

# Failure Analysis in Backtrack Search for Constraint Satisfaction

Tudor Hulubei and Barry O’Sullivan  
 Cork Constraint Computation Centre, Ireland  
 tudor@hulubei.net, b.osullivan@4c.ucc.ie

## 1 Introduction

We augment our traditional statistics-based approach to studying systematic search with a visualisation method that uses heatmaps, which can be more informative and present different views of the relationship between the depth at which a mistake occurred and the size of the refutation associated with it. We compare search algorithms on the basis of the number of, and effort required to recover from, *individual mistakes*. We highlight interesting differences between random and real-world problems, contradicting conventional wisdom that states that mistakes at the top of the search tree are much more expensive to refute than those made deeper in the tree. We also observe some interesting patterns in terms of where most of the search effort is consumed *over a large population of problem instances* and show that it is not always the case that extremely large mistakes account for most of the effort. Finally, we show that variable ordering heuristics alone can avoid making mistakes, but that their performance cannot be attributed exclusively to either fail-firstness or promise.

## 2 Experiments

We study the failure characteristics of backtrack search methods in constraint satisfaction problems. Our analysis is based on counting the number of times an assignment was made during search that took us off the path to a solution. We refer to such decisions as *mistakes* [4]. The set of nodes visited by the algorithm in order to recover from a mistake is the *refutation tree* of that mistake. The number of nodes in that tree is the *refutation size*, our measure of effort.

Our empirical analysis includes configurations of uniform Model B random binary problems and quasigroup completion problems, encoded using binary constraints. With the exception of the random  $17 \times 8$  problems (i.e. 17 variables with uniform domain size 8), where we used backtracking, all other experiments used MAC. Our data sets of random problems contain approximately 10,000 instances for each algorithm used. The QWH-10 data set includes a total of over 1,000,000 instances.

Rather than using the overall effort required to solve each instance, which we refer to as *instance-based effort*, as the basis of our analysis, we also considered the effort required to refute each mistake separately, which we refer to as *mistake-based effort*.

### 2.1 Distribution of Search Effort

Figure 1 shows heatmaps of the probability of encountering mistakes of a certain size at a certain depth (Figures 1(a) and 1(c)), as well as the proportion of effort spent at a certain depth in refutations of a certain size (Figures 1(b) and 1(d)). Colours represent a log-scale.

- While a survival function-based analysis [2] can demonstrate the presence or absence of heavy tails, heatmap visualisations also show precisely the depth where they occur.

- The QWH-10 plots in Figure 1 contradict the conventional wisdom – the largest mistakes occur at *intermediate* depths.
- Over a population of instances, for all the algorithms we used, the bulk of the effort is spent in refuting the extremely large number of small mistakes (4 to 100 nodes) that occur deep down in the search tree (Figure 1(c)).
- Random binary problems do exhibit exponential decay of the refutation size with depth (Figures 1(a), 1(b) and 2(d)). Near the phase transition, mistakes start occurring close to the root of the tree, and over a population of instances, the bulk of the effort (the dark colour in the heatmaps), shifts towards the top of the tree.
- The mean and median refutation sizes in Figure 1(d) fail to provide any information with respect to the wide range and distribution of refutation sizes encountered here.
- Heatmaps help in visually determining the absence of heavy tails by observing the variation in refutation sizes (Figure 1(d)) and the proportion of effort associated with the large refutations (Figure 1(b)).

