Running head: GIVING UP PROBLEM SOLVING

Giving Up Problem Solving

Stephen J. Payne and Geoffrey B. Duggan

University of Bath

Bath, United Kingdom

Correspondence concerning this article should be sent to: Stephen Payne, Department of Computer Science, University of Bath, Bath, BA2 7AY, UK. Email: s.j.payne@bath.ac.uk.

Abstract

How do people decide to abandon a problem? Participants were presented with unsolvable water-jar problems, having been accurately informed of the prior probability of solvability. Across three experiments, we discovered effects of prior probability of solvability and of problem-size (number of distinct problem states) on measures of effort and confidence. If a problem is more likely to be solvable, and allows more problem states then a problem solver spends longer trying to solve the problem. Giving-up decisions are informed by the same judgments of probability of success and costs of solution that inform move-choice in a rational model of problem solving.

Giving Up Problem Solving

In the everyday world and in the classroom, problem solvers may fail to complete the problems they attempt. Some problem-solving episodes end with the problem solver abandoning the current problem to move to another, or to do something else. We contend that an understanding of quitting is an important goal for a psychology of human problem solving, yet to our knowledge it has rarely been investigated. It is important not only because it is a widespread everyday phenomenon, but also because rational decisions to quit are an important component of effective behavior – there is little to be gained from continued work on a problem that you will never solve.

Giving up also plays an important role in successful problem solving. The decision to give up on one approach for another can be a critical determinant of success. For example, the model developed by MacGregor, Ormerod and Chronicle (2001) to explain performance on the nine-dot problem has the abandonment of hill-climbing as a crucial step in the insight process. According to their account, the unsuitability of hill-climbing explains the difficulty of insight problems: its failure sometimes prompts potential solutions or “insight”. MacGregor et al. (2001) suggest that problem solvers continue until some progress-monitoring criterion is broached. Kaplan and Simon (1990) also sketched a model of insight problem solving in which a strategy was abandoned when “no operators seem to yield progress” (*p*. 377). However, beyond actually exhausting a problem's possible states, these accounts do not provide “stopping rules” for deciding when insufficient progress has been made.

Stopping Rules in Cognition

Looking beyond problem solving per se, there is a recent coincidence of interest in understanding stopping rules in a variety of behaviors. Consider, for example, memory retrieval. If an experimental participant is asked to recall a list of words, say, they will usually be given a fixed time so to do. If, instead, they are free to choose when to stop trying, how do they behave? This question was addressed in a study by Dougherty and Harbison (2007). These authors suggested that the "exit latency" – the time since the last successful retrieval was a sensitive measure of persistence, and was affected by prior measures of motivation and personality, as well as by the difficulty of the task. They also reported that exit latencies decreased as a function of the number of items retrieved and argued this reflected a lower probability that later items would be worth retrieving. Alternatively, Laming (2009) argues that free recall leads to a state where the same item is retrieved repeatedly and that giving up occurs once this state is reached.

In research on decision making, stopping rules have been used to characterise the choice process (e.g. Gigerenzer & Goldstein, 1996). It is hard to relate these rules to classic problem solving tasks but of greater relevance are analyses of the information-gathering phase that precedes many decisions. Browne and Pitts (2004) consider rules that may underpin decision makers' ceasing to acquire information and moving to a decision. Such rules may be based on a priori quantity thresholds, e.g. a fixed number of bits of information, or on cumulative, relational properties, e.g. a decision-maker may search for information until each new item fails to make an above-threshold difference to their knowledge-base.

Stopping rules in Foraging Theory

These simple stopping rules suggested by Browne and Pitts (2004) for information gathering are similar to the rules-of-thumb that have been proposed in the foraging literature to explain how animals decide to abandon one patch for another (see Stephens & Krebs, 1986, Chapter 8). Moreover, the connection between information gathering and animal foraging has been emphasised by Pirolli and Card (1999).

Most work in optimal foraging theory analyses the optimal solution to an animal's problems, such as diet selection or patch leaving. Some work on patch-leaving instead explores heuristics that animals might actually compute. Optimal giving up decisions may require computations of marginal and cumulative rates of return (Charnov, 1976) that are too costly to be realistic, but rules-of-thumb may offer cheap approximations, by tracking time and encounters with food items.

One simple foraging rule would be to leave any patch after a fixed amount of time; another would be to leave after a fixed number of successes. Such rules are adaptive under some environmental conditions (Iwasa, Higashi, & Yamamura, 1981). A more flexible rule is to quit after a certain time since the last success, or to allow a certain amount of time for each patch, but adjust this upwards with every success.

Recent experimental work has investigated whether such heuristics may explain human decisions to abandon foraging-like tasks. Payne, Duggan and Neth (2007), assigned participants two sets of seven letters and asked them to generate as many words as possible in total, working on the sets of letters in sequence and switching between them as they preferred. The word-finding task is quite similar to the situations analysed in foraging in that the problem-solver is continuously accumulating units of a particular currency (i.e. words instead of food items). Payne et al.'s (2007) analysis of giving up therefore exploited concepts from foraging theory, such as time since last success and number of items encountered or generated (see also Hutchinson, Wilke & Todd, 2008; Wilke, Hutchinson, Todd & Czienskowski, 2009).

Payne et al. (2007) found that a rule based on time since last success – (what foraging theorists call "giving-up time" but Dougherty and Harbison referred to as "exit latency") could not account for their participants' switch decisions. Instead they modelled participants' behavior with a more complex heuristic that is sensitive to the rate at which items are generated, in concert with a probabilistic decision to switch tasks immediately after a subgoal success.

Obviously, classical problem solving tasks do not have the continuous accumulation character of foraging. Instead, their gain curve may be considered to be a step function, with success determined discretely by accomplishment of a single goal state. Nevertheless patch-leaving heuristics might apply, if problem solvers treat the attainment of certain problem states (perhaps all new problem states) as the accumulated currency. Thus a problem solver could quit a problem after a fixed amount of time, or after a fixed number of new problem states, or when the time since the last new problem state exceeds a threshold.

Rational analysis of problem solving

A different account of giving up problem solving might begin with an analysis of rational choice of moves. Anderson (1990) proposed that many important regularities in human cognition could be understood by assuming the mind is adapted to statistical properties of its environment, and by comparing behavior with optimal solutions to information-processing problems (as in foraging theory). Anderson's (1990) rational analysis of problem solving sketches a plausible account of how an optimal problem solver should approach problems, which incorporates an analysis of when to give up (although the quit decision is not emphasised in Anderson's work).

