tasks. *Acta Psychologica, 94*, 273-290.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-55.
- Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science, 12*, 467-472.
- Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 456-465.
- Lee, Y. (1995). Effects of learning contexts on implicit and explicit learning. *Memory and Cognition, 23*, 723-734.
- Lehmann, A. C., & Ericsson, K. A. (1997). Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students. *Psychomusicology, 16*, 40-58.
- Lerch, F. J., & Harter, D. E. (2001). Cognitive support for real-time dynamic decision making. *Information Systems Research, 12*, 63-82.
- Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*, 331-352.
- Litt, M. D. (1988). Self-efficacy and perceived control: Cognitive mediators of pain tolerance. *Journal of Personality and Social Psychology, 54*, 149-160.

Locke, E. A., & Latham, G. P. (2000). Building a practically useful theory of goal setting and task motivation. *American Psychologist*, 57, 705-717.

Marescaux, P-J., Luc, F., & Karnas, G. (1989). Modes d'apprentissage selectif et nonselectif et connaissances acquies au control d'un processus: Evaluation d'un modele simule. [Selective and nonselective learning modes and acquiring knowledge of process control: Evaluation of a simulation model] *Cahiers de Psychologie Cognitive*, 9, 239-264.

Osman, M. (2004). An Evaluation of Dual Process Theories of Reasoning *Psychonomic Bulletin & Review*, 11, 998-1010.

Osman, M. (in press). Observation can be as effective as action in problem solving. *Cognitive Science*.

Pintrich, P. R., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.

Rossano, M. J. (2003). Expertise and the evolution of consciousness. *Cognition*, 89, 207-236.

Sanderson, P. M. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation, and mental models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 729-747.

Sanderson, P. M., & Vicente, K. J. (1986). *Verbalizable knowledge and skilled task performance: Explaining association and dissociation*. (Technical Report EPL-86-04). Urbana, IL: University of Illinois at Urbana-Champaign. Engineering Psychology Research Laboratory.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stanley, W. B., Mathews, R. C., Buss, R. R. & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction, and practice in a simulated process CDCT. *Quarterly Journal of Experimental Psychology*, 41, 553-577.

- Stanovich, K. E. (2004). *The robot's rebellion*. Chicago: University of Chicago Press.
- 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 of learning. *Cognitive Science*, 12, 257-285.
- Sweller, J. (2003). Evolution of human cognitive architecture. *Psychology of Learning and Motivation: Advances in Research and Theory*, 43, 215-266.
- Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-end analysis and learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 8, 463-474.
- Trumpower, D. L., Goldsmith, T. E., & Guynn, M. (2004). Goal specificity and knowledge acquisition in statistics problem solving: Evidence for attentional focus. *Memory & Cognition*, 32, 1379-1388.
- VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513-539.
- Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity and systematicity of strategies on the acquisition of problem structure. *Cognitive Science*, 20, 75-100.
- Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, 10, 293-309.
- Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring Intellectual Skills. *Annual Review of Psychology*, 46, 155-181.
- Weinberg, R. S., Gould, D., Yukelson, D., & Jackson, A. (1981). The effect of preexisting and manipulated self-efficacy on a competitive muscular endurance task. *Journal of Sport Psychology*, 4, 345-354.

Acknowledgements

Preparation for this article was supported by Economic and Social Research Council ESRC grant RES-000-27-0119. The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was also part of the programme of the ESRC Research Centre For Economic Learning and Human Evolution. The author wishes to thank Yousef Osman, David Shanks, Maarten Speekenbrink, Chris Berry, Belen Lopez, Andrea Smyth, Yana Weinstein, Joaquin Moris, David Lagnado, Bob Hausmann, Bjoern Meder, Momme von-Sydow, York Hagmayer, and Michael Waldmann for their inspired comments and encouragement.

Figure Captions.

Figure 1. Water tank system with inputs (salt, carbon, lime) and outputs (oxygenation, chlorine concentration, temperature). The CDCT in Figure 1 is from Burns and Vollmeyer's (2002) task, which was based on a water tank purification plant, and will be used in the present study.

Figure 2. Mean Error scores (\pm SE) at Test 1 and Test 1 for each condition in Experiment 1. Successful performance is indicated by *lower* mean error scores.

Figure 3. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 1. Successful performance is indicated by *higher* structure scores.

Figure 4. Mean Error scores (\pm SE) at Test 1 and Test 1 for each condition in Experiment 2. Successful performance is indicated by *lower* mean error scores.

Figure 5. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 2. Successful performance is indicated by *higher* structure scores.

