Proceedings of the Tenth International Symposium on Combinatorial Search (SoCS 2017)

Cost-Based Heuristics and Node Re-Expansions Across the Phase Transition

Eldan Cohen, J. Christopher Beck
Department of Mechanical & Industrial Engineering
University of Toronto
Toronto, Canada
{ecohen, jcb}@mie.utoronto.ca

Abstract

Recent work aimed at developing a deeper understanding of suboptimal heuristic search has demonstrated that the use of a cost-based heuristic function in the presence of large operator cost ratio and the decision to allow re-opening of visited nodes can have a significant effect on search effort. In parallel research, phase transitions in problem solubility have proved useful in the study of problem difficulty for many computational problems and have recently been shown to exist in heuristic search problems. In this paper, we show that the impact on search effort associated with a larger operator cost ratio and the number of node re-expansions is concentrated almost entirely in the phase transition region. Combined with previous work connecting local minima in the search space with such behavior, these observations lead us to hypothesize a relationship between the phase transition and the existence of local minima.

1 Introduction

The phase transition in problem solubility has been an important tool in the study of problem difficulty for a number of computational problems (e.g., SAT (Mitchell, Selman, and Levesque 1992; Crawford and Auton 1996) and CSP (Smith and Dyer 1996; Prosser 1996)). In a recent work, we showed the existence of a rapid transition in problem solubility in heuristic search and the occurrence of the hardest problems during this transition (Cohen and Beck 2017).

In this paper, we empirically analyze how the phase transition phenomenon interacts with two algorithm design decisions: the use of cost-based heuristics on problems with differing operator cost ratio (Wilt and Ruml 2011; 2014; Cushing, Benton, and Kambhampati 2010; 2011) and whether or not to allow re-opening of closed nodes (Valenzano, Sturtevant, and Schaeffer 2014; Sepetnitsky and Felner 2015; Sepetnitsky, Felner, and Stern 2016). For Greedy Best First Search (GBFS) we show the following:

1. The effect of large operator cost ratio on the search effort is most significant in the phase transition region and diminishes outside.
2. The number of node re-expansions peaks in the phase transition and declines outside.
3. *exceptionally hard problems* (Gent and Walsh 1994a; Smith and Grant 1995) appear in the relaxed region of the phase transition where the median effort is relatively low.

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2 The Phase Transition Phenomenon

Cheeseman, Kanefsky and Taylor (1991) showed that many NP-complete problems exhibit a transition between regions of solubility and insolubility over a narrow interval of an appropriate problem generation control parameter. One region is under-constrained with high density of solutions and the other is over-constrained with low likelihood that a solution exists. Furthermore, the hardest problems occur in this interval. As noted, a significant body work in the 1990s developed these insights across a number of problem types.

Researchers have used the terms *mushy region* (Smith and Dyer 1996) to refer to the interval and *crossover point* (Crawford and Auton 1996) to refer to the point in which the probability that a problem is soluble is 0.5. We adopt these terms and quantify the mushy region as the range of the control parameter's values in which the observed portion of soluble problems is between 0.1% and 99.9%.

In a previous work, we showed that both the solubility and problem hardness aspects of the phase transition phenomenon can also be observed for GBFS on unit-cost domains (Cohen and Beck 2017). As this paper extends this work we present the foundational definitions.

Definition 1. (Observed connectivity density) *Let $G\langle V, E\rangle$ be an arbitrary transition graph. We define the observed connectivity density of this graph $\mathcal{P}(G) = \frac{|E|}{|V|\cdot(|V|-1)}$.*

Definition 2. (Restricted instance) *Let $G\langle V, E\rangle$ be an arbitrary transition graph. $\hat{G}\langle V, \hat{E}\rangle$ is considered a restricted instance of G if $\mathcal{P}(\hat{G}) < \mathcal{P}(G)$ and $\hat{E} \subseteq E$.*

Definition 3. (Relaxed instance) *Let $G\langle V, E\rangle$ be an arbitrary transition graph. $\hat{G}\langle V, \hat{E}\rangle$ is considered a relaxed version of G if $\mathcal{P}(\hat{G}) > \mathcal{P}(G)$ and $\hat{E} \supseteq E$.*

Model 1. (p-Constrained Benchmark Problems) *Given an existing problem's transition graph $G\langle V, E\rangle$ and the required connectivity density p, the class $R_{G,p}$ consists of all problem instances $\langle T, S_i, S_g\rangle$ such that:*

1. T, the transition graph, is a restricted instance of G if $p < \mathcal{P}(G)$, or a relaxed instance otherwise. $\mathcal{P}(T) = p$.

2. $S_i \in S$, *a randomly chosen initial state,* $\exists k : (S_i, S_k) \in T$.
3. $S_g \in S$, *a randomly chosen goal state such that* $S_g \neq S_i$ *and* $\exists k : (S_k, S_g) \in T$.

Finally, we use the following control parameter from which the required connectivity density p is derived:

$$\gamma := \frac{\text{Expected number of edges in the transition graph}}{\text{Number of states}}$$

3 A Domain-Specific Model

While Model 1 provides a domain-independent way to generate relaxed and restricted problems, for some domains one can find a domain-specific parameter that controls the constrainedness of the problem. For example, in the Grid Navigation domain, the probability of a blocked cell q clearly controls the constrainedness of the generated instances: a high q will produce more constrained state spaces with lower solution density while a low q produces the opposite.[1]

We therefore define a domain-specific model for Grid Navigation and use it in our analysis.

