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

Thomas C. Ormerod · James N. MacGregor ·
Edward P. Chronicle · Andrew D. Dewald · Yun Chu

Published online: 16 April 2013
© Psychonomic Society, Inc. 2013

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. In this article, we report two experiments in which we explored 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 toward the problem goal and minimization of the problem space in which attempts are sought. The effectiveness of these heuristics was 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 noninsight problems.

Keywords Planning · Problem-solving · Insight · Planning failures · Look ahead

Introduction

Planning is generally regarded as a prerequisite for successful cognitive performance (Hayes-Roth & Hayes-Roth,

1979; Morris & Ward, 2005; Newell & Simon, 1972; O'Hara & Payne 1998) and differentiates humans from many other species (Tomasello, Carpenter, Call, Behne, & Moll, 2005). By looking ahead mentally from a current state in order to anticipate a new state, an individual can evaluate the likely success of a move sequence before its execution. Such 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., Jones, 2003; MacGregor, Ormerod, & Chronicle, 2001).

While failures of planning have been implicated in a number of neurodegenerative 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 behaviors that in retrospect appear irrational, like sitting on a branch of a tree while you saw it off.

Over the last 40 years, evidence has amassed that human performance deviates from what is normatively rational (Kahneman & Tversky, 1972, 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 & Gaissmaier, 2011; Gigerenzer & Goldstein, 1996), 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, 1996, 2003). Recently, an information-gain explanation

T. C. Ormerod (✉)
Department of Psychology, Lancaster University,
Lancaster LA1 4YF, UK
e-mail: t.ormerod@lancaster.ac.uk

J. N. MacGregor
University of Victoria, Victoria, BC, Canada

E. P. Chronicle · A. D. Dewald · Y. Chu
University of Hawaii at Manoa, Manoa, Hawaii

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 did not appear to be readily explained by information gain. Here we propose and test an explanation of the behavior using a model of problem solving that 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 (Chronicle, MacGregor, & Ormerod, 2004; MacGregor et al., 2001; Ormerod et al., 2002).

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 eight- and nine-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 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 balls against any other two.²

For the eight-ball problem, Simmel (1953) reported that 78% of 58 participants selected an initial weigh of four versus four (4v4), while 22% selected another symmetrical weigh. For the nine-ball problem, the most frequent initial weighs were 4v4 (42%) and 5v4 (37%). The 5v4 weighs seem irrational: A weigh in which the numbers on each side of the balance are 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 was one made by over a third of Simmel’s university student participants.

Simmel (1953) 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 eight-ball problem, the two tendencies are mutually compatible, leading to a 4v4 comparison. For the nine-ball problem, the tendencies conflict and lead to an imperfect outcome.

Simmel’s (1953) proposed tendency of symmetry applies only to problems that allow for 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 fail to generate a new move. Other problem-specific heuristics have been proposed for the mutilated-checkerboard (Kaplan & Simon, 1990) and rings (Kotovskiy & Simon, 1990) problems, and for 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 problems. Second, the principles for switching between heuristics across move attempts tend to be arbitrary and/or problem-specific, themselves. Third, and critically for the present study, 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 nine-ball problem, yet in Simmel’s (1953) data, such weighs 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, in the present report we propose and test a general approach to explaining behavior in the n -ball task, including the selection of “irrational” unequal weighs.

CSP theory

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

¹ We are grateful to Stellan Ohlsson for pointing us toward this article.

² The alternative solution to the seven-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 one against one on the second weigh, or else select the two balls from the lower scale and weigh one against one.

Fig. 1 Solution paths to the seven-, eight-, and nine-ball problem variants. In all three cases, the solution requires a first weigh of three balls against three. If this weigh is not balanced, 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 seven-ball problem, the heavy ball is the remaining one, while for the eight- and nine-ball problems, it can be identified with one more weighing.