According to the analysis, a rational problem solver should choose among partial plans so as to maximise PG-C, where P is the probability of achieving a goal if a behavior is enacted, G is the gain or value assigned to that goal and C is the cost that will be incurred to achieve the goal, given the behavior. It is assumed that PG-C is estimated for a set of alternative problem solving plans, and used to choose among them, with a Bayesian update of the probability and cost evinced by the new problem states that are thus visited, and a monitoring of effort expended so far. The currencies for G and C are not specified, nor is the construct of effort. To simplify, one might consider time as the currency for all three, with the Gain of a problem being the total length of time it would be worth investing in order to solve it (see also Gray, Sims, Fu & Schoelles, 2006). The starting value of P depends on an assumed distribution incorporating a parameter for the a priori probability that the problem is solvable.

Anderson (1990) applies this rational analysis of problem solving to a handful of classic 'puzzles' from the problem solving literature: 'Missionaries & Cannibals' (e.g. Jeffries, Polson, Razran & Atwood, 1977); Water Jars (e.g. Atwood & Polson, 1976) and Tower of Hanoi (e.g. Kotovsky, Hayes & Simon, 1985). The theory is general to any problem solving situation, but the analysis is best developed when problem solvers use the hill-climbing heuristic, as Anderson assumes is true, with strong backing from the experimental literature cited above, for Missionaries & Cannibals and for Water Jars, the latter being the problems studied in this article. In such cases, the partial plans referred to above are single moves, and PG-C for each move is estimated based on an assessment each possible next state's similarity to the goal (for the water jars problem this is operationalised as the arithmetic difference of jar-contents between the state and the goal state) and its novelty (i.e., how many times the problem solver has already visited the state).

Our interest is in how readily this model accounts for giving up. The key claim is this: a rational problem solver should give up when estimated PG-C approaches zero for all considered partial plans. An elegant aspect of this analysis of giving-up is that it is completely integrated with problem solving per se. Deciding to give up is simply a special case, using the same environmental and metacognitive information as is used to decide between moves. (It is true that both prior probability and effort-so-far have an effect on the PG-C estimates of ALL competing moves: they nevertheless influence move-choice by affecting the relative influence of the similarity to the goal and the novelty factors).

Experiments and Predictions

The experiments in this article all use computerised versions of Water Jars problems, in which water must be transferred among a set of jars to obtain a goal quantity in one or two of the jars. Our main experimental innovation is simple – to present participants with unsolvable problems (that they know *might* be unsolvable) and measure the effort expended (in terms of time and number of moves) before giving up.

In the problems studied by Luchins (1942), there were three jars of different capacities, all of which began empty. The goal was to achieve a particular quantity in the largest jar, and the operators were to fill jars (from a tap), empty them, or transfer water from one jar to another until the destination jar was full. We call this type of problem "open".

In the studies modelled by Anderson (1990; Atwood & Polson, 1976), there were again three jars, but there was no tap, and no sink. Instead, the largest jar began full, the other two jars empty. The only operators were transfers and the goal was to divide the water equally among the two largest jars. We call this type of problem "closed".

In closed problems, the total amount of water distributed among jars is constant, and consequently the number of problem states is fewer than in corresponding open problems. Because exhausting problem states is a plausible stopping rule, we wanted to be sure to set some problems where this was unlikely to happen by default. Thus we begin our investigation with open problems.

We draw our experimental predictions from the PG-C account of problem solving. We contrast these predictions with those made by simple heuristics borrowed from the foraging literature.

The PG-C model predicts how much effort a problem solver will expend (i.e., how long they will persist) before quitting, in the face of failure. First, the model predicts an effect of the problem solver's estimated likelihood that the problem is solvable. A higher prior expectation that the problem is solvable will increase the value of P for all moves, making PG-C above zero for more moves for longer as problem solving continues. We test this prediction by presenting participants with a set of problems, and informing them, honestly, that a certain proportion will be unsolvable. By manipulating the proportion, we manipulate the prior probability. When a participant attempts an unsolvable problem, they have to give up, and the total time to reach this decision is one dependent variable (another measure of expended effort is the number of moves made – although this may not be a very good index of mental search, O'Hara & Payne, 1998). This prediction may be challenged by the finding that decision makers sometimes fail to make sufficient allowance for prior probabilities in the face of individual information about the current problem (Kahneman & Tversky, 1973)

The PG-C model also predicts that total time before quitting will increase when there are more problem states to consider. As noted above, PG-C is judged higher for states that have been less visited previously. Thus, where there are more new states available, PG-C is less likely to approach zero, which translates into a simple prediction that problem solvers should take longer to give up when the state space is bigger. This hypothesis is tested by manipulating the size of the state space of unsolvable problems in Experiments 1 and 2.

Whenever a participant abandons a problem, we ask them how confident they are that the problem is unsolvable. Throughout this article, for clarity in the face of a judgment that inverts readers' conventional expectations, we refer to this judgment as confidence-unsolvable.

The PG-C model does not come with a ready-made theory of confidence, but one intuitively clear assumption is that it will directly reflect the problem solver's current estimations of P (confidence-unsolvable will reflect 1-P). The model supposes that problem solvers estimate P for a set of partial plans, and quit when PG-C approaches zero for all such partial plans. In making a single judgment about the problem, the solver must somehow reduce these estimates of P, perhaps using the maximum value or the P of the move with highest PG-C. Given this reading of confidence, what does the PG-C model of problem solving predict about post-giving-up confidence-unsolvable?

According to the model, as the problem solver makes moves through the problem space, P and C are estimated for each new plan (each move, in the case of hill-climbing) that is considered. In particular P will increase and C will decrease according to both similarity to the goal and novelty of the next state. Furthermore, effort-so-far (which may be approximated by time spent problem solving) will have a moderating effect on both P and C, decreasing P and increasing C for all moves (see Anderson 1990, pp. 212-213). When the prior probability of solving a problem is low, PG begins lower, and PG-C will approach zero sooner for all moves in unsolvable problems, and therefore before C has had as much time to grow. Consequently, the model predicts that when prior probability of solution is lower, problem solvers will quit at lower levels of P and thus report higher levels of confidence-unsolvable.

This analysis exposes that the effect of low prior probability will be similar to the effects of a low G (i.e., a low evaluation by the problem solver of the worth of solving a problem). The model thus explains what to us seems intuitively correct - if a participant is less motivated to solve the problem, they will work at it for less time and report lower confidence-unsolvable when they quit. This tradeoff between time and confidence that is determined by participants' individual level of motivation will likely add noise to our experimental data.