Figure 6. Mean Error scores (\pm SE) at Test 1 and Test 1 for each condition in Experiment 3. Successful performance is indicated by *lower* mean error scores.

Figure 7. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 3. Successful performance is indicated by *higher* structure scores.

Figure 8. Mean number of inputs varied (\pm SE) for each block of each the learning phase by condition in Experiment 4

Figure 9. Mean Error scores (\pm SE) at Test 1 and Test 2 for each condition in Experiment 4. Successful performance is indicated by *lower* mean error scores.

Figure 10. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 4. Successful performance is indicated by *higher* structure scores.

Figure 1. Water tank system with inputs (salt, carbon, lime) and outputs (oxygenation, chlorine concentration, temperature). The CDCT in Figure 1 is from Burns and Vollmeyer's (2002) task, which was based on a water tank purification plant, and will be used in the present study.

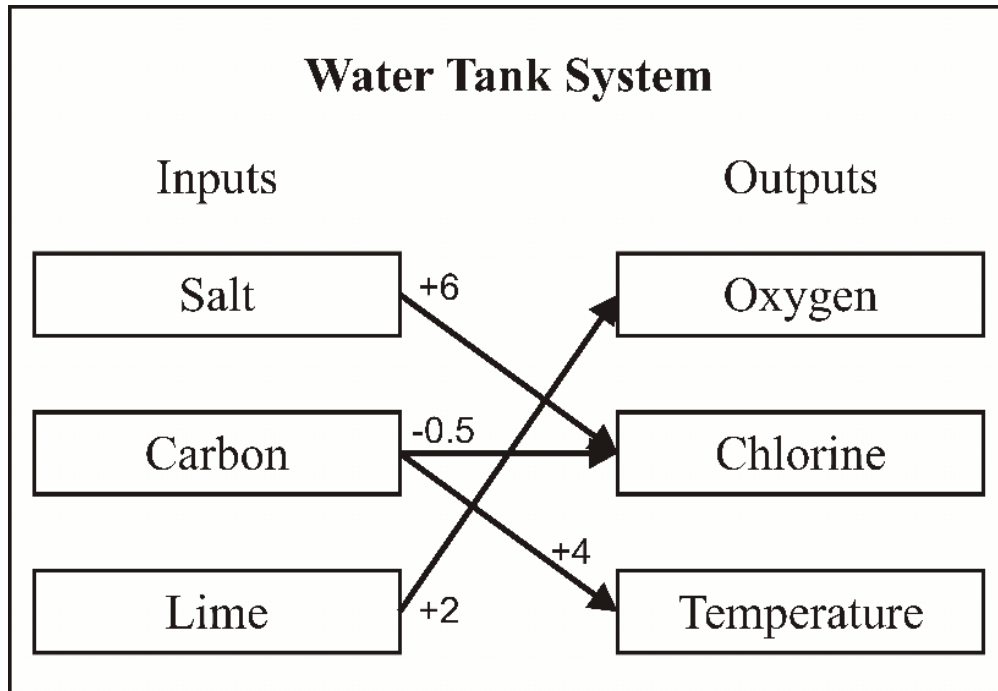


Figure 2. Mean Error scores (\pm SE) at Test 1 and Test 1 for each condition in Experiment