First weighing outcome
7-ball	Unweighed ball is heaviest		Select two balls from the lower pan for next weigh	
8-ball	Select the two unweighed balls for next weigh		Select two balls from the lower pan for next weigh	
9-ball	Select two of the three unweighed balls for next weigh		Select two balls from the lower pan for next weigh	
Second weighing outcome
7-ball	N/A	N/A	Unweighed ball is heaviest	Heaviest ball is on lower pan
8-ball	N/A	Heaviest ball is on lower pan	Unweighed ball is heaviest	Heaviest ball is on lower pan
9-ball	Unweighed ball is heaviest	Heaviest ball is on lower pan	Unweighed ball is heaviest	Heaviest ball is on lower pan

Maximization heuristic

The *maximization* heuristic operates to sample moves that appear to maximize progress toward 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 that 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 (see Fig. 2). 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. 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 in which the first line is given, as illustrated in Figs. 2a, b (MacGregor et al., 2001; Weisberg & Alba, 1981). Maximization operates in CSP theory to enable choices 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 an 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 a search

Fig. 2 The nine-dot problem, with first line given: “Here are nine dots arranged in a square matrix. Can you draw three further straight lines starting from one end of the given line, without lifting your pencil, through all nine dots?” (panels a and b). The most frequent second and third lines drawn in response to variants a and b are shown in panels c and d, respectively. See MacGregor, Ormerod, and Chronicle (2001).

for moves that might meet a criterion of 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 it fails to allow a solution. Eventually, individuals exhaust all the criterion-meeting moves and relax the minimization heuristic, which allows for 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.

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 is illustrated in Figs. 2c, d). By comparing observed performance with the CSP predictions, MacGregor et al. estimated the proportions of participants using lookahead of one, two, three, and four 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 will have to be considered before a satisfactory one can be 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 that zero and one lookahead will 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 would lead to qualitatively different move selections than would 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 that it eliminates. As we illustrate 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 are 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 nine-, eight-, and seven-ball problems.

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 nine-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 nine-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.

Table 1 Model assumptions (upper panel) and application of the model to the nine-, eight-, and seven-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, >1 weigh meets criterion, one is 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	First	Lookahead			
Version	Weigh	0	1	2	3
Nine-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%)
Eight-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%)
Seven-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%)

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 eight balls weighed would be eliminated, while if it is unbalanced, the four balls on the lighter side plus the unweighed ball would be 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 eight balls if balanced and five 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. The balanced and unbalanced outcomes would each eliminate six 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 numbers of balls eliminated are 5.33 for the 4v4 weigh (i.e., $8 \times .11 + 5 \times .89$) and 6 for the 3v3 weigh (i.e., $6 \times .33 + 6 \times .67$). 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 the eight- and seven-ball problems is essentially the same as for the nine-ball. However, some details are relevant to some of the predictions described later. For the eight-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 four balls. The criterion is met, lookahead is exhausted, and the weigh is selected (Assumption 5). Of note, in the eight-ball and nine-ball versions, the first weighs selected at one lookahead are physically the same but have different statuses. In the eight-ball version, the 4v4 weigh has been tested, while in the nine-ball version, it has not, which places the eight-ball problem one step closer to solution than the nine-ball problem for the same level of lookahead. This point is germane to one of the predictions developed below.

For the eight-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 would eliminate all six balls weighed, while an unbalanced outcome would eliminate five (the three on the lighter side plus the two unweighed). This exhausts lookahead, and the 4v4 weigh and 3v3 weigh are assumed to be selected with equal probability. However, if selection is proportional to the expected number of balls eliminated, then 43% of selections would favor the 4v4 weigh, and 57% the 3v3; both of these results again are relatively close to the 50% proposed by Assumption 5.

Finally, at three lookahead in the seven-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 six weighed balls if balanced, and four 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 four balls, and if unbalanced, it eliminates five, thereby meeting the criterion. The 3v3 and 2v2 weighs are therefore preferred equally. (On the basis of the expected number of balls eliminated, the 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 might find two different solutions to the seven-ball problem.

Using the model to estimate lookahead

Simmel's (1953) study reported the frequencies of initial weighs for the nine- and eight-ball problems, and it 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 proportions of participants operating with zero, one, two, and three lookahead, respectively.

According to our analysis of the nine-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 (1953) Table IV (nine-ball first condition), the proportion selecting the 5v4 weigh was 37%, from which we estimate that $w = 37\%$.