There is no similar prediction for an effect of problem size on reported confidence unsolvable. In bigger problems, PG-C will tend to reduce more gradually, because revisits to states will be less frequent; but such an effect will be due to both a decrease in P and an increase in C, so it is not clear why it should lead to any change in the value of P at which PG-C approaches zero.

Do any models make competing predictions? The foraging heuristics reviewed above offer a basis for prediction, if one allows an alternative conception of confidence as a post-hoc judgment about the problem, given the problem solving episode.

First, it seems quite plausible that people may utilise the exit-latency heuristic, quitting a problem after a certain time has passed since any new problem states have been discovered. This heuristic allows a problem solver to judge that the problem is exhausted; it is similar to Browne & Pitts's (2004) "difference threshold", and has also been considered for stopping memory retrieval by Dougherty & Harbison (2007). Use of an exit latency threshold predicts that time and number of moves will increase with problem size, but predicts no effect of prior probability. Assuming that sufficient extra time is given to bigger problems to reach a similar judgment of state-exhaustion, there should be no effect of problem-size on confidence-unsolvable. It is hard to infer a prediction for the effect of prior probability on confidence.

In the foraging literature, the exit latency heuristic has sometimes been viewed as a very simple way of computing marginal rates of return, thus allowing a simple approximation to the optimum behavior specified by Charnov's marginal value theorem. However, Iwasa et al (1981) have shown that using a simple threshold for total time or total number of items can actually improve rates of gain relative to exit-latency in certain environments, depending on the distribution of rewards in those environments. These rules therefore seem worth considering in an experimental situation, where participants may make simple assumptions about the variation in problems they are likely to encounter.

So, each participant could simply allocate a fixed amount of time to each problem. If such a time budget determines giving up then there will be no effects of problem size or of prior probability on time before quitting. However, because in a fixed time the smaller problem will be explored more thoroughly, with more visits to repeated states, this stopping rule predicts that confidence-unsolvable will be higher for smaller unsolved problems. Prior probability may additionally affect confidence, the account is neutral on that matter.

A second very simple stopping rule assumes problem solvers will quit after a fixed number of new, distinct states have been found without the goal being achieved. Such a rule would lead to problem solvers spending as long on smaller problems as on larger problems, or possibly longer (if new states are harder to find). There will be no effect of prior probability. Confidence-unsolvable will be higher for smaller problems, if solvers can somehow estimate the number of unvisited states.

Experiment 1

Method

Design

This study was a 2 × 2 between-subjects design. The first factor had two levels. For half the participants one of the four problems they were asked to solve was unsolvable, for the other half, three of four were unsolvable. Participants were informed of these proportions before they attempted the first problem. We label this factor "Prior Probability". The second factor was the number of states in the problem-space of the first problem, which in fact was always unsolvable (participants did not know that problem order was fixed); this factor also had two levels, and will be referred to as "Problem Size".

Participants

Forty eight undergraduates from Cardiff University were each paid £5.00 (US $7.91) or offered course credit to participate.

Materials

This experiment used open water jar problems. The capacity of the three jars decreased from B to A to C. Each jar began empty, and operators for filling and emptying each jar were provided. The goal state was a quantity in jar B, with the final quantity of the other jars irrelevant. We denote problems by the capacities of the three jars followed by the goal quantity for jar B. The "small" problem was A=18, B=36, C=9, Goal=14 and had 30 states in it's problem space; the "big" problem was A=15, B=36, C= 9, Goal=14 (218 states).

Problems were presented on a computer display (programmed in Hypercard), pictured in Figure 1. In the centre of the screen were three rectangles representing the three jars. Above each jar was displayed its capacity, and within each jar was displayed its current contents. Below the jars were a set of twelve buttons, one for each of the six possible transfer operators. A Restart button allowed participants to return to the starting state of the current problem.

Procedure

The procedure was implemented on a Macintosh computer which stepped participants through the procedure, self-paced. The experiment began with an explanation of the water jars problem, followed by a practice problem (A=14, B=50, C=11, Goal=33) for participants to solve. A solution sequence was presented in the top left of the screen, so that participants did not have to work out the solution, although they could if they wished. In either case the solution had to be entered interactively in exactly the same way as with the experimental problems.

Once participants had solved the practice problem they were informed that they would be asked to try to solve four problems, and informed the appropriate proportion of problems that were solvable (1 or 3 out of 4, depending on experimental condition).

The first, unsolvable problem was followed either by three solvable problems or by two more unsolvable problems and then a single solvable problem. The order of problems was fixed. Data from the final three problems are not considered in this article.

Each problem screen contained an “Abandon” button, which participants used to move to the next problem. Participants were instructed that problems must be attempted in strict sequence and that no problem could be re-visited. Participants were given a small cash award for all successfully solved problems [£1 (US $1.58) per problem].

Whenever a participant abandoned a problem, they were presented with a question: “How confident are you that the problem you just abandoned really is unsolvable? Please click on one of the buttons below to indicate your confidence.” Eleven response buttons were numbered from -5: “Although I abandoned the problem I am certain it is solvable” to 5: “I am certain the problem is unsolvable”.

Results

Untransformed means and standard deviations for each variable under each experimental condition are shown in Table 1. No participants visited all problem states of either problem. The analyses all relate to the first problem that participants attempted, which was always unsolvable. Three main dependent measures of performance are considered: time to give up, number of moves, post-hoc confidence that the problem really was unsolvable.

For each dependent variable we consider the effects of Problem Size and Prior Probability using between-subjects 2 × 2 ANOVA. Before ANOVA, times and number of moves were log-transformed [ln(N+1) was used as 8 participants made no moves]. These transforms made no difference to the pattern of effects.

Both time and number of moves were greater for the large problem space than for the small problem space [Time, $F(1,44) = 5.48$, $MS_e = .04$, $p < .05$, $\eta^2_p = .11$; Moves, $F(1,44) = 4.86$, $MS_e = .28$, $p < .05$, $\eta^2_p = .10$]. A high prior probability that the problem was solvable lengthened the time spent solving and increased the number of moves but neither of these effects were significant [Time, $F(1,44) = 1.16$, $\eta^2_p = .03$; Moves, $F(1,44) = .58$, $\eta^2_p = .01$]

Confidence-unsolvable was higher for the small problem space than the large problem space, $F(1,44) = 4.38$, $MS_e = 4.48$, $p < .05$, $\eta^2_p = .09$, and was higher for low prior probability that the problem can be solved, $F(1,44) = 8.25$, $MS_e = 4.48$, $p < .01$, $\eta^2_p = .16$.