1. Successful performance is indicated by *lower* mean error scores.

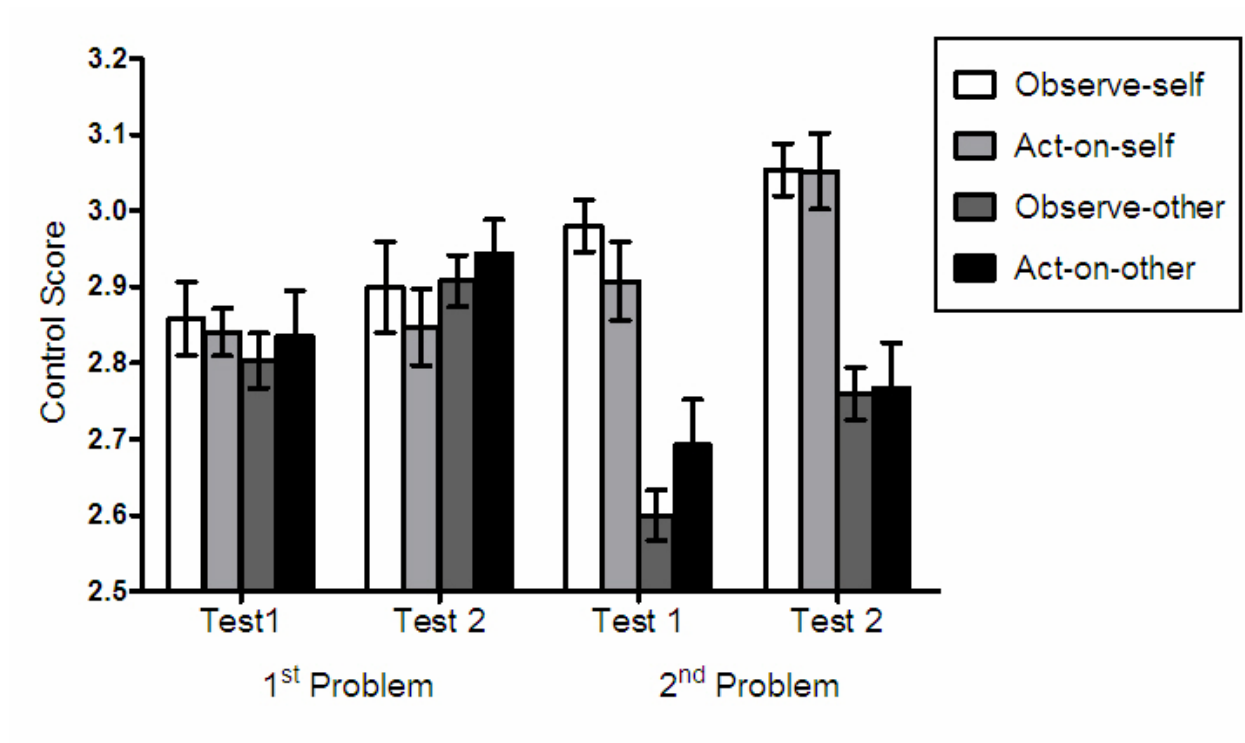


Figure 3. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 1. Successful performance is indicated by *higher* structure scores.

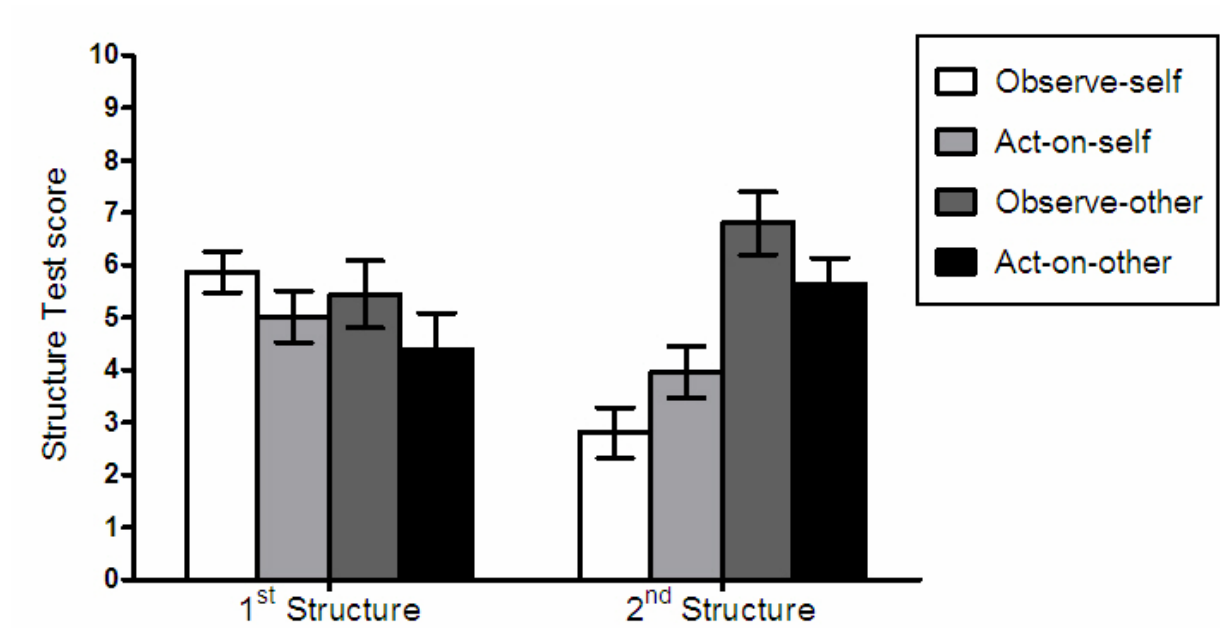


Figure 4. Mean Error scores (\pm SE) at Test 1 and Test 1 for each condition in Experiment 2. Successful performance is indicated by *lower* mean error scores.