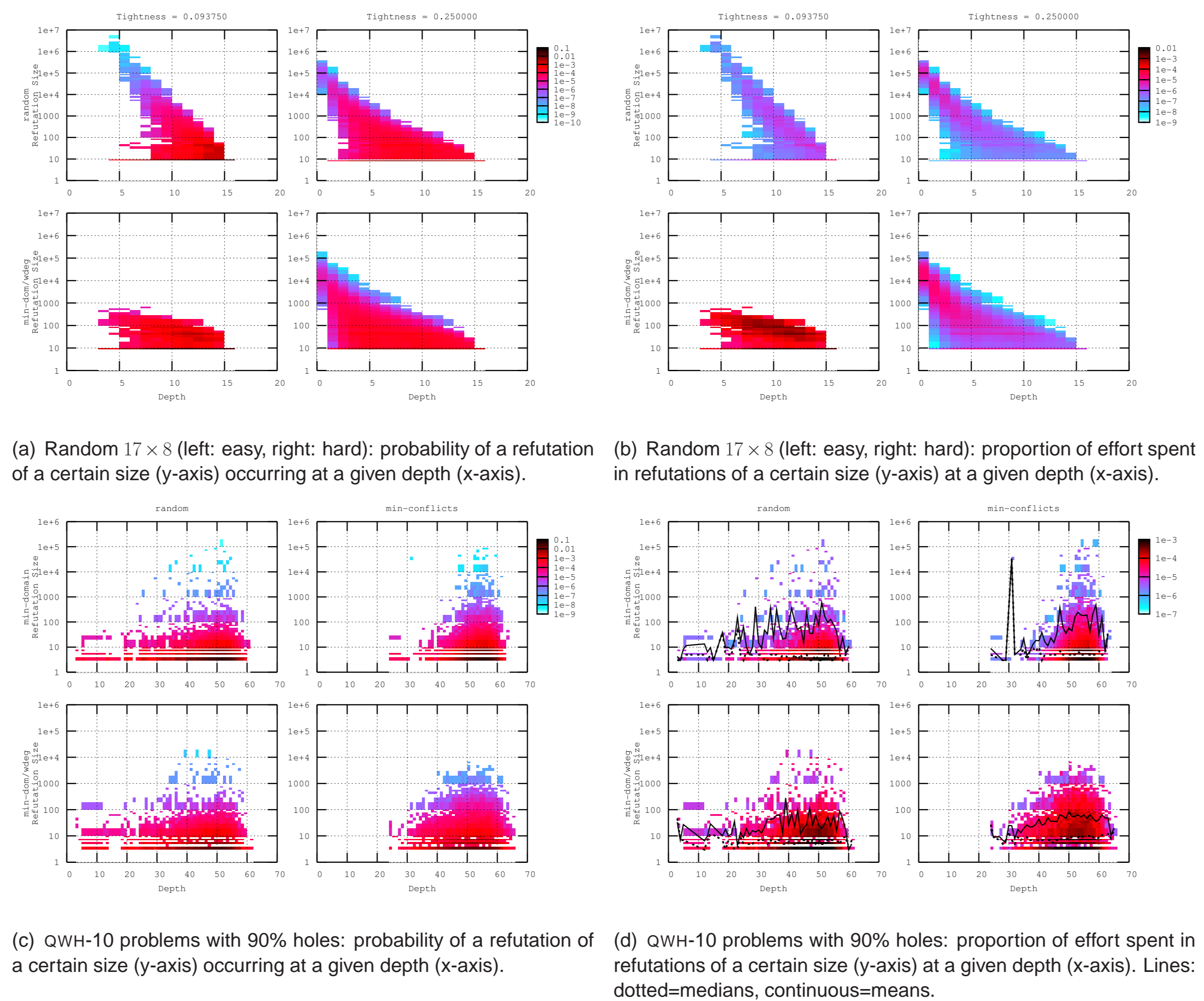


Figure 1: Heatmaps for random problems  $17 \times 8$  and QWH-10 with 90% holes.

### 2.2 Promise versus Fail-Firstness

From the heuristics studied here, min-dom/wdeg performs best not because it makes fewer mistakes, or because it refutes them with less effort, but because it strikes a good balance between promise [1] and fail-firstness [3].