Similarly, the analysis for the nine-ball problem in Table 1 indicates that the proportion selecting a 3v3 first weigh will consist of half of those using three lookahead, or $.5z$. The results shown in Simmel's (1953) 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 $x + y + .5z$. From Simmel's (1953) 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 (1953) 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 (1953) eight-ball results

As an initial test, we may use the estimates derived above from Simmel's (1953) nine-ball experiment to predict performance in her eight-ball conditions. From Table 1, for the eight-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 $w + x + y + .5z$, while those selecting a 3v3 first weigh will be half of those operating at three lookahead, or $.5z$. 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 (1953) Table II (eight-ball problem only and eight-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 eight-ball problem appear encouraging.

Later in this article, we will 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 seven-ball problem will be easiest to solve, because the predicted most common first weigh, 3v3, lies on a correct solution path. For the eight-ball and nine-ball problems, the most common first weighs are not on the solution path.

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

Third, of the two correct solutions to the seven-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 2×2 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 1–3, above, using the seven-ball, eight-ball, and nine-ball problems. We also examined the frequencies of selection of first weighs, to test the predictions from Table 1. Experiment 2 examined detailed performance

across ten trials of the seven- and eight-ball problems, to test predictions based on the proposed minimal expansion of the problem space.

Experiment 1

In Experiment 1, we tested several of CSP's predictions about n -ball performance: First, that the seven-ball problem will be simpler than the eight- or nine-ball problems; second, that the eight-ball problem will be simpler than the nine-ball problem; and third, that the 3v3 solution to the seven-ball problem will be more frequent than the 2v2 solution. Furthermore, the experiment allowed for 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% was 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 proportions of first weighs in the nine-, eight-, and seven-ball problems. For the nine-ball problem, the resulting predictions are necessarily similar to Simmel's (1953) results, since the estimates were derived from Simmel's nine-ball condition. However, the predictions for the eight- and seven-ball problems are independent of Simmel's results. To illustrate, the predictions for the seven-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 nine-ball, eight-ball, and seven-ball problems are summarized in Table 2.

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

Method

Participants A group of 80 unpaid undergraduate students volunteered to participate (identifiers were not collected, so

Table 2 Observed and predicted frequencies (percentage frequencies) of first weighs in Experiment 2, for the nine-, eight-, and seven-ball problems

First Weigh	Nine-Ball		Eight-Ball		Seven-Ball	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
5v4	9 (35)	9.6 (37)	N/A		N/A	
4v4	11 (42)	10.9 (42)	22 (79)	22.1 (79)	N/A	
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)
<i>N</i>	26		28		26	

age and gender are unknown). In all, 26 attempted the seven-ball problem, 28 attempted the eight-ball problem, and 26 attempted the nine-ball problem.

Materials The participants received a booklet containing problem and information sheets as a function of condition, with the following problem instructions: “You have [seven, eight, or 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 is to find out which one is heavier. You have a balance scale, and you can use it only twice.” These 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 min to complete the task.

Results and discussion

The proportions of participants solving each problem differed significantly, $\chi^2(2) = 6.01$, $p = .049$, $d = 0.27$. As predicted, the seven-ball problem was the simplest of the three (9/26 solutions, or 35%), followed by the eight-ball (5/28, 18%), and then the nine-ball (2/26, 8%). A chi-square test between the seven-ball condition and the eight- and nine-ball conditions combined was also significant in the expected direction, $\chi^2(1) = 5.14$, $p = .023$, $d = 0.52$. The difference in solution proportions between the eight- and nine-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 findings are consistent with the weighing preferences predicted from CSP, in that the preferred first weigh for the seven-ball problem lies on a correct solution path, while the preferred weighs in both the eight-ball and nine-ball versions do not.

For the seven-ball problem, CSP predicted that solving by means of a 3v3 weigh would be more frequent than by means of the equally valid 2v2 route. The observed frequencies of the two types of solutions 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 nine-ball, eight-ball, and seven-ball problems. Table 2 indicates that the degree of fit between the obtained and predicted values is relatively high, and for no problems did the predicted frequency distribution depart significantly from the one 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 $r = .96$ [$F(1, 16) = 177.22$, $p < .001$]. This result attests to a high degree of correspondence between the predicted and obtained values.

