Strategies, Heuristics and Biases in Complex Problem Solving

Frederic Dandurand (fdandu@ego.psych.mcgill.ca)
Department of Psychology, McGill University
1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada

Thomas R. Shultz (thomas.shultz@mcgill.ca)
Department of Psychology and School of Computer Science, McGill University
1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada

Kristine H. Onishi (konishi@psych.mcgill.ca)
Department of Psychology, McGill University
1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada

Abstract

How do instructions help people solving complex puzzles? We studied a problem solving task (Which of the 12 balls is heavier or lighter than the rest?) using detailed analyses of problem solving steps to assess what cognitive biases, heuristics and strategies were used. First, we found that all participants effectively used means-ends analysis. Second, in the absence of instructions or observation of successful solutions, participants preferred symmetrical and overly simple solution steps. Instructions and imitation effectively reduced these biases, which was important for correct solutions. Finally, instructions and imitation helped participants attend to less salient aspects of the task.

Keywords: Problem solving; strategies; heuristics; biases; means-ends analysis; cognition.

Introduction

Problem solving is an important mental activity, and a classic research area in cognitive psychology. Problem solving has received less attention recently, and is still largely dominated by the information processing paradigm where problems are construed in terms of states, transitions, operators and constraints (Holyoak, 1995).

Several factors may constrain how humans solve problems. First, given their limited cognitive capacity, humans tend to perceive and represent stimuli in large, regular chunks. Gestalt psychology has studied these phenomena extensively. For example, human perception appears to have a symmetry bias (Freyd & Tversky, 1984). Second, humans tend to repeat solution processes that have been previously successful, a phenomena known as "problem solving set" (Glass & Holyoak, 1986). Third, humans sometimes have difficulty finding new ways to use familiar tools, a phenomenon called "functional fixedness" (Duncker, 1945). Finally, problems solvers may impose unnecessary constraints to the problem. Those "conceptual blocks" (Adams, 1974) may also limit their ability to explore promising areas of the problem space.

Here we measured some factors involved in solving a difficult problem solving task. On the one hand, we posited that participants use means-ends analysis to guide their solution steps. On the other hand, we identified three major blocks or sets during preliminary analyses that may limit participants' ability to explore promising areas of the problem space: a symmetry bias, a simplicity bias, and a tendency to ignore some important aspect of the task.

We used a complex puzzle involving the weighing of balls. Participants had to find, with three uses of a scale, the one ball that was either heavier or lighter than the rest of a set of 12 balls. Any combination of balls was allowed on the scale. We used this task because it is well-defined, complex enough to avoid performance ceiling effects, and because similar problems have been used in earlier research in psychology (Simmel, 1953).

We previously found that participants were significantly more accurate when they had access to relevant information (instructions or demonstrations) than when they were only told if their answers were correct (Dandurand, Bowen, & Shultz, 2004). Demonstrations consisted in five correctly solved instances of the problem (i.e., five randomly selected branches of the solution tree shown in Fig. 2). Learning by watching demonstrations is sometimes referred to as imitation learning. It is also similar to vicarious learning, except that participants do not see the mentor (here implemented as a computational agent) get explicit rewards for correct solutions. Participants who read instructions had equivalent information in written form. Clearly, instructions and demonstrations improved performance, but it remained unclear why and how.

Previous analyses of our problem solving task presented general trends and group averages. Here, we present detailed analyses of problem solving steps to find clues to cognitive mechanisms involved in solving the task. We look for those clues by quantifying important mental blocks and sets, both when demonstrations or instructions were available and when they were not. This shows how general concepts like blocks and sets can be operationalized and measured in complex problems. Simulations eventually will be required to identify mechanisms capable of exhibiting these blocks and sets, and other aspects of the data.

Ball Weighing Task

Task Analysis

As illustrated in Figure 1, participants need to alternate between two subtasks: (a) select which balls to put on the scale and (b) update the ball labels depending on results of the weighing. Labels reflect the information problem solvers have about possible weights of balls. Each ball could be labeled as follows: unknown (U), heavy or light weight (HL), heavy or normal weight (HN), light or normal weight (LN), heavy (H), light (L), or normal (N).

