Memory & Cognition

Act first, think later: The presence and absence of inferential planning in problem-solving

Journal:	*Memory & Cognition*
Manuscript ID:	MC-ORIG-12-287.R1
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	n/a

Complete List of Authors:	Ormerod, Tom; Lancaster University, Psychology; MacGregor, James; University of Victoria, School of Public Administration Chronicle, Edward; University of Hawaii at Manoa, Psychology Dewald, Andrew; University of Hawaii at Manoa, Psychology Yu, Chun; University of Hawaii at Manoa, Psychology

Keywords:	problem solving

Short title: The presence and absence of planning

Act first, think later: The presence and absence of inferential planning in problem-solving

Thomas C. Ormerod

Lancaster University, UK

James N. MacGregor

University of Victoria, Canada

Edward P. Chronicle (deceased), Andrew D. Dewald and Yun Chu

University of Hawaii at Manoa, USA

Address for correspondence:

Thomas C. Ormerod

Department of Psychology

Lancaster University, LA1 4YF, UK

Tel. +44 (1524) 593164

Fax +44 (1524) 593744

t.ormerod@lancaster.ac.uk

Abstract

Planning is fundamental to successful problem-solving, yet individuals sometimes fail to plan even one step ahead when it lies within their competence to do so. This paper reports two experiments that explore variants of a ball-weighing puzzle, a problem that has only two steps yet nonetheless yields performance consistent with a failure to plan. The results fit a computational model in which a solver's attempts are determined by two heuristics: maximization of the apparent progress made towards the problem goal, and minimization of the problem space in which attempts are sought. The effectiveness of these heuristics is determined by lookahead, defined operationally as the number of steps evaluated in a planned move. Where move outcomes cannot be visualized but must be inferred, planning is constrained to the point where some individuals apply zero lookahead, which with *n*-Ball problems yields seemingly irrational unequal weighs. Applying general-purpose heuristics with or without lookahead accounts for a range of rational and irrational phenomena found with insight and non-insight problems.

Keywords

Planning; problem-solving; insight; planning failures; lookahead

Introduction

Planning is generally regarded as a prerequisite for successful cognitive performance (Newell & Simon, 1972; Hayes-Roth & Hayes-Roth, 1994; O'Hara & Payne 1998; Morris & Ward, 2005), and differentiates humans from many other species (Tomasello, Carpenter, Call, Behne, & Moll, 2005). By looking ahead mentally from a current state to anticipate a new state, an individual can evaluate the likely success of a move sequence before its execution. Planning can reduce effort and avoid potentially costly errors or irreversible commitments. As implemented by VanLehn (1989) in terms of lookahead (the number of steps/moves that an individual plans ahead mentally), planning has been shown to be an important moderator of performance in models of problem-solving (e.g., MacGregor, Ormerod & Chronicle, 2001; Jones, 2003).

While failures of planning have been implicated in a number of neuro-degenerative diseases (e.g., Stuss & Alexander, 2007), healthy individuals who possess the necessary cognitive resources (e.g., working memory capacity, relevant domain knowledge, reasoning skills) may also fail to plan when it is in their interests to do so. Some major disasters may be attributed to failures of planning (e.g., Fukushima: Cooper, 2011). Mundane examples of planning absence arise in everyday life (e.g., failing to warm the oven in advance of starting to cook a meal). A failure to plan can give rise to behaviours that in retrospect appear irrational, like sitting on a branch of a tree while you saw it off.

Over the last forty years, evidence has amassed that human performance deviates from what is normatively rational (Kahneman & Tversky, 1972; Kahneman & Tversky, 2000) and planning failures, in the form of impulsivity, have been implicated as a root cause (Kahneman, 2003). However, some have argued that what may appear irrational from one perspective may be rational from another (Gigerenzer & Goldstein, 1996; Gigerenzer & Gaissmaier, 2011), one example being the information gain explanation of choices in the Wason selection task. Under this criterion, it is argued, the common "wrong" responses become rational (Oaksford & Chater, 1994, 2003). Recently, an information gain explanation has been extended to choices in a weighing task, where the goal was to identify an underweight member in a set of otherwise identical objects (Wakebe, Sato, Watamura, & Takano, 2012). The example recalled a different weighing task (Simmel, 1953), where apparently irrational choices do not appear to be readily explained by information gain. Here we propose and test an explanation of the behaviour using a model of problem solving which accounts for the frequent "irrational" response, frequent "rational" but incorrect responses, and the correct response. The model represents a further extension of the Criterion of Satisfactory Progress theory (CSP), previously applied to insight problem solving (MacGregor et al, 2001; Ormerod, MacGregor, & Chronicle, 2002; Chronicle, MacGregor, & Ormerod, 2004).

The introduction proceeds with a description of the *n*-Ball problem, then provides an explanation of CSP, illustrated using the nine-dot problem, followed by the application of CSP to the *n*-Ball problem.

The *n*-Ball problem

Simmel (1953) [1] examined performance on 8- and 9-ball variants of the *n*-Ball problem (Simmel referred to coins rather than balls, but the principles are identical). The *n*-Ball problem is as follows. You have *n* balls, which look identical. One is slightly

[1] We are grateful to Stellan Ohlsson for pointing us towards this paper.

heavier than the others, but the difference cannot be discerned by picking up each ball. You have a balance scale, and you may use it only twice. How can you find the heavy ball? Figure 1 demonstrates how the solution is found with *n*=7, 8 and 9. For each variant, the correct first move is to weigh any three balls against any other three. There is also an alternative solution for *n*=7, which initially involves weighing any two against any other[2].

Insert Figure 1 about here

For the 8-ball problem, Simmel reported that 78% of 58 participants selected an initial weigh of 4v4, while 22% selected another symmetrical weigh. For the 9-ball problem, the most frequent initial weighs were 4v4 (42%) and 5v4 (37%). The 5v4 weighs seem irrational: a weigh where the number on each side of the balance is unequal (referred to here as an “unequal weigh”) cannot yield usable information since the additional ball in the lower pan necessarily masks the presence of a slightly heavier ball. It seems to be a move made without thought, yet it is one made by over a third of Simmel’s university student participants.

Simmel explained the “irrational” weigh from the coalescing of two independent tendencies; “totality”, to maximize the number of balls weighed; and “symmetry”, to make a balanced comparison. For the 8-ball problem, the two tendencies are mutually compatible, leading to a 4v4 comparison. For the 9-ball problem the tendencies conflict,

[2] The alternative solution to the 7-ball variant is to weigh two balls against two others, leaving three unweighed. If the scales balance, then select two from the three unweighed balls and weigh 1 against 1 on the second weigh, else select the two balls from the lower scale and weigh 1 against 1.

and lead to an imperfect outcome.