The Other category included two 1v1 weighs, one each for the nine-ball and eight-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 was required by the problem instructions.

For both the nine- and eight-ball problems, the biggest residual values between the predicted and obtained scores occurred because one individual in each condition selected a 2v2 weigh. While they are 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 might explain the unexpected finding. In deriving the 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 eight-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 nine-ball problem. Consideration of the 5v4 and 4v3 weighs would be eliminated, and the problem would become equivalent to the eight-ball problem, with two lookahead operating like three, and three lookahead like four.

For the seven-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 seven-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

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

Method

Participants A group of 32 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 4-g lead weight inside, and gluing them closed. To remove differences in

visual appearance, the standard-weight tennis balls were also sliced open and glued closed.

Design and procedure Participants were randomly assigned in equal numbers to the seven-ball and eight-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 seven or eight balls, as well as with the following written instructions:

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 others. 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 the 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 the choice of ball.

Results

The data from two participants in the seven-ball condition were dropped from the analysis due to a problem with the

videorecordings. The data from one other participant were affected on the second weigh of the first trial only, and the remaining data were included in the analyses below.

The experiment provided a further test of the prediction that the seven-ball problem is easier to solve than the eight-ball problem. In this case, the numbers of correct solutions on the first attempt did not differ significantly between the seven-ball (5/14, or 36%) and eight-ball (2/16, 13%) conditions, $\chi^2(1) = 2.25$, $p = .13$. Similarly, the number of correct solutions by the end of ten trials did not differ significantly between the seven-ball (9/15, 60%) and eight-ball (8/16, 50%) conditions, $\chi^2(1) = 0.87$, $p = .84$. However, those who solved did so significantly faster in the seven-ball condition ($M = 2.33$ trials, $SD = 1.80$) than in the eight-ball condition ($M = 5.00$ trials, $SD = 3.02$), $t(15) = 2.24$, $p = .04$, $d = 0.59$. Thus, the results provide partial support for the predicted differences between solving the seven-ball and eight-ball problems.

Discussion

The major purpose of Experiment 2 was to examine whether the search space in n -ball 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, the search space would expand to include weighs lower in the order of maximization. To test this, we counted the number of model-consistent first weighs across trials, on the basis of the following criteria: On the first trial, the weigh had to be one of those appearing in Table 1. Thus, for the eight-ball problem, 4v4, 4v3, and 3v3 weighs were counted as being model consistent, and for the seven-ball problem, 4v3, 3v3, and 2v2 weighs were counted. 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 “valid” and potentially model consistent. Thus, 2v2 and 1v1 weighs were included for the eight-ball, and 1v1 for the seven-ball, problem, following the assumption of expansion of the search space. (Only equal weighs were 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 in 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 would be required to learn that unequal weighs are uninformative). So, for example, in the eight-ball problem, the first appearance of a 4v3 weigh was considered model consistent (provided that 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 ten 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 being model consistent. Because more participants solved the seven-ball than the eight-ball problem, and did so more quickly, seven-ball trials typically terminated sooner than eight-ball trials, resulting in fewer total trials. For the eight-ball condition, 116 of 139 (84%) total weighs were model consistent, as defined above. For the seven-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 were randomly selected with replacement under the constraint of at least one ball being placed in each pan. For the eight-ball problem, the chance proportion of model-consistent first weighs was 25%, significantly lower than the observed proportion of 84%, $\chi^2(1) = 16.74$, $p < .001$. For the seven-ball problem, the proportion expected by chance was 34%, again significantly lower than the 75% observed, $\chi^2(1) = 14.36$, $p < .001$.

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 being placed on each side of the balance. On the basis of 50,000 replications, the results for the first trial indicated 25% and 34% model-

consistent moves for the eight-ball and seven-ball problems, respectively, identical to the theoretical calculations. For the remaining Trials 2–10, the simulation for the eight-ball problem indicated chance percentages of model-consistent weighs ranging from 7.9% to 8.1%, with a mean of 8.0%, as compared with the mean percentage of 85% produced by participants. For the seven-ball problem, the chance proportion of model-consistent moves ranged from 8.7% to 8.9% with a mean of 8.8%, as 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 the participants. One participant in the seven-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 eight-ball and three in the seven-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 it was relatively infrequent, at around 3% of attempts. In contrast, 1v1 weighs represented approximately 27% of the first weighs in the present experiment. While a 1v1 first weigh cannot lead to a guaranteed correct solution, as required by the instructions, it might 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 (i.e., ten trials of two weighs each). Alternatively, or additionally, the pressure of a 1-min time limit may have encouraged these participants to adopt a guessing strategy as a reasonable path to success.