None of the analyses revealed significant interactions (*F*s < 1).

Discussion

Size of problem space exerted a significant effect on both primary dependent variables – the larger problem led to more time and more moves before the giving-up decision. The effect of prior probability was in the predicted direction, but did not reach significance for either time or number of moves.

The results for confidence-unsolvable were significant and imply a direct effect of both independent variables upon confidence. A larger problem size and higher prior probability of solvability resulted in lower confidence-unsolvable ratings after giving up. The effect of prior probability is in line with the predictions of the PG-C account, if one assumes that confidence ratings directly reflect P (strictly, 1-P), but the effect of problem size on confidence-unsolvable appears to be outside the scope of the current theory.

The ideas that participants may give up after a fixed time or after a fixed number of new states do not predict an effect of problem size or prior probability upon time or number of moves. The data count strongly against these two simple models.

However, the idea that participants will use an exit-latency threshold is offered support by the data. Indeed this account can also explain the effect of problem size on confidence-unsolvable. Clearer evidence against the exit latency model would be reliable effects of prior probability on time and number of moves.

The absence of a significant effect of prior probability on total time is the main discrepancy between our results and the PG-C account. It is worth noting that the open water jars used in this study may afford a wider variety of strategies than do the "closed" variety that was studied by Atwood and Polson (1976). For example, it is possible to search mentally for algebraic combinations of the jars' capacities that achieve the target quantity. The PG-C account should still apply to such strategies; nevertheless a mix of strategies among participants will weaken the statistical power of the experiment, especially in light of the argument above that individual variation in G likely allows effort (time) and probability (reported confidence-unsolvable) to be traded off to some extent.

Experiment 2

To address this concern the second experiment used “closed” water jars (i.e. problems in which the only operators were transfer of water from one jar to another – jars could not be filled, nor could water be discarded). Algebraic strategies are less likely with closed jars, because fewer simple algebraic combinations of quantities are readily available. For example, in the open jar problem A=15, B=36, C= 9, it is easy to create the quantity 2A - C in B by filling A and transferring it to B twice, then filling C from B. It is similarly simple to create A+C, B-C, B-2C etc. When the total quantity of water is fixed, many fewer of these simple expressions can be achieved as subgoals. We hope this might reduce variation in strategy choice and increase the power of the experiment.

Method

Design

The design was as for Experiment 1. The precise problems were different, but the factors and their relative levels were unchanged.

Participants

Fifty two undergraduates from Cardiff University were paid £5.00 (US $7.91) or offered course credit to participate.

Materials

The experiment used “closed” water jar problems. In all problems used in this experiment the capacity of the jars decreased from A to B to C. The capacity of jar A was an even number of units. In the initial state A was full and B and C were empty, in the goal state the water from A was evenly distributed between A and B. Because of this fixed formula, problems can be specified by listing the capacities of the three jars.

Problems were presented on an interactive computer display as for Experiment 1 but without the Fill-and Empty-jar buttons.

Each participant attempted four problems, with the set of problems varying with experimental condition. The first problem was always unsolvable, and the experimental hypotheses relate to this problem. For half the participants the problem was A=10, B=9, C=8, which has 10 states in its problem space, for the remainder it was A=12, B=9, C=7 (25 states).

Procedure

The procedure was exactly as in Experiment 1, except for the difference in problems noted above, as well as a different practice problem, (A=18, B=11, C=5).

Results

Table 2 shows the untransformed means and standard deviations for all dependent measures according to experimental condition. Only 3 participants visited all the states of the smaller problem. Analyses were conducted as for Experiment 1.

Before ANOVA, times and number of moves were log-transformed [for number of moves, ln(N+1) was used as 1 participant made no moves]. These transforms made no difference to the pattern of effects.

Both time and number of moves were greater for the large problem space than for the small problem space [Time, $F(1, 48) = 10.75$, $MS_e = .04$, $p < .01$, $\eta^2_p = .18$; Moves, $F(1,48) = 9.20$, $MS_e = .10$, $p < .01$ $\eta^2_p = .16$]. A high prior probability that the problem was solvable lengthened the time spent solving, $F(1, 48) = 5.93$, $MS_e = .04$, $p < .05$, $\eta^2_p = .11$, and increased the number of moves, although this effect was marginal, $F(1,48) = 3.66$, $MS_e = .10$, $p = .06$, $\eta^2_p = .07$.

Confidence-unsolvable was higher for the smaller problem space and for the lower prior probability of solvability but these effects were not significant [Problem Size, $F(1,48) = 1.04$, $\eta^2_p = 02$; Prior Probability, $F(1,48) = 1.76$, $\eta^2_p = .03$].

No interactions were significant (*F*s < 1.6).

Discussion

Both main predictions for the rational PG-C model were supported. When problems were large and prior probability of solvability was high participants took longer to give up. The effects on confidence-unsovlable that were reported in Experiment 1 were not replicated, although the trends were in the same directions.

The results across both experiments count against all the simplest stopping rules. The effect of problem size upon time to give up directly contradicts the hypothesis that participants give up after a fixed time or a fixed number of new states. The effect of prior probability on time to give up in Experiment 2 suggests that the decision to quit is not entirely a matter of judging the problem to have been exhausted.

The main inconsistency in the results of the first two experiments is that the effect of prior probability is less robust than we expected. The establishment of this effect with closed problems supports our speculation that strategies for open jar problems may be more various. Additionally, the predicted effect of prior probability on confidence is unstable: reliable in Experiment 1 but not Experiment 2. Finally, a limitation of the method in the first two experiments is that moves were not time-stamped, preventing a direct test of the exit-latency rule. The next experiment addresses all these issues.

Experiment 3

To gain more insight into the relation between confidence judgments and giving-up decisions, and into the effects of prior probability in the two separate problem types we asked participants to provide regular ratings of confidence-unsolvable as they solved each problem (in addition to post-giving up confidence-unsolvable ratings). Following Metcalfe (1986) we asked for these ratings every fifteen seconds. Prior probability was manipulated as before but only the larger problem of each type was used. Problem size was not manipulated as it produced a reliable effect on time to give up for both open and closed jar problems. The larger problems were chosen to provide longer problem solving protocols.