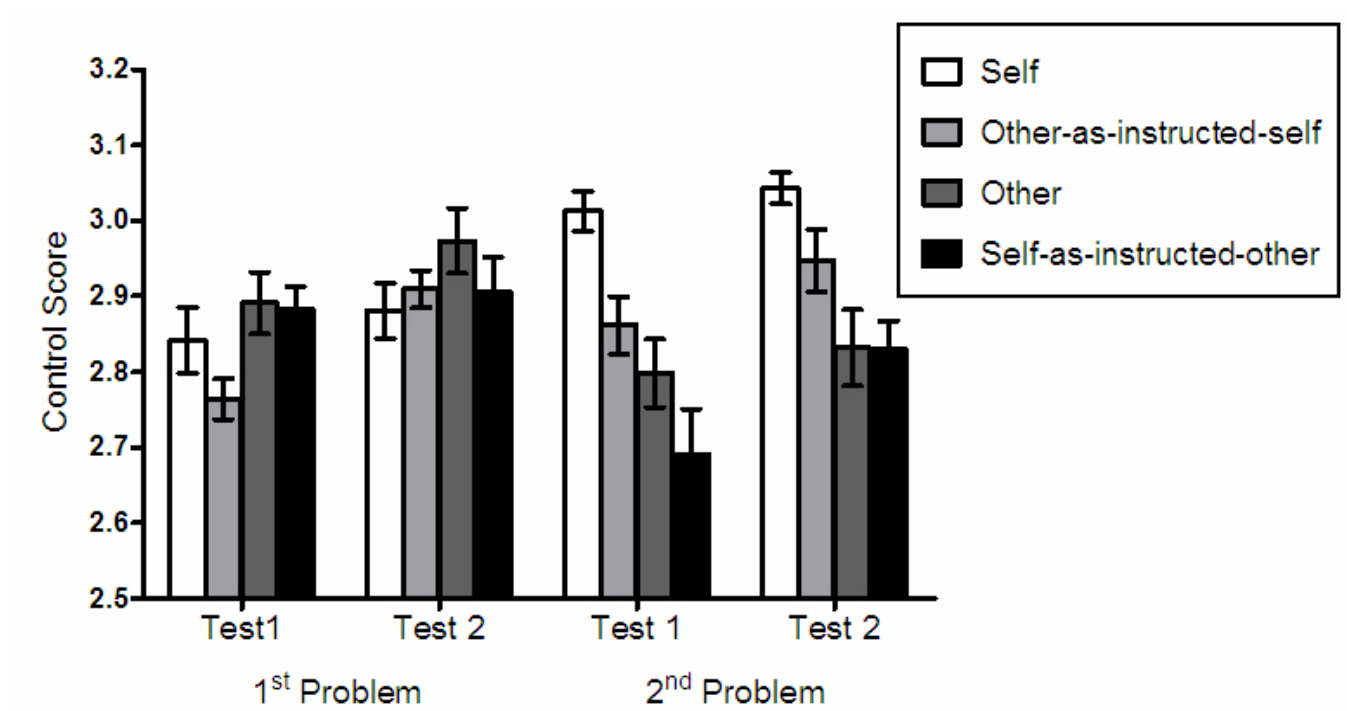


Figure 5. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 2. Successful performance is indicated by *higher* structure scores.