Initially, all balls are marked as Unknown. Participants alternated between selecting balls and updating labels until they either found the solution or used the three weighings and were forced to guess the target ball and its weight.

Optimal Problem Solution

There are 24 possible cases (12 balls x 2 weights, heavy/light) in the task and Figure 2 presents an ideal solution. For conciseness, two branches have been collapsed so 16 leaves are shown. The theoretical limit on the number of cases that can be discriminated using a balance with three outcome states (balanced, left heavier, right heavier) in three weighings is $3^3 = 27$. Therefore, the task is difficult because it is close to the theoretical limit. A general discussion of solutions to this class of problems is available in Halbeisen and Hungerbuhler (1995).

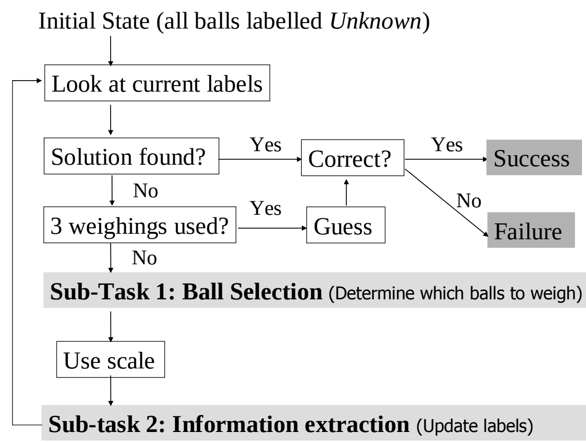

Figure 1: Task analysis of the ball weighing experiment

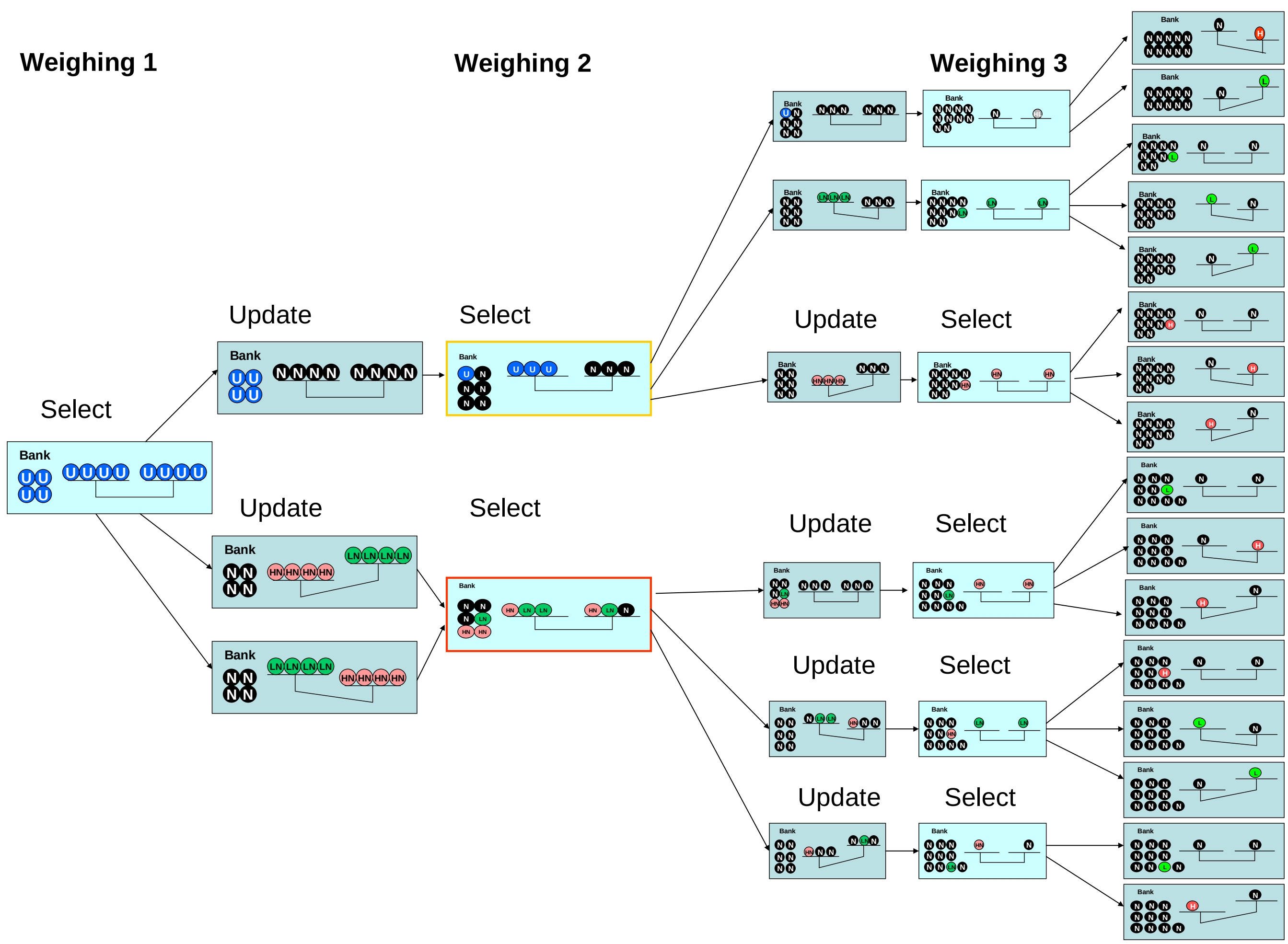

Figure 2: Optimal solution to ball weighing problem

Selections and updates are often complex and asymmetrical. In particular, the selection following the case where the scale tipped on the first weighing involves placing balls with different labels on the right and left sides of the scale. Also, updates have to be made to the balls in the bank in addition to those on the scale. Preliminary analyses suggested that participants may have the conceptual block of assuming that only one kind of item can be installed on the scale.

Methods

Participants