Our theoretical assumption is that reported confidence-unsolvable reflects participants' judgments of the problem during problem solving. But in the first two experiments, by asking for these judgments only after a quit decision, we have run the risk that the judgments will be adjusted post-hoc, using remembered aspects of the problem solving experience. By instead collecting judgments every fifteen seconds we maximise the chances of reports reflecting just those metacognitive states that are active in the problem solving process; we can also test the assumption that confidence-unsolvable will increase as problem solving continues. Additionally, by prompting participants to focus on the probability that the problem is solvable, we hope that the regular confidence ratings may encourage participants to work for longer on problems, so as not to quit having just reported low confidence-unsolvable. This would effectively increase G, in that quitting would only occur when P is low and confidence-unsolvable is relatively high.

Further, in this experiment we time-stamped all interactive behavior. This additionally allowed us to look at the temporal distribution of new-state visits during each problem solving episode. From these data we can directly ask whether participants' behavior is well-characterized by an exit-latency heuristic – quitting once a threshold of elapsed time since a new state has been passed.

Method

Design

Once again a 2 × 2 between-subjects design was employed. One factor was "Prior Probability", manipulated in the same way as in Experiments 1 and 2. The other factor was the "Problem Type" as half the participants were given an Open problem as in Experiment 1 and the other half received a Closed problem as in Experiment 2.

Participants

Sixty four students from the University of Manchester were each paid £5.00 (US $7.91) for participation in the experiment.

Materials

A Visual Basic program replicated the interfaces described in Experiments 1 and 2. Every time the participant clicked a button the program recorded and time-stamped the event. All interactions with the program were recorded and time-stamped. The problems with a large problem size from Experiments 1 and 2 were used. Thus, the Open jar problem was 15, 36, 9, 14 and the Closed jar problem was 12, 9, 7.

Procedure

Participants were informed that while attempting a problem they would be interrupted every 15 seconds and asked to make a judgement about how confident they were that the problem was unsolvable. To familiarise participants with this process they were given a computerised 5 disk Tower of Hanoi problem and interrupted in the same way as during the experimental problem. The Tower of Hanoi problem was terminated by the program after 2 minutes (none of the participants solved the problem in this time).

Participants were then given instructions, a practice water jar task (in which they also were interrupted to practise giving confidence ratings) and the Prior Probability manipulation in the same way as in Experiments 1 and 2 before attempting the experimental problem.

The program interrupted participants 15 seconds after beginning the experimental problem and asked participants to make a confidence-unsolvable judgement. When they had completed this judgement the program returned to the problem solving task for another 15 seconds before interrupting again in the same way. This process continued until participants abandoned the task whereupon they were required to make another final confidence-unsolvable judgement.

During interruptions the interface displaying the problem was replaced by the question requiring a confidence-unsolvable judgement. This question was the same as in Experiments 1 and 2. For the second interruption and all subsequent confidence-unsolvable judgements the confidence level from the immediately preceding judgement was highlighted in yellow.

After abandoning the first problem participants were not given the remaining 3 problems and instead were debriefed as to the purpose of the experiment.

Results and Discussion

Because of their different structure, closed and open jar problems were analysed separately throughout. The effects of prior probability on time, number of moves, and confidence-unsolvable were examined using independent t tests. Times and number of moves were log-transformed before conducting analyses but, as in the previous experiments, this made no difference to the pattern of effects. All times reported in this section were log transformed before analysis and all significant ($p < .05$) comparisons or correlations are included.

Time spent making the confidence judgements every 15 seconds was excluded from time measures; however, inclusion of these time data did not affect the pattern of significance. For the initial set of analyses "confidence-unsolvable" was the final confidence judgement that was made after abandoning the task.

The means and standard deviations for each condition are given in Table 3. Higher prior probability of solvability lengthened problem solving time for both problem types [Closed jar, $t(30) = 2.10$, $SE = .07$, $p < .05$, $\eta^2_p = .13$; Open jar, $t(30) = 3.51$, $SE = .11$, $p < .01$, $\eta^2_p = .29$] and increased the number of moves for both problem types, although the effect on moves was marginal [Closed jar, $t(30) = 1.71$, $SE = .10$, $p = .098$, $\eta^2_p = .09$; Open jar, $t(30) = 2.04$, $SE = .15$, $p = .051$, $\eta^2_p = .12$.].

Confidence-unsolvable after giving up was higher for low prior probability of problem solvability but this effect was not significant for the closed jar problem, $t(30) = 1.37$, $\eta^2_p = .06$, or the open jar problem, $t(30) = 1.11$, $\eta^2_p = .04$.

For both problem types there was an effect of prior probability on time and number of moves but no effect upon final judgments of confidence-unsolvable. For the closed jars this replicates the findings from Experiment 2, whereas for the open jars it replicates the direction of the trends, but different comparisons reached significance. In Experiment 1, there was a reliable effect of prior probability on confidence-unsolvable but not upon time or number of moves. The main difference in procedure between Experiments 1 and 3 was that participants were required to make confidence ratings throughout problem solving in Experiment 3. Because participants used these ratings to reflect changes in confidence throughout a problem solving episode, and because they had no way of estimating how long this episode would last, we suspect that the effect on post-hoc confidence unsolvable is compromised by ceiling effects. This argument is supported by the fact that confidence-unsolvable scores after giving up were considerably higher in Experiment 3 than in Experiments 1 and 2.

Confidence Progression over Time

The confidence-unsolvable ratings every 15 seconds are plotted in Figure 2 as means for each of the four conditions. Because participants gave up at different times some participants did not contribute to the means at the later time points. To provide a broad indication of the proportions: 91% of participants made confidence ratings at 75 seconds, 69% at 150s, 34% at 225s and 23% at 300s.

Considering the ratings from all participants who made at least 10 ratings (an arbitrary cut-off leaving at least 9 participants in each cell) , we computed 2 (Prior Probability) * 10 (Time of rating) ANOVAs for closed and open problems with repeated measures on the second variable. In both ANOVAs there were significant main effects of probability and of time. Open problems, Probability: $F(1, 23) = 6.58$, $MS_e = 43.82$, $p < .05$, $\eta^2_p = .22$; Time: $F(9, 207) = 8.74$, $MS_e = .72$, $p < .001$, $\eta^2_p = .28$; Interaction $F(9, 207) = 1.27$, $MS_e = .72$, $p = .077$, $\eta^2_p = .07$. Closed problems, Probability: $F(1, 19) = 5.30$, $MS_e = 48.81$, $p < .05$, $\eta^2_p = .22$; Time: $F(9, 171) = 20.16$, $MS_e = .83$, $p < .001$, $\eta^2_p = .52$; Interaction $F < 1$.