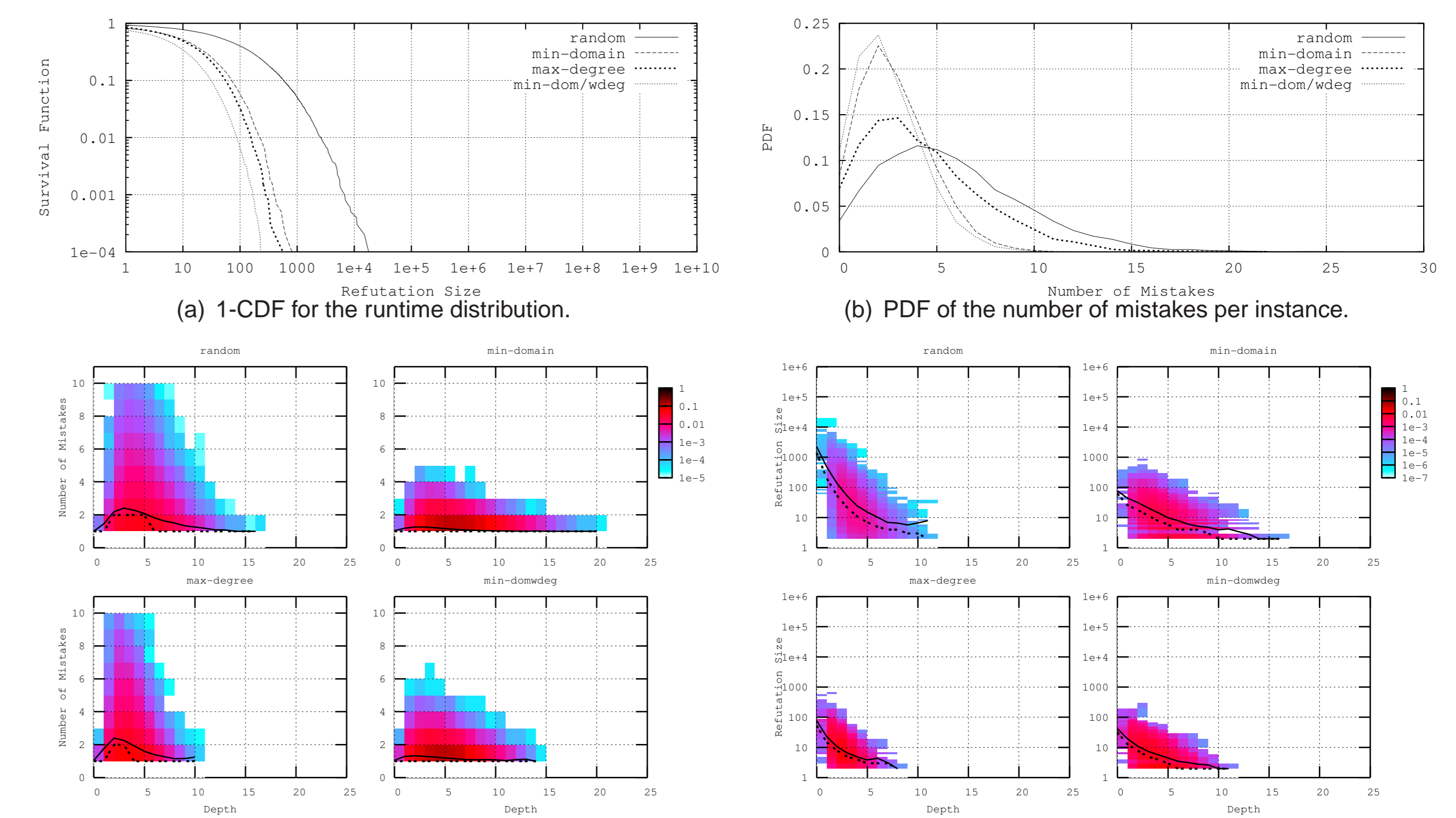


Figure 2: Sparse random  $30 \times 10$ : Promise vs fail-firstness; 4 variable orderings, random values.

## 3 Conclusions

Our novel use of heatmaps nicely complements the use of survival functions and allows a more granular view of the complex interaction between a heuristic’s ability to avoid mistakes and its ability to recover from them. Heatmaps have helped show very clearly that the effort required to recover from mistakes is not always correlated with the depth where they occur, and better search heuristics do not necessarily make fewer mistakes, or have the ability to quickly recover.

## Acknowledgments

This work was supported by Science Foundation Ireland (Grant 00/PI.1/C075). We thank the Boole Centre For Research in Informatics for providing access to their Beowulf cluster.

## References

- J.C. Beck, P. Prosser, and R.J. Wallace. Variable ordering heuristics show promise. In *Proceedings of CP-2004*, LNCS 3258, pages 711–715, 2004.
- C.P. Gomes, C. Fernández, B. Selman, and C. Bessière. Statistical regimes across constrainedness regions. *Constraints*, 10(4):317–337, 2005.
- R.M. Haralick and G.L. Elliott. Increasing tree search efficiency for constraint satisfaction problems. *Artificial Intelligence*, 14(3):263–313, 1980.
- T. Hulubei and B. O’Sullivan. The impact of search heuristics on heavy-tailed behaviour. *Constraints*, 11(2–3):157–176, 2006.