Simmel's proposed tendency of symmetry applies only to problems that allow move selection on the basis of geometric properties, and may be viewed as an example of a problem-specific heuristic. A similar 'balance' tendency was proposed as one of three heuristics by Simon and Hayes (1976) to account for performance on the Missionaries and Cannibals problem, used when the general heuristics of means-ends analysis fails to generate a new move. Other problem-specific heuristics have been proposed for the mutilated checkerboard (Kaplan & Simon, 1990), and Rings (Kotovsky & Simon, 1990) problems, and Sudoku (Lee, Goodwin, & Johnson-Laird, 2008).

Problem-specific heuristics can provide a good fit to data, but they suffer three problems as models of problem-solving performance. First, they lack theoretical parsimony, since the set of heuristics must be extended for each new class of problem. Second, the principles for switching between heuristics across move attempts tend to be arbitrary and/or problem-specific themselves. Third, and critically for the current paper, problem-specific heuristics tend to be too powerful, in that they predict rational move selection under evaluation. For example, competition between totality and symmetry tendencies ought to preclude the selection of 5v4 weighs in the 9-ball problem, yet in Simmel's data these account for over a third of all first attempts. Similarly, participants often make imbalanced move attempts (e.g., sending two cannibals across the river as their first move) before exhausting the set of balanced moves (Greeno, 1974).

As an alternative to problem-specific heuristics, the present paper proposes and tests a general approach to explaining behavior in the *n*-Ball task, including the selection of "irrational" unequal weighs.

Criterion for Satisfactory Progress theory

In seeking a problem-general account of heuristics selection we have proposed a criterion for satisfactory progress (CSP) theory that characterizes both insight and non-insight problem-solving in terms of generic cognitive processes (MacGregor et al, 2001; Ormerod, MacGregor, & Chronicle, 2002; Chronicle, MacGregor, & Ormerod, 2004). CSP proposes that problem-solvers apply two general heuristics to novel problems; maximization and minimization.

Maximization heuristic

The *maximization* heuristic operates to sample moves which appear to maximize progress towards a goal, and the degree of progress is subsequently evaluated against a criterion of progress derived from the problem statement. Throughout a problem-solving attempt, moves are sampled from the same problem representation if they make adequate progress as judged by a criterion of satisfactory progress. For example, with the classic nine-dot problem, described below, the initial problem representation may be one which considers only lines drawn within the limits of the initial dot array, and this representation is maintained so long as moves continue to make satisfactory progress.

With insight problems, maximization gives rise to impasse, when no moves can be found within the current representation to meet the criterion. The maximization heuristic is effectively an instantiation of hill-climbing (Newell & Simon, 1972) but with the addition of a criterion below which moves are not selected even if they appear to make some progress.

To illustrate the operation of maximization, consider the nine-dot problem.

Typical naïve attempts show that people aim to cancel as many dots as possible with each line, for example by using the first three available lines to draw around three sides of the nine-dot figure (see Figure 2). This maximizing heuristic meets a criterion of progress – that, on average, 9/4 dots must be cancelled with each line – until the fourth and final line is considered. The success of the maximizing heuristic for the first three lines seems to be compelling: many unsuccessful rotationally symmetric attempts often ensue, and a state of impasse is thus reached.

Minimization heuristic

The operation of the maximization heuristic explains why the nine-dot problem is so resistant to solution, yet people do occasionally solve it, and they are able to solve variants where the first line is given, illustrated in Figure 2 (Weisberg & Alba, 1981; MacGregor et al, 2001). Maximization operates in CSP theory to enable choice of moves from a given representation. A second heuristic, *minimization,* operates to create and change mental representations of problems, and changing the representation of the problem is crucial to eventual solution (Newell & Simon, 1972). Minimization dictates that individuals limit the initial representation and subsequent expansion of the problem space to the minimum required to allow search for moves that might meet a criterion for satisfactory progress. Like maximization, minimization is a general heuristic that impacts on other tasks such as deductive reasoning (e.g., Ormerod & Richardson, 2003).

With the nine-dot problem, minimization constrains the initial problem space to the dot array. This representation allows the discovery of moves that meet the criterion for satisfactory progress but fails to allow a solution. Eventually, individuals exhaust all the criterion-meeting moves and relax the minimization heuristic, which allows discovery of a different problem space. For the nine-dot problem, this includes the space around the dot array, as well as other possibilities that may lead to illegal moves.

Insert Figure 2 about here

Lookahead in CSP

In applying CSP to the nine-dot problem we proposed that different participants employ different levels of lookahead (MacGregor et al, 2001).. A lookahead of one encompasses the selection of one line and its evaluation against the criterion, while two-lookahead comprises selection and evaluation of two successive lines (as illustrated in 2c and d). By comparing observed performance with CSP predictions, MacGregor et al estimated the proportion of participants using lookahead of 1, 2, 3 and 4 to be 32%, 32%, 36% and 0%, respectively.

Move selection and move evaluation involve different theoretical processes. Move selection utilizes maximization and minimization while move evaluation involves comparison against a criterion. The distinction makes it possible to have selection without evaluation, which we define here as a lookahead of zero. A lookahead of zero means the automatic selection of a maximizing move without evaluation against the criterion. With the nine-dot problem, first moves under zero- and one lookahead would be indistinguishable, since both would result in a single line through three dots.

In applying lookahead in the nine-dot problem, we assumed that the recognition of a line that intersects the maximum number of dots is immediately apparent and requires little or no cognitive processing. As a result, finding the best move requires no horizontal search (breadth) through the problem representation, leaving all lookahead capacity to search vertically through subsequent moves (depth). For the *n*-ball problem, however, the situation is different. The outcome of weighing different subsets of balls is not likely to be perceptually given, making it possible that more than one mental weigh has to be considered before a satisfactory one is found for the first weigh. That is, lookahead may be applied to search the problem space in breadth first, before being applied in depth. Also, we anticipate zero and one lookahead to arise more frequently than two and three lookahead in the *n*-Ball problem because the effects of weighs cannot be visualized; they must be inferred. Moreover, application of zero lookahead with the *n*-Ball problem leads to qualitatively different move selections than application of one-lookahead, as we outline below.

Applying CSP to the *n*-Ball problem.

For the *n*-Ball problem, we define a maximizing weigh as one that maximizes the number of balls in each pan. We propose that weighs are preference ranked in descending order of maximization. Since one ball must be isolated after two weighs, *n*-1 balls must be eliminated in two weighs, giving a criterion of progress of an average of (*n*-1)/2 balls eliminated with each weigh. We assume that if more than one planned weigh meets the criterion, selection occurs on the basis of chance. An alternative assumption is that the probability of each weigh being selected is proportional to the expected number of balls it eliminates. As illustrated below, the two assumptions lead to highly similar predictions, and so we have retained the simpler of the two.

Process model

The model and its application to the *n*-Ball problem is summarized in Table 1. The upper panel of the table presents seven model assumptions, while the lower panel provides predictions for the first weighs selected under these assumptions. The analysis is applied to 9-, 8-, and 7-ball problems.