The effect of Probability was significant for the first judgement after 15 seconds, Open problems [$t(30) = 2.15$, $SE = .84$, $p < .05$, $\eta^2_p = .13$]; Closed problems [$t(30) = 3.61$, $SE = .67$, $p < .01$, $\eta^2_p = .30$] and for the tenth judgement after 150 seconds, Open problems [$t(23) = 2.83$, $SE = 1.07$, $p < .01$, $\eta^2_p = .26$]; Closed problems [$t(19) = 2.47$, $SE = 1.00$, $p < .05$, $\eta^2_p = .24$]. Overall these effects offer support for our hypothesis that initial confidence-unsolvable will be affected by prior probability and that confidence-unsolvable will increase as problem solving continues.

To further understand the *process* of giving up we also consider measures of task performance over time and their relation to giving up.

Time-course measures and quit decisions

Water jar problems do not have such an unambiguous measure of task performance as do the word-finding tasks used by Payne, Duggan, & Neth (2007). However, Atwood & Polson (1976) demonstrated that for closed jar problems finding a new "state" or combination of water across the three jars is treated as the primary subgoal. We therefore used "new states found" as a proxy for task performance and used this measure to relate performance on the water jar problems to the decision to quit the task.

Table 4 provides descriptive statistics of some of the measures analysed in Payne et al. (2007). *Exit latency* refers to the amount of time between the last new state found and the decision to quit. Also included is the longest interval between finding new states – all participants had at least one between-states interval that was longer than their time to make a first move.

For both problem types the effect of prior probability of solvability upon number of new states found and exit latencies was tested. Higher prior probability of problem solvability meant that more new states were found for both the closed jar problem, $t(30) = 2.66$, $SE = 1.06$, $p < .05$, $\eta^2_p = .19$, and the open jar problem, $t(30) = 2.36$, $SE = 5.35$, $p < .05$, $\eta^2_p = .16$, and that exit latencies were longer for the open jar problem, $t(30) = 2.60$, $SE = .15$, $p < .05$, $\eta^2_p = .18$.

The exit-latency heuristic was tested by comparing the exit latency for each participant with their longest time to find a new state. In all four conditions the longest between-states interval was longer than the exit latency: High Prior Probability closed jar, $t(15) = 2.15$, $SE = .13$, $p < .05$, $\eta^2_p = .24$; Low Prior Probability closed jar, $t(15) = 2.61$, $SE = .16$, $p < .05$, $\eta^2_p = .31$; High Prior Probability open jar,

$t(15) = 3.15$, $SE = .11$, $p < .05$, $\eta^2_p = .40$; Low Prior Probability open jar, $t(15) = 5.77$, $SE = .09$, $p < .001$, $\eta^2_p = .69$. Therefore, these data appear to rule out a simple heuristic based on exit latency (this was also the conclusion of Payne et al., 2007; Hutchinson et al., 2008 and Wilke, Hutchinson, Todd & Czienskowski, 2009).

General Discussion

Across three experiments we have gathered reliable evidence that the decision to give up problem solving is affected by the size of the problem and by the prior probability that the problem is solvable.

That problem solvers take longer to quit a bigger problem can be explained by a rational model: when a problem contains many intermediate states, problem solvers' considered partial plans are more likely to be judged as worth pursuing. In particular, when the hill-climbing strategy is used, single-step plans in a bigger problem are more likely to lead to novel states, which receive higher estimates of P and lower estimates of C. Thus, if giving-up decisions reflect PG-C approaching zero for all currently applicable moves they will be delayed when the problem space is bigger.

The prior probability effect can be explained by the same model: when the prior probability of a problem being solvable is lower, all subsequently updated estimates of PG-C for partial plans are reduced. As a direct consequence, a situation where PG-C approaches 0 for all partial plans is reached in fewer moves. These effects are therefore consistent with the idea that problem solvers are continually monitoring the quality of next-states in terms of the probability of them leading to the goal, as assumed by Anderson's rational analysis of problem solving by hill climbing.

The results for confidence-unsolvable are less stable. When participants were asked to make confidence ratings every 15 seconds (Experiment 3) these intermediate ratings behaved in a way consistent with our assumption that they would reflect the same judgments of probability of success that underpin move-choice and giving-up decisions. That is to say, initial confidence-unsolvable is affected by prior probability and increases with experience with the unsolvable problem. Similarly, the tendency for final post-giving up ratings of confidence-unsolvable to be affected by prior probability is in line with predictions, but these effects are only significant in Experiment 1, where there was an additional effect of problem size on confidence-unsolvable. Perhaps this is a simple matter of experimental power, together with the measurement noise in how people interpret the confidence scales. Nevertheless our data show that confidence judgments hold some promise as a way of tracking the timecourse of quit decisions.

Our data count against some of the simplest stopping rules that might have been supposed to be operational in the decision to quit a problem. Participants evidently do not simply allow themselves a fixed amount of time to succeed, or a fixed number of new intermediate states, nor do they always keep working on a problem until they have exhausted all the possible states (Experiments 1, 2, 3). Participants do not set a simple exit-latency threshold – quitting a problem after a certain time in which no new states are discovered (Experiment 3).

One attractive aspect of the account of giving up problem solving that we have sketched is that it is integrated with a rational account of move-choice during ongoing problem solving. The account supposes that no special reasoning is required to inform the giving-up decision. Rather, this decision is a by-product of the evaluation of moves: A problem solver quits when the evaluation of every considered move fails to justify the cost of its enactment.

Note, however, that this estimation of probabilities and costs of particular partial plans incorporates some global parameters of the problem, such as prior probability of solvability, and of the problem solving episode, such as effort expended. These parameters affect the judgment of all partial plans and may play a relatively minor role in the choice between partial plans (although, according to the rational analysis, such a role does exist).

An interesting possibility which might be explored in future research is that overall effort may not in fact be the most appropriate global summary of the problem solving episode by which to adjust judgments of probability and cost of success. Perhaps, instead, some of the slightly more complex foraging heuristics for estimating 'patch potential' would work better in this role. For example, effort so far might be moderated by some measure of success (such as how many new states closer to the goal have been found) as it is in foraging heuristics like Green's rule (Green, 1984; Payne, Duggan & Neth, 2007).