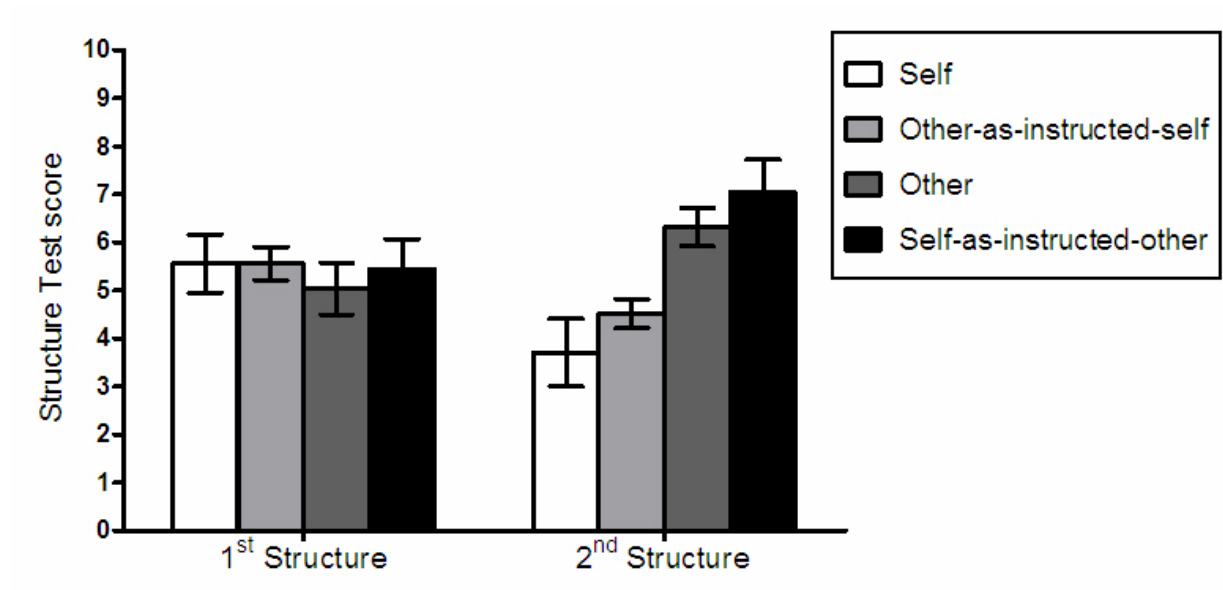


Figure 6. Mean Error scores (\pm SE) at Test 1 and Test 1 for each condition in Experiment 3. Successful performance is indicated by *lower* mean error scores.

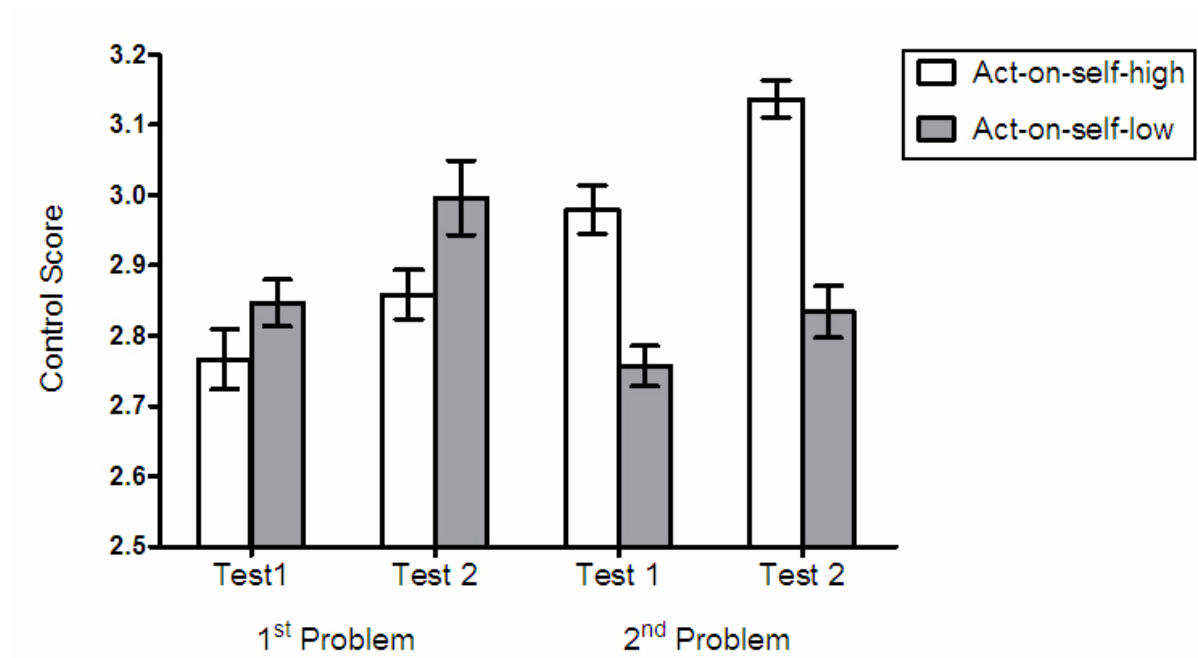


Figure 7. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 3. Successful performance is indicated by *higher* structure scores.