General discussion

In this article, we have 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, in which 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 that 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 selected unequal weighs (e.g., 5v4). We suspect that many behaviors 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 nonplanning 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 have understood the problem (e.g., that the scope for a second weigh would depend on 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, 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:

1. Unequal weighs appeared in both experiments reported here, 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 that made explicit the effects of adding more balls to one side (cf. Zhang, 1997, 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.
2. If lack of problem understanding explained the unequal weighs, they should arise on all three problems, but they arose primarily on the seven- and nine-ball problems, and rarely on the conceptually isomorphic and superficially similar eight-ball problem.
3. If lack of problem understanding explained the unequal weighs, 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, nine eventually solved, as compared with eight solutions from the 15 participants who produced only equal weighs, $\chi^2(1) = 0.03$, $p = .86$.

Another explanation is that participants may have adopted a deliberate strategy of seeking counterintuitive moves to “perturb” the problem space when they 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, one might expect unequal moves to arise with all three problems, but as noted above, they were found mainly with the seven- and nine-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, what causes this lack of understanding? We suggest, the application of zero lookahead.

Whether zero lookahead was encouraged by the situation or reflected 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; Kahneman & Frederick, 2002). Frederick (2005) proposed a three-item “cognitive reflection test” to measure an inclination to make impulsive decisions without reflection, and he found that scores were related to differences in delaying gratification and taking risks. If zero lookahead reflects individual characteristics, the cognitive reflection test may be a 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 often found in the problem-solving literature (e.g., Jones, 2003; Ohlsson, 2011) and seems to yield satisfactory results. With the nine-dot problem, MacGregor et al. (2001) modeled lookaheads of one, two, and three 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 that lookahead is generally likely to be low, partly because the problem only contains two steps (as compared with four in the nine-dot problem). Also, to evaluate the results of lookahead in the n -ball problem effectively would be complex: It would require 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 nine-dot problem, one need only envisage and count the dots that remain uncanceled. We suggest that the complexity of executing lookahead in the n -ball problem is the reason why individuals often attempt unequal 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 nine-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 canceled). Others (e.g., the six-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) argued that 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 think-ahead (planning via inference).

Author note Edward P. Chronicle sadly deceased prior to publication of this article.

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, and Cognition*, 30, 14–27.
- Cooper, M. (2011). The implications of Fukushima. *Bulletin of the Atomic Scientists*, 67, 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. *Journal of Personality and Social Psychology*, 71, 390–405.
- Evans, J. S. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, 21, 313–326.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal 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.
- Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, 6, 270–292.
- Hayes-Roth, B., & Hayes-Roth Perrault, F. (1979). A cognitive model of planning. *Cognitive Science*, 3, 275–310. doi:[10.1207/s15516709cog0304_1](https://doi.org/10.1207/s15516709cog0304_1)
- 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 T. Frangsmyr (Ed.), *Les Prix Nobel: The Nobel Prizes 2002* (pp. 449–489). 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: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. New York: Cambridge University 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, 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: Prentice-Hall.
- 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.
- Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, 103, 381–391. doi:[10.1037/0033-295X.103.2.381](https://doi.org/10.1037/0033-295X.103.2.381)
- Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review*, 10, 289–318.
- 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, 791–799.
- Ormerod, T. C., & Richardson, J. (2003). On the generation and evaluation of inferences from single premises. *Memory & Cognition*, 31, 467–478.
- Simmel, M. L. (1953). The coin problem—A study in thinking. *American Journal of Psychology*, 66, 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? *Philosophical Transactions of the Royal Society B*, 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.
- VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science*. Cambridge: 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, 169–192.
- Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21, 179–217. doi:[10.1207/s15516709cog2102_3](https://doi.org/10.1207/s15516709cog2102_3)