A similar refinement to the PG-C theory is suggested by our finding (Experiment 1) that confidence-unsolvable is higher for smaller problem spaces (although it must be borne in mind that this effect was not replicated in Experiment 2). In the current version of the theory, both similarity to the goal state and the frequency of visitation affect estimates of both P and C for next states. But it seems possible that the number of times a state has been visited should decrease estimates of the probability of success from that state more than it should increase estimates of cost of success. Cost of success may rather be dominated by the similarity of the state to the goal. If this were the case, if PG-C is to approach zero because of the number of revisits to a state, this will require relatively low levels of P. This might explain why confidence-unsolvable is higher for smaller problems.

In summary, then, the current experiments favour an account of giving up as a special case of move-choice in problem solving, and count against the simplest stopping heuristics that have been considered in foraging theory. Future work may profitably consider a fuller integration of these two approaches, for example, by allowing foraging heuristics to offer summaries of the problem solving episode that play a similar role in the computation of PG-C to 'effort so far' but which are more sensitive to other aspects of problem solving experience than pure effort.

There is much work remaining to develop a full theory of giving up problem solving. Our aim in this article has been to introduce this goal as an important one, and to provide some basic empirical phenomena that constrain future theories; and a suggestion that such developments may fruitfully begin from a rational account of problem solving. Further, we propose that our simple experimental innovation of presenting participants with problems that may or may not be solvable offers one important and efficient way of gathering data in support of this enterprise.

References

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

Atwood, M.E., & Polson, P.G. (1976). A process model for water jug problems. *Cognitive Psychology, 8,* 191-216.

Browne, G. J., & Pitts, M. G. (2004). Stopping rule use during information search in design problems, *Organizational Behavior and Human Decision Processes, 95,* 208-224.

Charnov, E. L. (1976). Optimal foraging: the marginal value theorem. *Theoretical Population Biology, 9* 129-136.

Dougherty, M. R., & Harbison, J. I. (2007). Motivated to retrieve: How often are you willing to go back to the well when the well is dry. *Journal of Experimental Psychology: Learning, Memory & Cognition, 33*, 1108-1117.

Gigerenzer, G., & Goldstein, D. (1996).Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461-482.

Green, R. F. (1984). Stopping rules for optimal foragers. *American Naturalist, 123,* 30-43.

Hutchinson, J. M. C., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: Can a generalist adapt its rules to dispersal of items across patches? *Animal Behavior, 75*, 1331-1349.

Iwasa, Y., Higashi, M., & Yamamura, N. (1981). Prey distribution as a factor determining the choice of optimal foraging strategy. *American Naturalist 117* 710-723.

Jeffries, R.P., Polson, P. G., Razran, L. & Atwood, M. (1977). A process model for missionaries-cannibals and other river-crossing problems. *Cognitive Psychology,* 9, 412-440.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

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

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems so hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248-294.

Laming, D. (2009). Failure to recall. *Psychological Review, 116*, 157-186.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54, 6* (Whole No. 248).

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.

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 623-634.

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.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental*

Psychology: General, 136 (3), 370-388.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review. 106 (4)*, 643-675.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory.* Princeton, NJ: Princeton University Press.

Wilke, A., Hutchinson, J. M. C., Todd, P. M., & Czienskowski, U. (2009). Fishing for the right words: Decision rules for human foraging behavior in internal search tasks. *Cognitive Science, 33,* 497-529.

Acknowledgements

The authors would like to thank Andrew Howes, Michael Dougherty, several anonymous reviewers and most especially Erik Altmann for comments on previous drafts.

Table 1.

Experiment 1: Means and Standard Deviations for Time to Abandon Problem, Number of Moves, Number of States Visited and Confidence that a Problem was Unsolvable as a Function of Prior Probability of Problem Solvability and Problem Size.

Prior Probability Problem Solvable	Problem Size (No. of states)	Time to Abandon Problem (s)		Number of Moves		Number of States Visited		Confidence Unsolvable	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
.75	30	198.28	108.80	11.17	12.55	6.08	5.00	0.48	2.44
.75	218	273.78	113.10	27.00	18.85	15.75	11.30	−0.54	2.09
.25	30	171.00	70.25	13.42	15.58	6.17	5.37	2.50	1.81
.25	218	213.80	67.57	14.00	10.15	8.92	5.66	0.95	2.07

Table 2.

Experiment 2: Means and Standard Deviations for Time to Abandon Problem, Number of Moves, Number of States Visited and Confidence that a Problem was Unsolvable as a Function of Prior Probability of Problem Solvability and Problem Size.

Prior Probability Problem Solvable	Problem Size (No. of states)	Time to Abandon Problem (s)		Number of Moves		Number of States Visited		Confidence Unsolvable	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
.75	10	166.33	72.90	26.46	18.40	8.92	1.85	0.69	2.95
.75	25	285.42	149.55	36.00	19.38	10.69	3.90	0.31	2.46
.25	10	149.93	108.85	18.46	18.25	6.85	3.24	2.08	3.01
.25	25	181.43	63.02	30.77	17.44	9.31	3.12	0.92	2.40

Table 3.

Experiment 3: Means and Standard Deviations for Time to Abandon Problem, Number of Moves and Confidence that a Problem was Unsolvable as a Function of Prior Probability of Problem Solvability for Open and Closed Water Jar Problems.

Problem Type	Prior Probability Problem Solvable	Time to Abandon Problem (s)		Number of Moves		Confidence Unsolvable	
		M	*SD*	*M*	*SD*	*M*	*SD*
Closed	.75	204.49	106.08	30.00	15.76	3.06	2.26
Closed	.25	140.15	49.84	22.63	14.64	3.94	1.18
Open	.75	419.55	242.54	56.69	47.44	2.75	2.82
Open	.25	222.14	247.59	38.38	34.71	3.69	1.89

Table 4.

Experiment 3: Number of states visited, time between last new state and quitting and longest time between finding new states.

Problem Type	Prior Probability Problem Solvable	Number of States Visited		Exit Latency (s)		Longest between states interval (s)	
		M	*SD*	*M*	*SD*	*M*	*SD*
Closed	.75	10.19	3.27	34.02	31.80	59.13	56.65
Closed	.25	7.38	2.68	29.42	35.09	44.78	24.90
Open	.75	28.44	17.93	41.36	51.66	70.09	63.30
Open	.25	15.81	11.70	14.39	13.28	40.69	25.54

Figure Captions

Figure 1. Screenshot of the interface for the open jars problem from Experiment 1.

Figure 2. Experiment 3: Progression of confidence over time with high and low prior probability of solvability for open and closed water jars. Closed = Closed jar problem, Open = Open jar problem, Hi = High prior probability of solvability, Lo = Low prior probability of solvability.