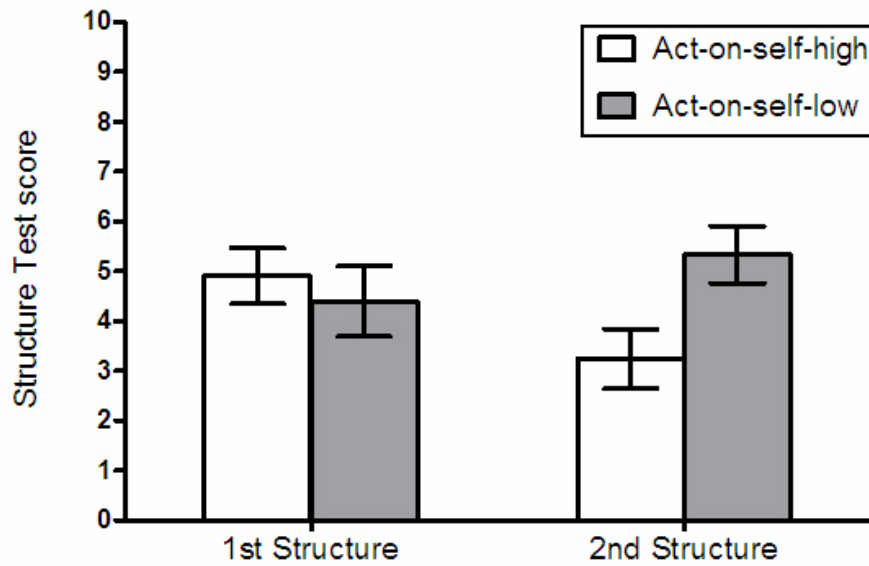


Figure 8. Mean number of inputs varied (\pm SE) for each block of each the learning phase by condition in Experiment 4

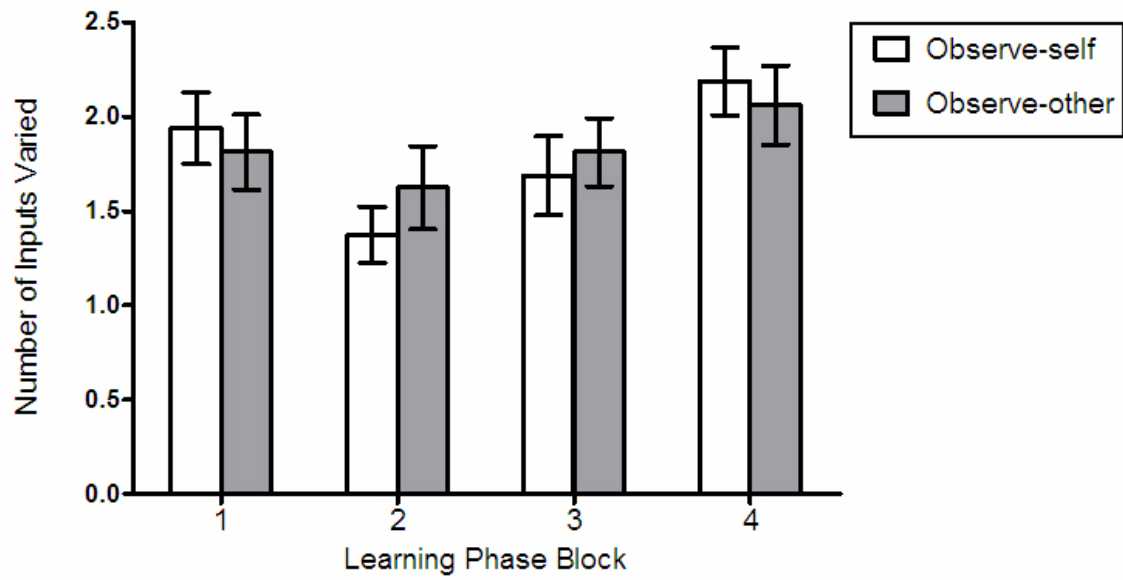


Figure 9. Mean Error scores (\pm SE) at Test 1 and Test 2 for each condition in Experiment 4. Successful performance is indicated by *lower* mean error scores.