Insert Table 1 about here

Four levels of lookahead are considered and, even at the highest level, the process does not reach the second of the two weighs allowed in the task. Although in principle the analysis may be extended to include the second weigh, we considered that lookahead of greater than three will be relatively rare, and we have limited the analysis accordingly.

Here we illustrate the process using the 9-ball problem. Under zero-lookahead, a participant simply selects and places the maximum number of balls possible in each pan without evaluation. Effectively, a participant at zero lookahead is maximizing under trial and error, resulting in a first weigh of 5v4.

Under one-lookahead, a participant evaluates the outcome of the sampled weigh prior to selecting it, considering both balanced and unbalanced possibilities. For the 9-ball problem, the unequal 5v4 weigh is first sampled and mentally tested, revealing an unbalanced outcome that eliminates no balls. The weigh is rejected and the next weigh in order of decreasing maximization, 4v4, is sampled. However, at this point lookahead is exhausted, so the 4v4 weigh is selected without testing (Assumption 7). Also, because an unequal weigh has been sampled and rejected, all further unequal weighs drop from the maximization ranking for this participant, under Assumption 4.

Under two-lookahead, a 5v4 weigh is sampled, tested, and rejected. Next, a 4v4 weigh is sampled and tested. If the outcome is balanced, all of the 8 balls weighed are eliminated, while if it is unbalanced, the four balls on the lighter side plus the unweighed ball are eliminated. Both possible outcomes meet the criterion of eliminating (9-1)/2 balls. This exhausts lookahead, and the 4v4 weigh is selected, under Assumption 5.

Under three-lookahead, a 5v4 weigh is sampled, tested, and rejected. Next, a 4v4 weigh is sampled, tested, and found to eliminate 8 balls if balanced and 5 balls if unbalanced, in both cases meeting the criterion. With the remaining lookahead, one more weigh is sampled and tested. Under Assumption 4, a 4v3 weigh is no longer eligible, so 3v3 is sampled. Balanced and unbalanced outcomes will each eliminate 6 balls, meeting the criterion. Under Assumption 6, 3v3 and 4v4 weighs are selected with equal likelihood. An alternative assumption is that selection is proportional to the expected number of balls eliminated. In this case, the expected number of balls eliminated is 5.33 for the 4v4 weigh (i.e. 8 x 0.11 + 5 x 0.89) and 6 for the 3v3 weigh (i.e. 6 x 0.33 + 6 x .67). This results in a 47% selection rate for the former and 53% for the latter, both of which are close to the 50% proposed by Assumption 6.

Application of the model to 8-ball and 7-ball problems is essentially the same as for the 9-ball. However, there are some details relevant to some of the predictions described later. For the 8-ball problem at one-lookahead, the 4v4 weigh is sampled first and tested (see Table 1). Because the heavy ball is included in the weigh, the outcome is necessarily unbalanced and eliminates 4 balls. The criterion is met, lookahead is exhausted, and the weigh is selected (Assumption 5). Of note, in the 8-ball and 9-ball versions, first weighs selected at one-lookahead are physically the same but have a different status. In the 8-ball version the 4v4 weigh has been tested, while in the 9-ball version, it has not, which places the 8-ball problem one step closer to solution than the 9-ball for the same level of lookahead. The point is germane to one of the predictions developed below.

For the 8-ball problem at three-lookahead, a 4v4 weigh is first sampled, tested and found to meet the criterion. Next, a 4v3 weigh is sampled, tested, and found to be uninformative. Then, a 3v3 weigh is selected and tested. A balanced outcome eliminates all 6 balls weighed while an unbalanced outcome eliminates 5 (the 3 on the lighter side plus the 2 unweighed). This exhausts lookahead and the 4v4 weigh and 3v3 weigh are assumed to be selected with equal probability. However, if selection was proportional to the expected number of balls eliminated, then 43% of selections would favour the 4v4 weigh, 57% the 3v3, again relatively close to the 50% proposed by Assumption 5.

Finally, at three-lookahead in the 7-ball problem, a 4v3 weigh is sampled, tested and rejected. A 3v3 weigh is next sampled, tested, and found to meet the criterion, by eliminating all 6 weighed balls if balanced and 4 if unbalanced. Since unequal weighs were eliminated by the 4v3 weigh, the final unit of lookahead samples and tests the 2v2 weigh. If balanced, this eliminates 4 balls, and if unbalanced, eliminates 5, thereby meeting the criterion. The 3v3 and 2v2 weighs are therefore equally preferred. (Based on the expected number of balls eliminated, preference weightings would be 48% and 52%, respectively.) Of note, both the 3v3 and 2v2 weighs are on a correct solution path, opening up the possibility that participants may find two different solutions to the 7-ball problem.

Using the model to estimate lookahead

Simmel's (1953) study reported the frequencies of initial weighs for the 9- and 8-ball problems, and provides a basis for estimating the likely distribution of zero-, one-, two-, and three-lookahead in a participant sample. To do so, let w, x, y, and z be the proportion of participants operating with zero-, one-, two-, and three-lookahead, respectively.

According to our analysis of the 9-ball problem in Table 1, the only participants who select a 5v4 weigh are those operating at zero-lookahead. Thus, the proportion of participants selecting 5v4 as a first weigh provides an estimate of w, the proportion applying zero lookahead. From Simmel's Table IV (9-ball first condition) the proportion selecting the 5v4 weigh was 37%, from which we estimate that w=37%.

Similarly, the analysis for the 9-ball problem in Table 1 indicates that the proportion selecting a 3v3 first weigh will consist of half of those using three-lookahead, or $0.5z$. The results shown in Simmel's Table IV indicate that 10.5% selected the 3v3 weigh, from which we estimate that the proportion operating at three-lookahead, z, was 21%.

Finally, from Table 1, those selecting a 4v4 first weigh will consist of those using one-lookahead, those using two-lookahead, and half of those using three-lookahead , or $x + y + 0.5z$. From Simmel's Table IV the proportion selecting 4v4 weighs was 42%, from which we may estimate that the proportion of one- and two-lookahead combined, $x + y$, was 31.5% (42%-10.5%).

The remaining 10.5% of Simmel's participants selected a 1v1 weigh, which we will treat as belonging to an "Other" weigh category. A 1v1 weigh has a possibility of solving the problem in one weigh if it includes the heavy ball, although it cannot guarantee solving in two weighs. Its selection may therefore represent another form of maximization under zero-lookahead (maximizing the speed of solving, without testing that it guarantees a solution).