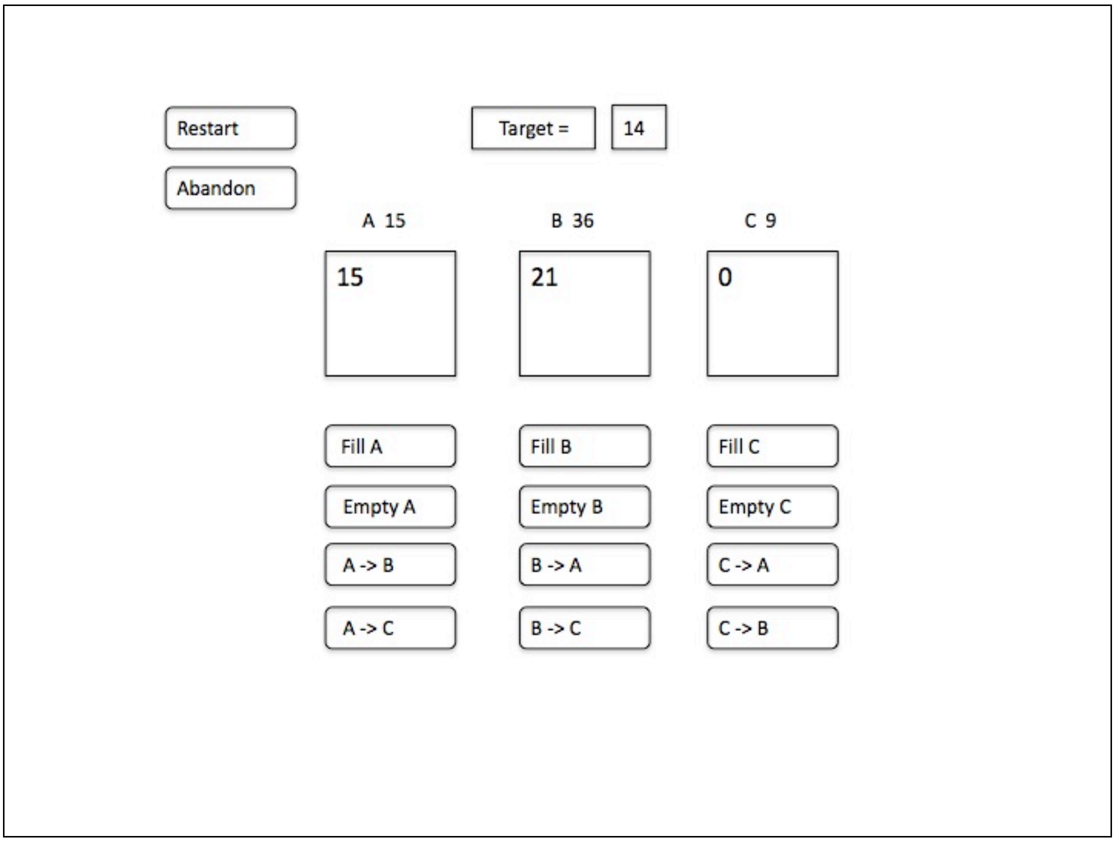
Restart
Abandon
Target =
14
A 15
B 36
C 9
15
21
0
Fill A
Fill B
Fill C
Empty A
Empty B
Empty C
A -> B
B -> A
C -> A
A -> C
B -> C
C -> B

Figure 1

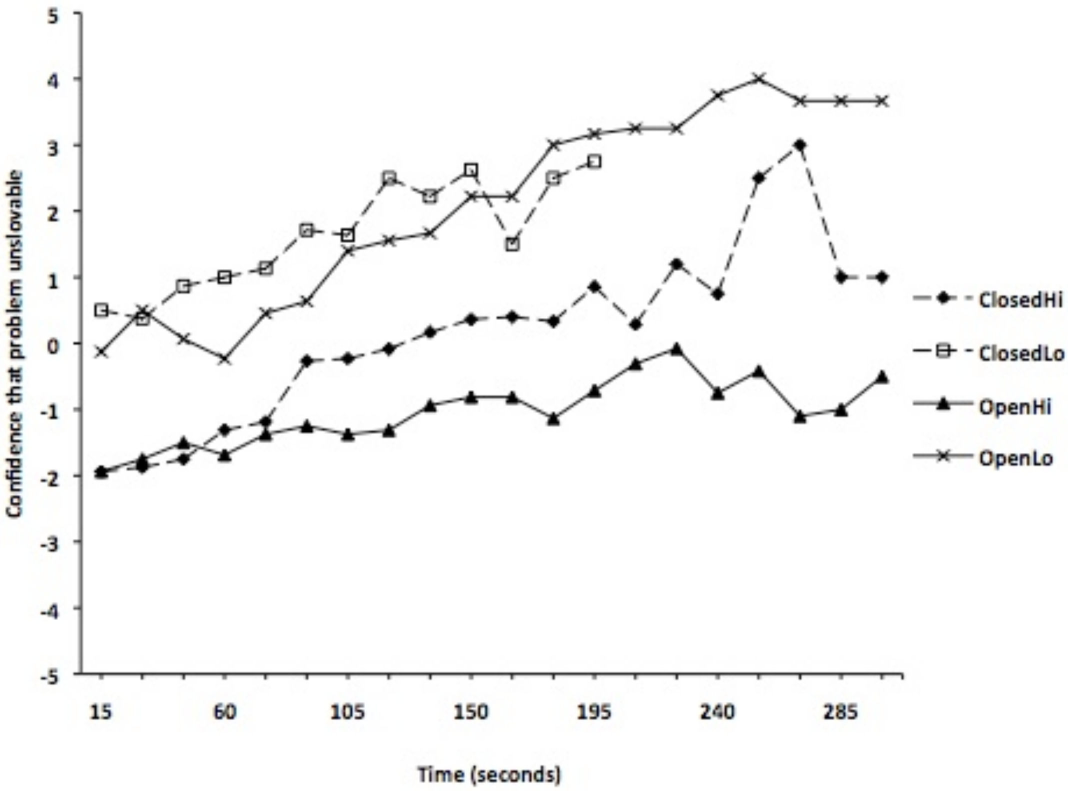
5
4
3
2
1
0
-1
-2
-3
-4
-5
Confidence that problem unslovable
15
60
105
150
195
240
285
Time (seconds)
ClosedHi
ClosedLo
OpenHi
OpenLo

Figure 2