Sixty eight McGill undergraduate and graduate students participated and five were excluded for not finishing the warm-up task within 30 minutes (n = 3) or as statistical outliers on a q-q plot (n = 2). As we compared number of correct to number of error responses, we excluded participants who generated only one type of response (n = 8) to avoid missing cells.

Procedure

Participants were randomly assigned to a learning condition. *Reinforcement* group participants were told if their answers were correct or not. By contrast, prior to working on solving problems, *imitation* group participants watched five demonstrations of problems being successfully solved. Those in the *instruction* group studied written instructions on how to solve ball-weighing problems.

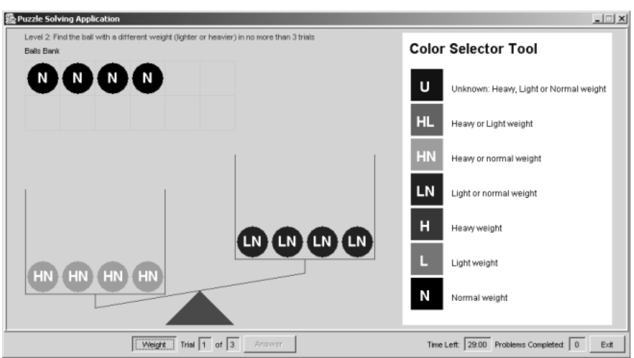

Figure 3: Screenshot of ball-weighing task

Participants worked on problem trials for 30 minutes, solving a mean of 18.6 trials (range: 6-37). They were told to label balls to reflect the information gained after each weighing. Figure 3 shows a screenshot of the Java program that presented the task and recorded participant data. Details of the experiment are available in Dandurand et al. (2004).

We collected the following information for each step taken:

- *Selection* subtask: how many balls with each label were installed on the right and on the left side of the scale?
- *Labeling* subtask: in each container (bank, left and right sides of the scale), how many balls of each label had their label updated?