Applying the model and estimates to predict Simmel's 8-ball results

As an initial test, we may use the estimates derived above from Simmel's 9-ball experiment to predict performance in her 8-ball conditions. From Table 1, for the 8-ball problem, the proportion selecting a 4v4 weigh will be all of those operating at zero-, one-, and two-lookahead plus half of those at three-lookahead, , or $w + x + y + 0.5z$, while those selecting a 3v3 first weigh, will be half of those operating at three-lookahead, or $0.5z$. Using the estimates of w, $x + y$, and z, above, the predicted proportions are therefore 37%+ 31.5% + 10.5%, or 79%, selecting a 4v4 weigh, 10.5% selecting 3v3, with the remaining 10.5% selecting "Other".

From Simmel's Table II (8-ball problem only, and 8-ball problem first conditions combined, n= 39), the observed proportions selecting 4v4 and 3v3 were 87% and 10.3%, respectively. Given that the estimates of lookahead were based on only 19 participants, the predicted results for the 8-ball appear encouraging.

Later in the article, we use these estimates of lookahead to test predictions concerning the relative distribution of first weighs. In addition, the CSP model leads to several other predictions. First, the 7-ball problem will be easiest to solve because the predicted most common first weigh, 3v3, lies on a correct solution path. For the 8-ball and 9-ball problems, the most common first weighs are not on the solution path.

Second, the 8-ball version will be easier to solve than the 9-ball version. The rationale concerns both zero- and one-lookahead. At zero-lookahead, the predicted first weigh in the 9-ball problem, of 5v4, is two steps away from a weigh of 3v3, the first weigh on the correct solution path. The corresponding first weigh for the 8-ball problem, of 4v4, is only one step away. At one-lookahead, while a 4v4 weigh is predicted for both 9-ball and 8-ball versions, in the former, the weigh is selected without testing, whereas in the latter, it has been tested. The solution process in the 8-ball version is therefore one operation ahead of the 9-ball version.

Third, of the two correct solutions to the 7-ball problem, the solution with a 3v3 first weigh will occur more frequently than the solution with a 2v2 first weigh. This is because a 3v3 weigh is selected by all of those operating at one- and two-lookahead and by half of those at three-lookahead (an estimated 79%) . In contrast, a 2x2 weigh is selected by only half of those using three-lookahead (an estimated 10.5%).

Below we present two experiments examining human performance with *n*-Ball problems. Experiment 1 tested predictions 1 through 3, above, using the 7-ball, 8-ball and 9-ball problems. It also examined the frequency of selection of first weighs to test the predictions from Table 1. Experiment 2 examined detailed performance across 10 trials of 7- and 8-ball problems, to test predictions based on the proposed minimal expansion of the problem space.

Experiment 1

Experiment 1 tested several of CSP's predictions about *n*-Ball performance: first, that the 7-ball problem will be simpler than 8- or 9-ball problems; second, that the 8-Ball problem will be simpler than the 9-ball problem; and third, that the 3v3 solution to the 7-

ball problem will be more frequent than the 2v2 solution. Further, the experiment allowed tests of the theory's detailed predictions about weighing frequencies. The derivation of these detailed predictions is reviewed below.

Our estimates of the relative frequencies of participants operating under each lookahead were 37% at zero-lookahead, 31.5% at one- and two-lookahead combined, and 21% at three-lookahead. (The remaining 10.5% is considered to result in the selection of weighs in the "Other" category.) Applying these estimates to the sequence of events predicted in Table 1 allows us to predict the proportion of first weighs in the 9-ball, 8-ball and 7-ball problems. For the 9-ball, the resulting predictions are necessarily similar to Simmel's (1953) results, since the estimates were derived from Simmel's 9-ball condition. However, the predictions for the 8-ball and 7-ball problems are independent of Simmel's results. To illustrate, the predictions for the 7-ball problem are that 37% of first weighs will be 4v3 (all of those employing zero-lookahead), 42% will be 3v3 (all of those operating at one- and two-lookahead plus one-half of those at three-lookahead), 10.5% will be 2v2 (one-half of those operating at three-lookahead), and 10.5% will be "Other" (the remaining percentage). The CSP model's predictions for the frequencies of weigh selections in the 9-ball, 8-ball and 7-ball problems are summarized in Table 2.

Because a higher frequency is predicted to start on a correct solution path in the 7-ball than in either the 9-ball or 8-ball problems, a further prediction is that the 7-ball problem will be the simplest to solve of the three problems. Further, because more participants in the 8-ball than in the 9-ball problem are predicted to start only one step away from the correct solution path, with 79% choosing a 4v4 weigh in the 8-ball compared with 42% in the 9-ball, the 8-ball problem is predicted to be the simpler of the two. That is, solution rates should follow a pattern of 7-ball > 8-ball > 9-ball.

Method

Participants.

Eighty unpaid undergraduate students volunteered to participate (identifiers were not collected, so age and gender are unknown). Twenty-six attempted the 7-ball problem, 28 attempted the 8-ball problem, and 26 attempted the 9-ball problem.

Materials.

Participants received a booklet containing problem and information sheets as a function of condition, with the following problem instructions: "You have {seven; eight; nine – according to condition} balls that look identical. However, one is slightly heavier than the others (but the difference is too small to detect just by picking them up). Your task it to find out which one is heavier. You have a balance-scale, and you can use it only twice." Instructions were followed by a drawing of a balance scale, on which participants were asked to draw the balls for the first weighing. Thereafter, space was provided for participants to draw or explain their second weighing.

Design and Procedure.

Participants were tested individually, and were given a maximum of 5 minutes to complete the task.

Results and discussion

The proportions of participants solving each problem differed significantly, χ^2 [2]= 6.01, p=.049, ϕ_c.= 0.27. As predicted, the 7-ball problem was the simplest of the three (9/26 solutions, or 35%), followed by the 8-ball (5/28: 18%), then the 9-ball (2/26: 8%). A chi-square test between the 7-ball condition and the 8-ball and 9-ball conditions combined was also significant in the expected direction, χ^2 [1]= 5.14; p=.023, *d*=0.52. The difference in solution proportions between the 8-ball and 9-ball problems, while in the predicted direction, was not significant (p=.24 by the Fisher Exact Test used because of low expected cell frequencies). These finding are consistent with the weighing preferences predicted from CSP, in that the preferred first weigh for the 7-ball problem lies on a correct solution path, while the preferred weighs in both the 8-ball and 9-ball versions do not.

For the 7-ball problem, CSP predicted that solving through a 3v3 weigh would be more frequent than through the equally-valid 2v2 route. The observed frequencies of the two types of solution were 8 (31%) and 1 (4%), respectively, significantly different from what would be expected if the two solution types were equally likely (p=.02, by the Fisher Exact Test, used because of low expected cell counts). While both weighs lie on a correct solution path, the 2v2 weigh occurs lower in the hierarchy of preferred weighings, and will be considered only by participants operating at three-lookahead. This accounts for the rarity of this valid solution.