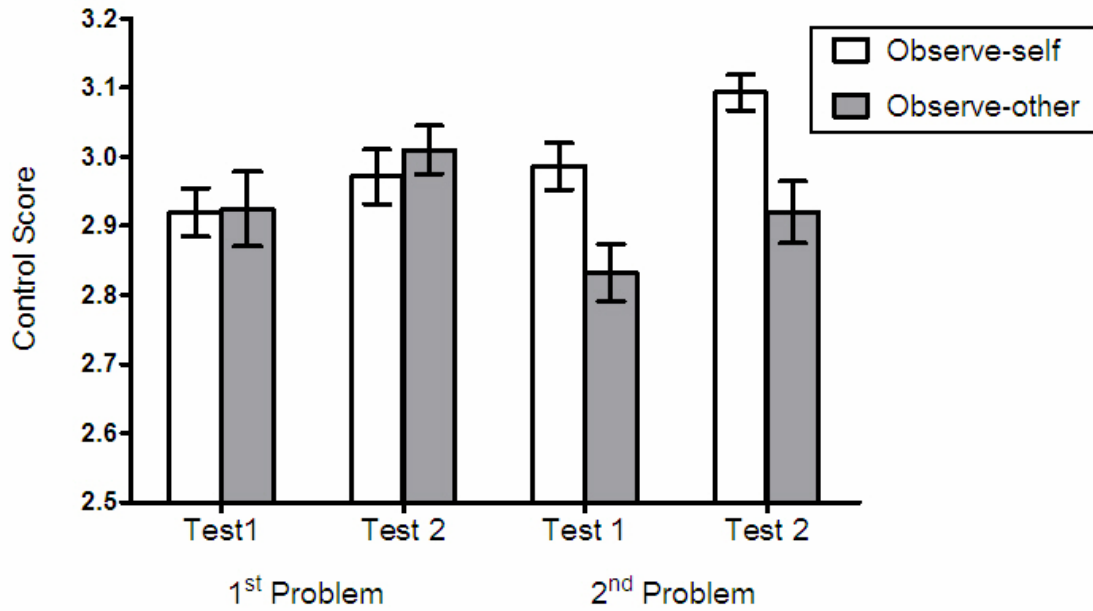
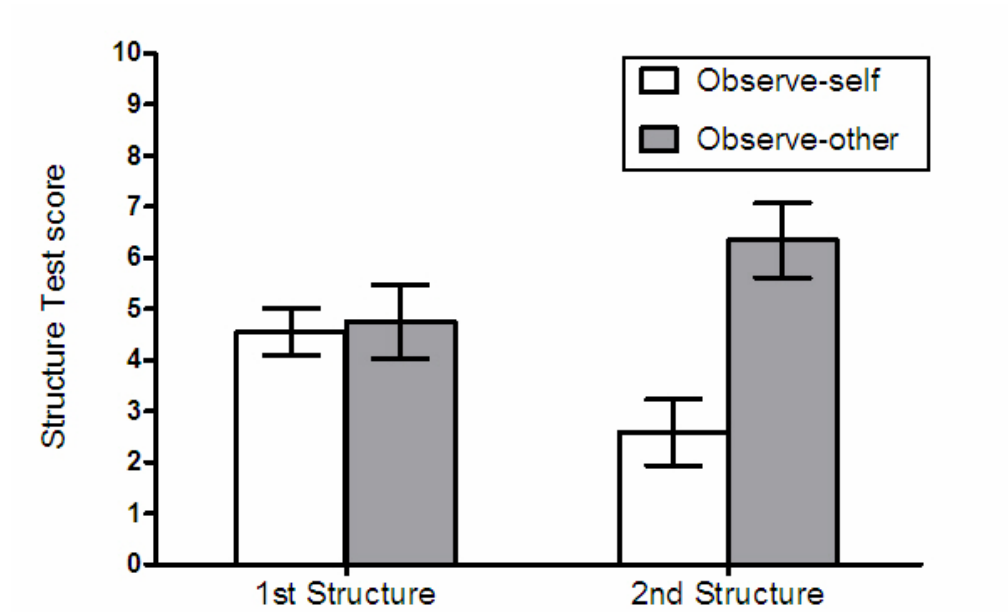


Figure 10. Structure scores (\pm SE) averaged across Structure Test 1, 2, 3, and 4 for each condition in Experiment 4. Successful performance is indicated by *higher* structure scores.



Appendix

Water Purification Tank Control System

Action instructions: You are a trainee laboratory technician working in a water filtration unit. As part of your training you will learn to control the water tank system by managing three water quality measures: Oxygenation; Chlorine CL concentration; Temperature. The quality measures are known as outputs and are used to monitor three system inputs: Salt; Carbon; Lime. In the following task you will be presented a total of 12 trials in which you will see a diagram of the 'Malwart' water filtration unit, which you will learn to control. You can modify the quality measures by manipulating the amount of Salt, Carbon, or Lime inputs; this can be done by moving the slider corresponding to the input either to the left or to the right. For each trial, you should try to change only one input; however this is only a recommendation and you may choose to use a different strategy. Once you have changed the value of an input you can then check the output levels by pressing the button labeled 'show me readings'; this will reveal the concentration levels of the quality measures. After you have studied these you should press the 'restart' button to begin the next trial. You should try and pay close attention to the values of the inputs you enter into the system and the output levels because this will help you to learn about the system. Good Luck!