To explore problem solving strategies, we (1) did a means-ends analysis, (2) measured bias including complexity and asymmetry of ball selection, and complexity of updates, and (3) assessed whether attention was paid to the bank.

Means-ends analysis

Based on the labels that participants assigned, we performed a means-ends analysis (Newell & Simon, 1972) in which we measured the distance between the current state and the end state. Table 1 indicates the distance of each label to the solution. The total distance was computed by summing the distances across the 12 balls. Thus the initial distance was 24 (12 balls as Unknown) and the final distance was 0.

Table 1: Distance to solution used for means-ends analysis

Label	*Distance*
Unknown (U)	2
Heavy or Light (HL)	1
Heavy or Normal (HN)	1
Light or Normal (LN)	1
Heavy (H)	0
Light (L)	0
Normal (N)	0

Bias measures

As we have seen in Figure 2, correct steps are often complex and asymmetrical. We computed measures of complexity and asymmetry for each of the sub-tasks.

For the selection subtask, we counted how many balls with each label were placed on each side of the scale and calculated indices of complexity and asymmetry based on the number of labels, not the number of specific items with each label.

To measure complexity, we summed the total number of labels present on each side of the scale. To measure asymmetry, we counted the total number of differences in labels between left and right sides of the scale, i.e., whenever a label was present on one side of the scale but not on the other, one unit of asymmetry was added. Table 2 shows examples of complexity and of asymmetry measures.

For the labeling subtask, complexity was calculated by summing how many balls with each label, in the bank and on each side of the scale, had their label updated. Thus, one unit of complexity was added whenever a given label in a given location (bank, or side of scale) was changed. Examples are given in Table 3.

Table 2: Example complexity and asymmetry index calculations for the Selection subtask

	Example	Index
Complexity	HN HN (1) vs. N N (1)	2
	HN LN LN (2) vs. HN LN N (3)	5
Asymmetry	HN HN vs. N N	2
	LN LN LN vs. LN LN LN	0
	HN LN LN vs. HN LN N	1

Table 3: Example complexity index calculations for the Labeling subtask

Labels before	Labels after	Index
Bank: U U U U Left: U U U U Right:U U U U	Bank: U U U U (0) Left: HN HN HN HN (1) Right: LN LN LN LN (1)	2
Bank: U U U U Left: U U U U Right: U U U U	Bank: N N N N (1) Left: HN HN HN HN (1) Right: LN LN LN LN (1)	3
Bank: HN HN LN N N N Left: HN LN LN Right: HN LN N	Bank: N N N N N N (2) Left: HN N N (1) Right: N LN N (1)	4

Attending to the bank

Finally, we computed an index of how often attention was paid to the bank by calculating how many label updates were made in the bank. For the examples in Table 3, the number of label updates in the bank was 0, 1 and 2, respectively.

Results

We separated correct and error trials because solutions leading to correct answers are likely to differ from those leading to errors in terms of strategies and biases. We performed three-way mixed ANOVAs with one independent factor (learning condition: reinforcement, imitation or instruction) and two repeated factors: correctness (correct or error response) and weighing number (1, 2 or 3). We focus here on effects significant at $p < 0.01$.

Figure 4 presents the ideal solution (for comparison) and the results of the means-ends analysis, that is, the distance to the goal state after weighings 1 and 2. Distance before weighing 1 was 24, and distance after weighing 3 was always 0 as participants were forced to answer. There were main effects of correctness, $F(1, 52) = 1894$, and weighing, $F(2, 51) = 795$, and an interaction between correctness and weighing, $F(2, 51) = 232$. Distance from the goal was smaller after each weighing, and trials leading to correct answers were closer to the goal and approached it faster than trials leading to erroneous answers.

Figure 5 presents results for the measure of selection complexity. There were main effects of learning condition, $F(2, 52) = 12$, and weighing, $F(2, 51) = 78$, and interactions between learning condition and weighing, $F(4, 104) = 8.1$. Participants in the instruction and imitation groups produced selections that were more complex than the reinforcement group, especially on weighing 2.