Table 2 reports the observed and predicted raw and percentage frequencies of first weighs for the 9-ball, 8-ball and 7-ball problems. Table 2 indicates that the degree of fit between obtained and predicted values is relatively high, and for no problems did the predicted frequency distribution depart significantly from the obtained, by Kolmogorov-Smirnov tests (all p-values >.20). Table 2 reports 18 pairs of predicted and observed scores. Regressing the observed on the predicted scores results in a regression equation with an intercept of -0.08, not significantly different from zero, (t[16] = -0.15, p=.88), a slope of 1.02, not significantly different from 1, (t[16]= 0.30, p=.77), and r=0.96

(F[1,16] = 177.22, $p < .001$). The result attests to a high degree of correspondence between predicted and obtained values.

The "Other" category included two 1v1 weighs, one each for the 9-ball and 8-ball problems, 2.5% of all first weighs, consistent with the 3.4% in Simmel's (1953) results. First weighs of 1v1 may reflect a guessing strategy, since if the pair weighed contains the heavy ball, the problem is solved in one weigh. However, a guessing strategy is not guaranteed to solve in two weighs, as is required by the problem instructions.

Insert Table 2 about here

For both 9-ball and 8-ball problems, the biggest residual values between predicted and obtained scores occurred because one individual in each condition selected a 2v2 weigh. While inconsistent with the present specific predictions, these outcomes are not necessarily inconsistent with the model. Someone operating at four-lookahead would consider a 2v2 first weigh, according to the model, which may explain the unexpected finding. In deriving model predictions we did not consider four-lookahead to be probable, and a potentially more likely explanation is that some participants automatically discount an unequal initial weigh without applying lookahead. In the 8-ball problem, this would mean that the 4v3 would not be considered. This would effectively make two-lookahead have the same result as the present three-lookahead, and the present three-lookahead, the same result as a four-lookahead. The same would hold for the 9-ball problem. Consideration of the 5v4 and 4v3 weighs would be eliminated, and the problem would become equivalent to the 8-ball problem, with two-lookahead operating like three, and three-lookahead like four.

For the 7-ball problem, there were two relatively large residuals, with fewer participants selecting the 4v3 weigh and more selecting the 3v3 weigh than predicted. Conceivably, the fewer balls in the 7-ball problem might allow for a similar automatic elimination of unequal weighs without lookahead, and just a few "zero-lookaheads" operating in this way would account for the discrepant findings.

Experiment 2

Experiment 2 used a multiple trials format with the 8-ball and 7-ball problems to track how weigh selections changed across trials. If participants learn over trails that some weighs are unsuccessful, then we would expect such weighs to be eliminated and replaced by weighs that compare fewer balls. Thus, we anticipate that weigh selections will either remain unchanged (if the participant fails to learn or forgets) or will move systematically down the ranking in order of decreasing maximization. In addition, the experiment compared performance on the 8-ball and the 7-ball problems to further test the prediction that the latter will be easier to solve than the former.

Method

Participants.

Thirty-two further undergraduate students were paid $4 each to participate (identifiers were not collected, so age and gender are unknown).

Materials.

A Classroom Products Balance Scale (Villa Park, IL) was used by participants to make their weighs. Heavier balls, whose increased weight was undetectable by hand, were created by cutting open tennis balls, fixing a 4g lead weight inside, and gluing them closed. To remove differences in visual appearance, standard-weight tennis balls were also sliced open and glued closed.

Design and Procedure.

Participants were randomly assigned in equal numbers to 7-ball and 8-ball conditions. Before the start of the experiment proper, participants practiced with the balance scale using everyday objects. Then the experimenter presented the participant with the 7 or 8 balls as well as written instructions as follows:

> "In front of you is a balance scale and 7 (or 8) tennis balls. Each of the balls is identical in shape and size, but 1 of the 7 (or 8) balls is slightly heavier than the other 6. Only the balance scale is sensitive enough to detect the difference. By holding the balls in your hand, you cannot detect the difference in weights (you can't just pick them up, feel and guess). You have to use the scale in 2 weighings (i.e., 2 uses of the scale) to determine which one is the heavy ball.
>
> You must use the scales as follows. For your first weighing, load the ORANGE pan with the balls you have chosen. Then load the YELLOW pan, with the balls you have chosen. Watch what the scale does. Then the experimenter will remove the balls from the pans and you will then begin your second weighing. At the end of the second weighing you are to tell the experimenter which you think is the heavy ball and then say if you are confident in this decision or if this is just a guess. There is a way to solve this in two weighings without guessing. You will have one minute to attempt the problem. The experimenter will let you know when to start."

To limit incidental learning of ball weights between trials, at the end of each trial, the experimenter collected the balls, left the room, and then brought a new set of balls into the room for the next trial. Participants continued until they had solved the problem correctly three trials in a row, or until ten trials had been completed. To be scored as correct, a solution required that the ball be identified in two weighs. If participants identified the heavier ball by chance (by selecting 1 vs. 1 on a first weigh) they were instructed to continue until they found a way of guaranteeing they could find the heavier ball in two weighs. Participants' weighs were videotaped. At the end of each trial, the experimenter manually recorded the participant's self-rated confidence in their choice of ball.

Results

The data from two participants in the 7-ball condition were dropped from the analysis due to a problem with video-recordings. Data from one other participant was affected on the second weigh of first trial only, and the remaining data were included in the analyses below.

The experiment provided a further test of the prediction that the 7-ball problem will be easier to solve than the 8-ball problem. In this case, the number of correct solutions on the first attempt did not differ significantly between 7-ball (5/14; 36%) and 8-ball conditions (2/16: 13%), χ^2 [1] = 2.25, *p*=.13. Similarly, the number of correct solutions by the end of ten trials did not differ significantly between 7-ball (9/15: 60%) and 8-ball (8/16: 50%) conditions, χ^2 [1]= .87, *p*=.84. However, those who solved did so significantly faster in the 7-ball condition (mean = 2.33 trials; sd = 1.80) than in the 8-ball condition (mean = 5.00 trials; sd = 3.02), *t*[15]=2.24, *p*=.04, *d*=.59. Thus, the results provide partial support for predicted differences between solving the 7-ball and the 8-ball problem.

The major purpose of Experiment 2 was to examine whether the search space in *n*-balls problems conforms to the minimization principle proposed by the model. The minimization principle suggests that, should an initial weigh be rejected on the basis of actual or projected failure to solve, then the search space will expand to include weighs lower in order of maximization. To test this, we counted the number of model consistent first weighs across trials, based on the following criteria. On the first trial, the weigh had to be one of those appearing in Table 1. Thus, for the 8-ball problem, 4v4, 4v3 and 3v3 weighs were counted as model consistent and, for the 7-ball problem, 4v3, 3v3 and 2v2 weighs. For subsequent trials, any weigh appearing in Table 1 plus any weigh lower in the hierarchy that involved an equal number of balls was considered as "valid" and potentially model consistent. Thus 2v2 and 1v1 weighs were included for the 8-ball and 1v1 for the 7-ball, following the assumption of expansion of the search space. (Only equal weighs are assumed to be incorporated by the expanding problem space because unequal weighs would have been eliminated earlier in the process, under Assumption 4 and as illustrated in Table 1.)