Specific goal action instructions: *For each trial, you should try to change one input, but this is only a recommendation and you may chose to use a different strategy. Once you have done this you can check the output levels by pressing the button labeled 'show me readings'; this will reveal the concentration levels of the quality measures. After you have studied these you should press the 'restart' button to begin the next trial. Your task will be to change the output levels so that Oxygenation = 50, Chlorine CL Concentration = 700, Temperature = 900. Try to get as close to these levels as possible, and once you have done this try to maintain these levels throughout. Good Luck!*

Observation instructions: You are a trainee laboratory technician working in a water filtration unit. As part of your training you will learn to control the water tank system by managing three water quality measures: Oxygenation; Chlorine CL concentration; Temperature. The quality measures are known as outputs and are used to monitor three system inputs: Salt; Carbon; Lime. In the following task you will be presented with a series of trials in which you will see a diagram of the 'Malwart' water filtration unit, which you will learn to control. The system is set so that the quality measures change according to the values chosen by one of the workers of the water plant. You will see the amount of Salt, Carbon, and Lime inputs change automatically according to those set by the worker; this is indicated by a slider corresponding to each input moving either to the left or the right. You will see a total of 12 trials divided into two short sessions of 6 each. For each trial, you should watch carefully the changes to the inputs. When you have examined the changes to the inputs you can check the output levels by pressing the button labeled 'Output readings.' This will reveal the concentration levels of the quality measures. After you have studied these you should press the 'Input levels' button to begin the next trial. You should try and pay close attention to the values of the inputs that are entered and the output levels; this is because you will be required to imitate the worker's behavior later. Good Luck!

Specific Goal Observation Instructions: *For each trial, you should watch carefully the changes to the inputs. When you have examined the changes to the inputs you can check the output levels by pressing the button labeled 'Output readings.' This will reveal the concentration levels of the quality measures. After you have studied these you should press the 'Input levels' button to begin the next trial. For each trial, your task will be to assess how well the worker of the water plant successfully achieved the following output levels so that Oxygenation = 50, Chlorine CL Concentration = 700, Temperature = 900. Good Luck!*

Ghost hunting Control System

General Instructions: Newspaper Report: Hillside, NJ Investigations, Utah State Library
 Library worker John, his brother, and wife all reported seeing odd shadows out of the corner of their eyes. Most unusual was the report of the phone calls that came at 7:15 AM, certain mornings, that were riddled with static and no one on the other end of the call. The team of paranormal investigators went to investigate yesterday and was fully equipped with a Trifield meter, a Anemometer, a GGH meter. The investigation took place from 6:30 AM till approx 8:30 AM. Regular recordings were made. You were part of the team. You've done all the hard work and are back at the lab processing the data from the difference pieces of equipment you have used. Since you are new to this you aren't quite sure which of the three pieces of equipment (GGH meter, Anemometer, Trifield meter) actually registers air pressure, radio waves, and the electro magnetic field – which are all disrupted when a ghost is present.

Standard action Instructions: You have a total of 12 trials in which you can test the equipment by altering the values of the meters and examining the computer readout for each of the output values: air pressure, radio waves, and electro magnetic field. For each trial, you should try to change only one input; however this is only a recommendation and you may choose to use a different strategy. Once you have changed the value of an input you can then check the output readings levels by pressing the button labeled 'show me readings'; this will reveal the computer readings. After you have studied these you should press the 'restart' button to begin the next trial. You should try and pay close attention to the values you chose for the meters and the effects on the output readings. Good Luck!

Specific Goal Action Instructions: *For each trial, you should try to change one input, but this is only a recommendation and you may chose to use a different strategy. Once you have done this you can check the output levels by pressing the button labeled 'show me readings'; this will reveal the readouts of the phenomena. After you have studied these you should press the 'restart' button to begin the next trial. Your task will be to change the readouts so that Radio waves = 50, Electro Magnetic Field = 700, Air Pressure = 900. Try to get as close to these levels as possible, and once you have done this try to maintain these levels throughout. Good Luck!*

Standard Observation Instructions: You have a total of 12 trials in which you will observe the equipment being tested by one of your team. This will be done by altering the values of the meters and examining the computer readout for each of the output values of Air pressure, radio waves and electro magnetic field. You will be presented with the different levels of the meters and the values of the three output readings. For each trial, you should watch carefully the changes to the inputs. When you have examined the changes to the inputs you can check the output levels by pressing the button labeled 'Output readings.' This will reveal the computer output readings. After you have studied these you should press the 'Input levels' button to begin the next trial. You should try and pay close attention to the values that are chosen for the meters and the effects on the output readings. Good Luck!

Specific Goal Observation Instructions: *For each trial, you should watch carefully the changes to the inputs. When you have examined the changes to the inputs you can check the output levels by pressing the button labeled 'Output readings.' This will reveal the concentration levels of the quality measures. After you have studied these you should press the 'Input levels' button to begin the next trial. For each trial, your task will be to assess how well the worker of the water plant successfully achieved the following output levels so that Oxygenation = 50, Chlorine CL Concentration = 700, Temperature = 900. Good Luck!*