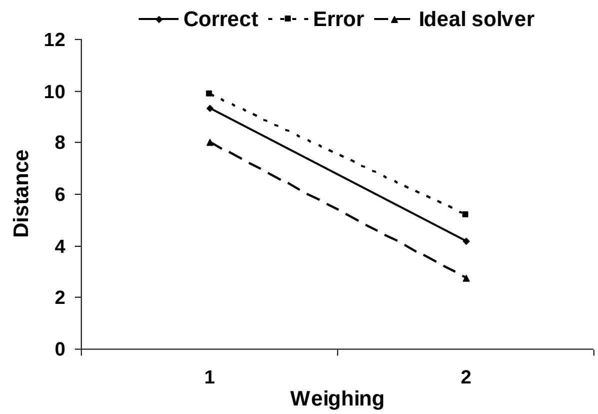

Figure 4: Means-ends analysis; distance to the goal as a function of answer correctness and weighing

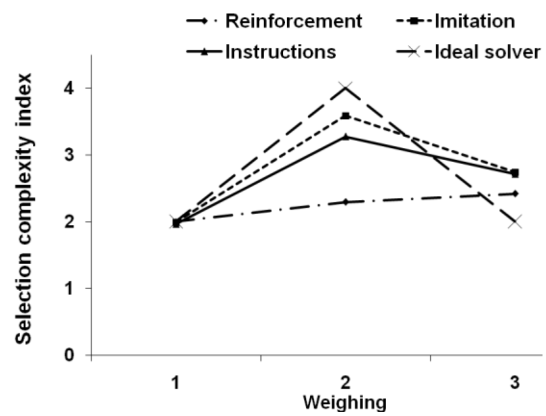

Figure 5: Selection complexity as a function of learning condition and weighing

Figure 6 presents results for the measure of selection asymmetry. There were main effects of correctness, $F(1, 52) = 70$, and weighing, $F(2, 51) = 11$, and an interaction between correctness and weighing, $F(2, 51) = 78$. Weighings 2 and 3 were more asymmetrical than weighing 1, and on weighing 3, trials leading to errors were more asymmetrical than those leading to correct answers.

Figure 7 presents results of the measure of labeling complexity. There was a main effect of weighing, $F(2, 51) = 205$, and an interaction between correctness and weighing, $F(2, 51) = 26$. Complexity was higher on weighings 2 and 3 than on weighing 1. It was also higher on trials leading to errors than trials leading to correct answers, but only on the third weighing.

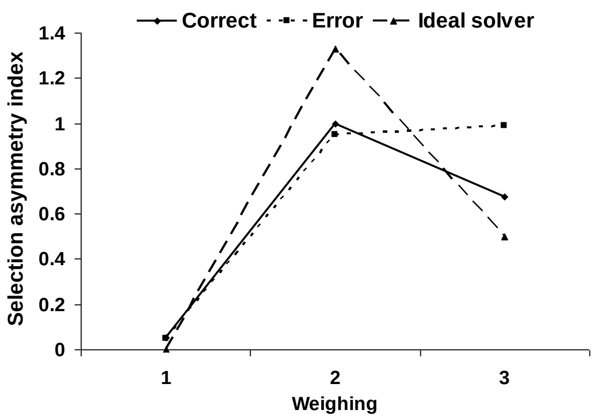

Figure 6: Selection asymmetry as a function of correctness and weighing

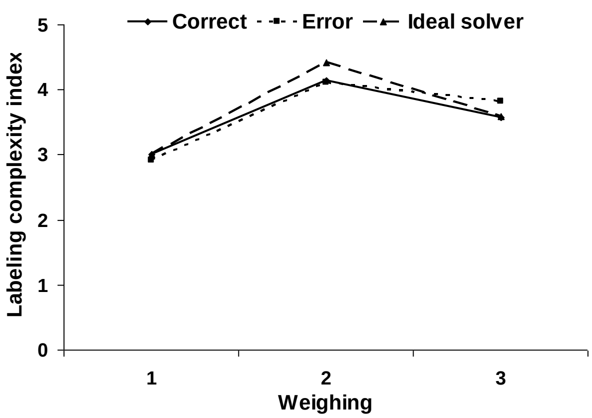

Figure 7: Labeling complexity as a function of correctness and weighing

Figure 8 presents results of the measure of bank attention. There was a main effect of weighing, $F(2, 51) = 78$. More bank updates were made on the third weighing.