To meet the requirement of minimal expansion, weighs had to appear in the same or decreasing order of number of balls weighed across trials. To illustrate, in a sequence of first weighs of 4v4 on Trial 1, 4v4 on Trial 2, and 2v2 on Trial 3, all three weighs were considered model consistent. Of note, this allowed for repetition of a valid weigh (since we did not know how quickly people would eliminate weighs that were not on a solution path) and for skipping levels of the maximization hierarchy, such as going from 4v4 to

2v2 with no intervening 3v3 weigh. (We adopted this approach because we did not know how much lookahead was present and how much thinking took place between trials.) Weighs that were valid but that involved more balls than the immediately preceding trial were deemed to be model inconsistent. Thus, in the sequence 3v3 on Trial 1 and 4v4 on Trial 2, only the Trial 1 weigh would be considered model consistent. However, we treated the second weigh in such a sequence as a "system reset", so that the immediately subsequent weigh was counted as consistent if it was a valid weigh involving the same or a lower number of balls. Thus, weighs in a sequence of trials such as 3v3, 4v4, 4v4, and 3v3, only the second weigh in the sequence would be considered inconsistent.

Finally, an exception concerning repetition of valid weighs was made in the case of unequal weighs, where any repetition was deemed to be model inconsistent (following Assumption 4, that only one trial is required to learn that unequal weighs are uninformative). So, for example, in the 8-ball problem, the first appearance of a 4v3 weigh was considered model consistent (provided it met the other criteria), but subsequent repetitions were not. For example, in the sequence 4v4, 4v3, 4v3, only the first two weighs would be considered model consistent.

The count of model consistent weighs was conducted across all 10 trials for participants who failed to reach the criterion of three consecutive correct trials. For participants who met the criterion, the count stopped after the first of the three consecutive correct trials. This avoided counting repetitions of known successful weighs as model consistent. Because more participants solved the 7-ball problem than the 8-ball, and did so more quickly, 7-ball trials typically terminated sooner than 8-ball, resulting in fewer total trials. For the 8-ball condition, 116 of 139 (84%) total weighs were model consistent, as defined above. For the 7-ball condition, 63 of 87 (72%) were consistent.

To provide a chance model for comparison, we first examined the first trial only for the number of model consistent results that would be expected by chance, assuming that weighs are randomly selected with replacement under the constraint of at least one ball placed in each pan. For the 8-ball problem, the chance proportion of model consistent first weighs is 25%, significantly lower than the observed proportion, of 84%, χ^2 [1] = 16.74, p<.001. For the 7-ball problem, the proportion expected by chance is 34%, again significantly lower than the 75% observed, χ^2 [1] = 14.36, p<.001.

To provide a chance model for weighs on subsequent trials we conducted a Monte Carlo simulation of the procedure, using the criteria identified above to identify model consistent weighs, again using weighs randomly selected under the constraint of at least one ball placed on each side of the balance. Based on 50,000 replications, the results for the first trial indicated 25% and 34% model consistent moves for 8-ball and 7-ball problems respectively, identical to the theoretical calculations. For the remaining trials 2 through 10, the simulation for the 8-ball problem indicated chance percentages of model consistent weighs ranging from 7.9% to 8.1% with a mean of 8.0% compared with a mean percentage of 85% produced by participants. For the 7-ball problem, the chance proportion of model consistent moves ranged from to 8.7% to 8.9% with a mean of 8.8%, compared with a mean of 71% by participants.

While the results strongly support the predicted pattern of search space expansion under minimization and maximization heuristics, there were some notable departures, contributed by a minority of participants. One participant in the 7-ball condition produced a first weighing of 4v3 on all ten trials, only the first of which was counted as model consistent under our criteria. Another did so on the final seven trials, none of which counted as model consistent. Nevertheless, although not counted, these 16 weighs could be interpreted as persistent, if perverse, commitments to maximizing under zero-lookahead.

A different type of departure from model behavior was the choice by eight participants of a 1v1 weigh on the very first trial (chosen by five participants in the 8-ball and three in the 7-ball conditions). The unexpected choice of a 1v1 weigh on a first attempt appeared in Simmel's (1953) data and in the present Experiment 1, but was relatively infrequent, at around 3% of attempts. In contrast, 1v1 weighs represented approximately 27% of first weighs in the present experiment. While a 1v1 first weigh cannot lead to a guaranteed correct solution, as required by the instructions, it may appear to some participants to be a reasonable gamble in a multiple-trial study. The present experiment's repeated trials format meant that a participant adopting this approach had available a total of 20 weighs to find a solution by chance (10 trials of two weighs each). Alternatively, or additionally, the pressure of a one minute time limit may have encouraged these participants to adopt a guessing strategy as a reasonable path to success.

General Discussion

This paper examined whether the CSP theory of problem solving can predict rational and irrational move selections. First, we described how maximization and minimization heuristics constrain and control search, and analyzed their operation in the context of the *n*-ball problem. Second, we proposed the concept of zero lookahead, where a problem-solver selects a maximizing move without evaluating its consequences, essentially a combination of trial and error, maximization and minimization heuristics. In the context of the *n*-ball problem, such decisions may manifest themselves as unequal weighs. Two experiments were reported which provided evidence for both extensions to the theory, in terms of solution rates, or relative frequencies of different weighs, or both.

The concept of zero lookahead explains trials in which participants select unequal weighs (e.g., 5v4). We suspect that many behaviours associated with seeming lack of engagement or awareness in a problem's task environment reflect the operation of zero-lookahead. A zero-lookahead approach is increased when the context tolerates failure and allows rapid feedback on performance, as in Experiment 2, in which case an individual might see strategic advantages in a non-planning approach to maximizing progress – in other words, an 'act first, think later' approach to selecting potential moves.

It is possible that unequal weighs arose, not through zero-lookahead, but through some other mechanism. For instance, participants may simply not understand the problem (e.g., that the scope for a second weigh would depend upon the outcome of the first; or that an unequal weigh would not discriminate between the effect of one balance pan having more balls than the other and the effect of one balance pan containing the slightly heavier ball). Zero-lookahead is indistinguishable from initial misunderstanding, since it leads participants to select a move that appears to make progress without considering the consequences of that move for what follows. However, if participants failed to understand the problem, then we would expect misunderstandings to continue to affect their performance relative to participants who had not produced unequal weighs. Three pieces of data speak to this question:

i. Unequal weighs appeared in both experiments reported in the paper, which were conducted at different times with different groups of participants and under somewhat different procedures. Critically, they arose in Experiment 2, where problem understanding was arguably simplified by the provision of working scales which make explicit the effects of adding more balls to one side (cf. Zhang, 1993, who demonstrated how making problem constraints explicit can facilitate performance). Importantly, the proportion of 5v4 weighs reported by Simmel (1953) was 37%, which is comparable with our findings.

ii. If lack of problem understanding explains the unequal weighs, they should arise on all three problems, but they arose primarily on 7- and 9-ball problems, and rarely on the conceptually isomorphic and superficially similar 8-ball problem.

iii. If lack of problem understanding explains the unequal weighs, then one would expect participants who produced them to be significantly less likely to solve overall. This was not the case: Of the 16 participants who produced unequal weighs on at least one trial, 9 eventually solved, compared with 8 solutions from 15 participants who produced only equal weighs, χ^2 [1] = 0.03, p=.86.

Another explanation is that participants may have adopted a deliberate strategy of seeking counter-intuitive moves to ‘perturb’ the problem space when they have reached impasse. This strategy would be an implementation of an exhortation to “think outside the box” or to think laterally (DeBono, 1967). Two pieces of evidence speak against this explanation. First, if this perturbation strategy were used to overcome impasse, then one might expect unequal moves to arise with all three problems, but as noted above, they were found mainly with 7- and 9-ball problems. Second, one would expect an equal distribution of unequal moves, but as Table 2 shows, unequal weighs mainly involved maximizing the number of balls weighed.

The high frequency of unequal weighs is an unexpected result that begs for an explanation, and we believe that the zero-lookahead hypothesis is a viable explanation. To look at this issue another way, given that the *n*-Ball problem clearly lies within the competence of adult college student participants, if unequal weighs reflect a lack of problem understanding, then what causes this lack of understanding? We suggest it is the application of zero-lookahead.

Whether zero-lookahead was encouraged by the situation, or whether it reflects more enduring characteristics of the individual is an issue that our research does not address. However, a distinction is often made between fast, spontaneous, cognitive processes and those that are slower and more deliberate, and there is evidence that the latter correlates with individual differences in general intelligence (Stanovich & West, 2000). In the area of social cognition, a similar distinction underlies instruments designed to test "need for cognition" (Cacioppo & Petty, 1982) and intuitive versus analytical thinking (Epstein, Pacini, Denes-Raj, & Heier, 1996). It has been proposed that different involvement of the two types of processes underlies differences in reasoning, judgment, decision-making, and risk-taking (Evans, 2010; Kaheneman & Frederick, 2002). Frederick (2005) proposed a three item "Cognitive Reflection Test" to measure an inclination to make impulsive decisions without reflection, and found that scores were related to differences in delaying gratification and taking risks. If zero-lookahead reflects individual characteristics, then the Cognitive Reflection Test may be promising way to assess it.

We have not explicitly measured lookahead but simply inferred its presence from the selection of move sequences. Measuring lookahead is problematic: either one must use a concurrent measure that might interfere with task performance, or infer it from measures of individual differences in capacity that ignore contextual factors such as motivation. Our approach is found often in the problem-solving literature (e.g., Jones, 2003; Ohlsson, 2011) and seems to yield satisfactory results. With the 9-dot problem, MacGregor et al modeled lookaheads of 1, 2 and 3, and showed how empirical data could be fitted across a sample with different degrees of lookahead. In the case of the *n*-Ball problem, we suspect lookahead is generally likely to be low, partly because the problem only contains two steps (compared with four in the 9-dot problem). Also, to evaluate effectively the results of lookahead in the *n*-Ball problem is complex: it requires consideration of a complex nested logical premise of the form:

If *choice (X balls against Y)* on the first weigh, then

either scales balance, in which case heaviest must be in unweighed balls

or side X/Y drops, in which case heaviest must be in X/Y

In contrast, to evaluate the effects of any level of lookahead in the 9-dot problem, one need only envisage and count the dots that remain uncancelled. We suggest that the complexity of executing lookahead in the *n*-Ball problem is the reason why individuals often attempt non-balancing weighs: they are acting rather than thinking, because action gives results that are easier to evaluate than the products of lookahead.

We have argued previously (Ormerod et al, 2002) that some problems (e.g., the 9-dot problem) are amenable to planning by visualization because progress can be evaluated using subitization (i.e., one can see at a glance how many dots are cancelled). Others (e.g., the 6-coin problem) are harder to visualize, since progress can only be evaluated by inspecting each component of the problem array separately (i.e., checking each coin to see how many others it touches). In the same way that Ormerod et al (2002) argue one can differentiate between problems according to their amenability to planning via visualization, one might differentiate between problems in terms of their amenability to planning via inference. In the case of the *n*-Ball problem, the inferences required to capitalize upon planning ahead are complex. They involve disjunctions nested within conditionals, which are known to be a source of difficulty (Johnson-Laird, 1993). One might predict that problems involving evaluation via simple inferences would be more amenable to planning. In future research, it may be useful to distinguish between lookahead (planning via visualization) and thinkahead (planning via inference).

References

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*,116-131.

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*, 14-27.

Cooper, M. (2011). The implications of Fukushima. *Bulletin of the Atomic Scientists*, 67, p.8.

DeBono, E. (1967). *New Think: The Use of Lateral Thinking in the Generation of New Ideas*. New York: Basic Books.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. J*ournal of Personality and Social Psychology*, *71*, 390-405.

Evans, J. St. B. T. (2010). Intuition and Reasoning: A Dual-Process Perspective. *Psychological Inquiry*, *21*, 313-326.

Frederick, S. (2005). Cognitive reflection and decision making. J*ournal of Economic Perspectives*, *19*, 25-42.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451-482.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way:models of bounded rationality. *Psychological Review*, *104*, 650-669

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive*

Science, 3, 275-310.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology, 6*, 270-292.

Johnson-Laird, P.N. (1993). *Mental models*. Cambridge: Cambridge University Press.

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1017-1027.

Kahneman D. (2003). Maps of bounded rationality: a perspective on intuitive judgment and choice. In *Les Prix Nobel: The Nobel Prizes 2002*, ed. T Frangsmyr, pp. 449-89. Stockholm: Nobel Foundation.

Kahneman, D.,&Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: Ajudgment of representativeness. *Cognitive Psychology*, *3*, 430-454.

Kahneman D, Tversky A, eds. 2000. *Choices, Values, and Frames.* New York: Cambridge Univ. Press/Russell Sage Foundation.

Kaplan, C. A. & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 373-419.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: explorations in the problem space of difficulty. *Cognitive Psychology, 22*, 143-183.

Lee, L., Goodwin, G., & Johnson-Laird, P.N. (2008). The psychological puzzle of Sudoku. *Thinking and Reasoning*, 14, pp 342-364.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information-processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 176-201.

Morris, R. & Ward, G. (Eds.). (2005). *Cognitive Psychology of Planning*. London: Psychology Press.

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

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, 103, 381-391. doi:10.1037/0033-295X.103.2.381

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin and Review,* 10, 289-318.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology, 35*, 34-70.