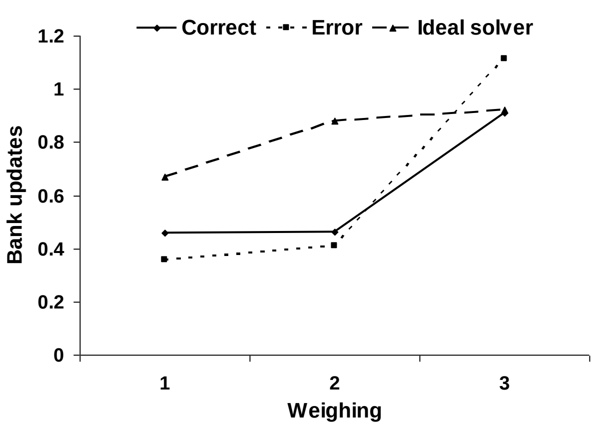

Figure 8: Bank attention index as a function of correctness

To sum up, more complex and asymmetrical steps were made on weighing 2. Selections were more complex, but not more asymmetrical in the imitation and instruction groups. Instructions did not appear to help participants attend to the bank.

Discussion

Participants in all experimental groups appear successful at reducing distance to solution across weighings, i.e., using means-ends analysis. However, reducing the distance to the goal does not necessarily yield a correct solution in three weighings. With sub-optimal solutions, participants can successfully exclude many possibilities (and thus get closer to the goal) but nevertheless have to guess among a few possible cases left after the third weighing. Only optimal solutions ensure that there is only one possibility (the correct solution) left at the end.

In ideal solutions (Figures 2, 4-9), the second weighing requires the most complex and asymmetrical steps. When weighing 2 is optimal, there are at most three not-fully-determined balls remaining for weighing 3, resulting in less complex and asymmetrical steps.

Compared to the optimal solution, participants use less complex and asymmetrical steps at weighing 2 and also pay less attention to the bank. By failing to pay attention to the bank, participants ignore information that may allow them to mark balls as normal thus excluding them from the possible solutions. However, bank neglect is not the only factor that explains the difference in Figure 9 between human and ideal solver performance. Participants told us that they sometimes gathered information about the weights of balls in the bank that they did not use to update ball labels.

Because participants have to give an answer after weighing 3, they may be forced to perform labeling operations that do not add any information. These forced operations inflate measures of complexity, asymmetry and attention to the bank for weighing 3, explaining why some of those measures are higher at weighing 3 for trials leading to erroneous answers than for trials leading to correct answers.

In short, two factors explain why human solutions were less than ideal. First, people tended to prefer simple and symmetrical selections and updates, whereas the optimal solution required more complex and asymmetrical choices. Second, important information can be inferred about items in the bank, but people sometimes ignored or failed to mark this information, perhaps because labeling balls in the bank requires inference. For example, if the scale tips, the target ball cannot be in the bank.

Why do participants show simplicity and symmetry biases? There is a long tradition of research about symmetry and simplicity biases in perception stemming back to Gestalt psychologists (Freyd & Tversky, 1984). Such perceptually grounded biases may impact cognition, for example, "…symmetry in an image allows it to be perceived and coded abstractly and economically." (Freyd

& Tversky, 1984, p. 112). In addition, the scale is physically symmetrical and participants may have had experiences in which choosing one side of the scale or the other did not matter. Finally, humans may prefer simple and symmetrical problem solving steps merely because they are cognitively easier.