Ohlsson, S. (2011). *Deep learning*. Cambridge: Cambridge University Press.

Ormerod, T. C., MacGregor, J. N., & Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 791-799.

Ormerod, T. C., & Richardson, J. (2003). On the generation and evaluation of inferences from single premises. *Memory and Cognition, 31*, 467-478.

Simmel, M. L. (1953). The Coin Problem - a Study in Thinking. *American Journal of Psychology, 66*(2), 229-241.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs.

Cognitive Psychology, 8, 165-190.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*, 701-726.

Stuss, D.T., & Alexander, M.P. (2007). Is there a dysexecutive syndrome? *Phil. Trans. R. Soc.* B *29,* 362, 901-915.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences,* 28, 675-735.

Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. Science 185:1124–31

VanLehn, K (1989). Problem solving and cognitive skill acquisition. In M.I. Posner (Ed.), *Foundations of Cognitive Science*, Cambridge, MA: MIT Press.

Wakebe, T., Sato, T., Watamura, E., & Takano, Y. (2012). Risk aversion in information seeking, *Journal of Cognitive Psychology*, 24, 125-133.

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General, 110*(2), 169-192.

Zhang, J. (1993). The nature of external representations in problem solving . *Cognitive Science*, 179-217.

Table 1.

Model assumptions (upper panel) and application of the model to the 9-ball, 8-ball and 7-ball problems

ASSUMPTIONS

1. Weighs are sampled in order of maximization.	2. Lookahead=number of weighs mentally tested to find how many balls are eliminated.	3. A lookahead episode comprises selection and test until lookahead is exhausted	4. One trial only is required to learn that unequal weighs are uninformative
5. If at the end of a lookahead episode one weigh meets criterion it is selected.	6. If at the end of a lookahead episode > one weigh meets criterion, one selected at random	7. If at the end of a lookahead episode no weigh meets criterion then the next ranked weigh is selected without being tested.	

APPLICATION

Problem version	First weigh	Lookahead 0	Lookahead 1	Lookahead 2	Lookahead 3
9-ball	5v4	Select 5v4	Test,reject 5v4	Test,reject 5v4	Test,reject 5v4
	4v4		Select 4v4	Test,select 4v4	Test,hold 4v4
	3v3				Test,hold 3v3
					Select 3v3 (50%) and 4v4 (50%)
8-ball	4v4	Select 4v4	Test,select 4v4	Test,hold 4v4	Test,hold 4v4
	4v3			Test,reject 4v3	Test,reject 4v3
	3v3			Select 4v4	Test,hold 3v3
					Select 3v3 (50%) and 4v4 (50%)
7-ball	4v3	Select 4v3	Test,reject 4v3	Test,reject 4v3	Test,reject 4v3
	3v3		Select 3v3	Test,select 3v3	Test,hold 3v3
	2v2				Test,hold 2v2
					Select 2v2 (50%) and 3v3 (50%)

Table 2

Observed and predicted frequencies (percentage frequencies) of first weighs in Experiment 2, for the 9-ball, 8-ball and 7-ball problems.

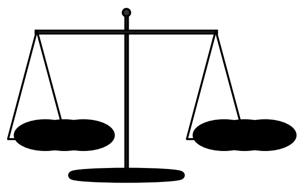
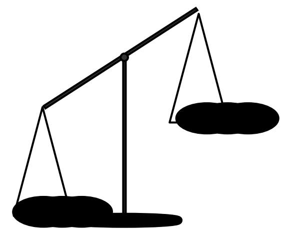
First	9-ball		8-ball		7-ball	
Weigh	Observed	Predicted	Observed	Predicted	Observed	Predicted
5v4	9 (35)	9.6 (37)	na		na	
4v4	11 (42)	10.9 (42)	22 (79)	22.1 (79)	na	
4v3	0 (0)	0 (0)	0 (0)	0 (0)	6 (23)	9.6 (37)
3v3	2 (8)	2.7 (10.5)	3 (11)	2.9 (10.5)	17 (65)	10.9 (42)
3v2	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2v2	1 (4)	0 (0)	1 (4)	0 (0)	2 (8)	2.7 (10.5)
Other	3 (12)	2.7 (10.5)	2 (7)	2.9 (10.5)	1 (4)	2.7 (10.5)
N	26		28		26	

Figure Captions

Figure 1. Solution paths to the 7-, 8- and 9-ball problem variants. In all three cases the solution requires a first weigh of three balls against three. If this weigh is imbalanced, the balls from the lower pan are selected and one is weighed against one, which identifies the heavy ball directly. If the first weigh is balanced, the heavy ball is in the unweighed group. For the 7-ball problem, it is the remaining ball, while for the 8-ball and 9-ball problems it can be identified with one more weighing.

Figure 2. First lines given in the 9-dot problem (panels a and b), and most frequent second and third lines drawn in response (panels c and d), MacGregor, Ormerod and Chronicle (2001).

Figure 1

<table>
<tr><td>First weighing outcome</td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>7- ball

8- ball

9- ball</td><td colspan="2">Unweighed ball is heaviest

Select the two unweighed balls for next weigh

Select two of the three unweighed balls for next weigh</td><td colspan="2">Select two balls from the lower pan for next weigh

Select two balls from the lower pan for next weigh

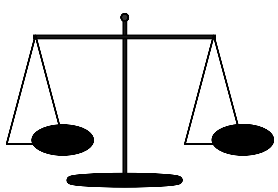
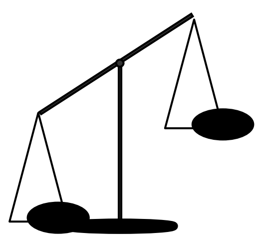
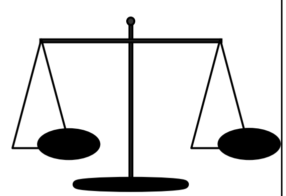
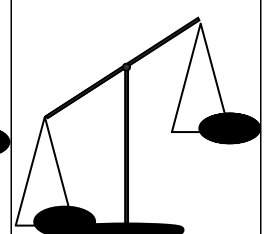
Select two balls from the lower pan for next weigh</td></tr>
<tr><td>Second weighing outcome</td><td></td><td></td><td></td><td></td></tr>
<tr><td>7- ball</td><td>N/A</td><td>N/A</td><td>Unweighed ball is heaviest</td><td>Heaviest ball is on lower pan</td></tr>
<tr><td>8- ball</td><td>N/A</td><td>Heaviest ball is on lower pan</td><td>Unweighed ball is heaviest</td><td>Heaviest ball is on lower pan</td></tr>
<tr><td>9- ball</td><td>Unweighed ball is heaviest</td><td>Heaviest ball is on lower pan</td><td>Unweighed ball is heaviest</td><td>Heaviest ball is on lower pan</td></tr>
</table>

Figure 2