However, despite evidence that they cannot correctly solve all problems and may have to revert to guessing, why do people persist in using simple and symmetrical solutions in the reinforcement condition? Aside from strong perceptual and perhaps conceptual biases, solutions can vary on many different dimensions besides simplicity and symmetry. Even when participants realize that their solutions do not correctly solve all problems, it can be hard for them to figure out which dimension to change. In fact, most participants first explore various possibilities, for example, the number of balls to weigh. We have previously reported that participants in the reinforcement learning group were the most likely to explore different combinations of balls on their first weighing (Dandurand et al., 2004). In most cases, participants would need to change many different dimensions to find a solution. This could be difficult given time and motivation constraints.

Furthermore, people are known to seek satisfactory, but not necessarily optimal solutions. Mean accuracy in the reinforcement group was close to 60%, suggesting that participants often had to guess between two possibilities and only rarely were able to uniquely identify the target. Given a chance accuracy of 4.2% (i.e., 1/24), this performance might be satisfactory to participants.

Giving participants demonstrations or instructions helps them do better (Dandurand et al., 2004) and here we showed that this improvement is due to (a) reduction of symmetry and simplicity biases, allowing participants to explore new areas of the solution space, and (b) paying more attention to the bank, maximizing the information gained from each weighing. In this paper, we have better characterized how demonstrations and instructions improve performance, but further research is still necessary to determine what cognitive mechanisms may explain this difference. An interesting follow-up question is whether this change is due to insight, or whether it can be explained by memorizing solution steps. Similarly, the increase in bank updates could be the result of insight, or could have simply been memorized or primed by instructions or demonstrations. One way to address these questions would be using Think Aloud Protocols to seek evidence for the cognitive processes participants might be engaging in: memory recall, insight, reasoning, understanding, etc. Computational modeling is another possible approach. If a cognitive mechanism is used, incorporating it in models should increase the fit to human data.

To sum up, the ball weighing problem is difficult because (a) the number of cases to discriminate (24) is close to the theoretical limit of the system (27), (b) solutions are counter-intuitive because they involve complex and asymmetric ball selections and updates, (c) solutions require attending to the bank, which is less common because it is less salient than the scale and requires inference, and (d) because solutions vary on many dimensions, the search space is large and cannot be exhaustively explored within the time allocated for the experiment (30 minutes).

This work suggests how abstract concepts such as blocks and sets may be operationalized and measured as specific cognitive biases.

Acknowledgments

We thank Melissa Bowen for her help collecting the data. The research was supported by a McGill Major scholarship to F. D. and a grant to T. R. S. from the Natural Sciences and Engineering Research Council of Canada.

References

Adams, J. L. (1974). *Conceptual blockbusting: A guide to better ideas*. Addison-Wesley: Reading, MA.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences, 21*, 667-721.

Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding “prior intentions” enables two-year-olds to imitatively learn a complex task. *Child Development, 73*(5), 1431-1441.

Dandurand, F., Bowen, M., & Shultz, T. R. (2004). Learning by imitation, reinforcement and verbal rules in problem solving tasks. *Proceedings of the Third International Conference on Development and Learning*, La Jolla, CA.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58:5.

Freyd, J., & Tversky, B. (1984). Force of symmetry in form perception. *American Journal of Psychology, 97*(1), 109-126.

Glass, A. L., & Holyoak, K. J. (1986). *Cognition, 2nd edition*. Random House: New York, NY.

Halbeisen, L., & Hungerbuhler, N. (1995). The general counterfeit coin problem. *Discrete Mathematics, 147*(1), 139-150.

Holyoak, K. J. (1995). Problem solving. In E. E. Smith & D. N. Osherson (Eds.) *Thinking: An Invitation to Cognitive Science, 2nd edition, Vol. 3* (pp. 267-296). MIT Press: Cambridge, MA.

Katona, G. (1940). *Organizing and memorizing*. Columbia University Press: New York, NY.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Prentice-Hall: Englewood Cliffs, NJ.

Tversky, B. (2005). Visuospatial reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 209-240). Cambridge University Press: New York, NY.