Model 2. (q-Constrained Grid Navigation Problems) *Given grid dimensions* $n \times m$, *we denote by* $G_{nm}\langle V, E\rangle$ *the transition graph of an* $n \times m$ *grid navigation problem and by* q *the probability of a blocked cell. We define the class* $R_{n,m,q}$ *that consists of all problem instances* $\langle T, S_i, S_g\rangle$ *such that:*

1. T, *the transition graph, is an instance of* G_{mn} *in which each cell is blocked with probability* q.
2. $S_i \in S$, *a randomly chosen initial state,* $\exists k : (S_i, S_k) \in T$.
3. $S_g \in S$, *a randomly chosen goal state such that* $S_g \neq S_i$ *and* $\exists k : (S_k, S_g) \in T$.

4 Cost-based Heuristics

Definition 4. (Cost-based search; Cushing, Benton, and Kambhampati, 2011) *A best first search in which* $g(x) = g_c(x)$, *the cost to reach state* x, *and* $h(x) = h_c(x)$, *an estimation of the cost of the cheapest path from state* x *to a goal state.*

Definition 5. (Size-based search; Cushing, Benton, and Kambhampati, 2011) *A best first search in which* $g(x) = g_d(x)$, *the distance (i.e., the number of actions) to reach state* x, *and* $h(x) = h_d(x)$, *an estimation of distance of the shortest path from state* x *to a goal state. Also called distance-based search.*

Definition 6. (Operator cost ratio; Wilt and Ruml, 2011) *The ratio of the largest edge weight in the graph to the smallest edge weight in the graph.*

Wilt and Ruml (2011) showed, empirically and theoretically, that a larger operator ratio can have a negative effect on the search effort of various best-first search algorithms, including GBFS. Their analysis showed that cost-based versions of sliding puzzle and the pancake problem become intractable as the operator cost ratio is increased. The cost-based grid navigation problem, however, does not suffer from a significant increase in search effort and the authors attribute it to size-bounded local minima and the large number of duplicates.[2] To mitigate the negative effect of large operator cost ratio, a number of authors suggest using size-based search (Cushing, Benton, and Kambhampati 2011; Wilt and Ruml 2011; 2014).

In this section, we present an empirical analysis of size-based and cost-based search on versions of these three domains plus TopSpin, across the phase transition. For each of the domains, we propose a cost function that is flexible enough to allow us to control the operator cost ratio and examine the change in search effort as we manipulate it.

All the experiments in this section use GBFS configured to not re-open closed nodes and to randomly break ties in h-values. We generated random problem instances for 25 γ values in $[0, |nodes| - 1]$, with higher density inside the phase transition. For each γ value we generated 1000 random instances. For each instance, we record its solubility and the number of nodes expanded in order to find a solution or prove that none exists.

4.1 Median-Case Analysis

Grid Navigation. We consider a 500×500 Grid Navigation Problem based on Model 2. We define $C_m(s, a)$, the cost of applying action a on state s, using a parameter m:

$$C_m(s,a) = \begin{cases} 1^m, & \text{if } a = \textit{up} \\ 2^m, & \text{if } a = \textit{down} \\ 3^m, & \text{if } a = \textit{left} \\ 4^m, & \text{if } a = \textit{right} \end{cases}$$

The parameter m controls the operator cost ratio. As the smallest operator cost is fixed to 1, the operator cost ratio is then 4^m. The heuristic function is based on Manhattan Distance, weighted accordingly.

Figure 1 shows the probability of solution and Figure 2 shows the median number of expanded nodes, for $m \in \{1, 2, 4, 10\}$, plotted against the probability that a cell is not blocked $(1 - q)$.

We see a rapid transition in the problem solubility and the peak of the median search effort is located in close proximity to the crossover point, as been shown for Model 1 (Cohen and Beck 2017). Even as we increase the operator cost ratio, the median effort peak remains at the crossover point.

To directly compare the effect of the different cost functions across the phase transition, we analyze the relative number of expanded nodes. Since the operator cost ratio has no effect on the search effort required to solve insoluble instances (i.e., every node that is accessible from the initial state will be expanded exactly once to prove insolubility), we focus only on the soluble instances. To avoid bias due to small sample size, we only consider the points in the phase transition in which at least 10% of the problems are soluble.

Figure 3 shows the median ratio of the number of expanded nodes when using a cost function with higher operator cost ratio (i.e., $m \in \{2, 4, 10\}$) to the search effort when using the lowest operator cost ratio heuristic ($m = 1$).

[1] With $q = 0$ the state space is still highly constrained, as there are only four actions per state. It is possible to relax the domain further by allowing more actions (e.g., diagonal moves, jumps, etc.).

[2] See Fan, Müller, and Holte (2017) for recent work.

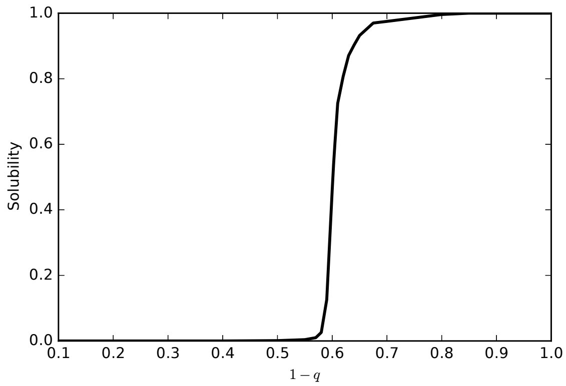

Figure 1: 500 × 500 Grid Navigation: Probability of a solution vs. probability of an unblocked cell.

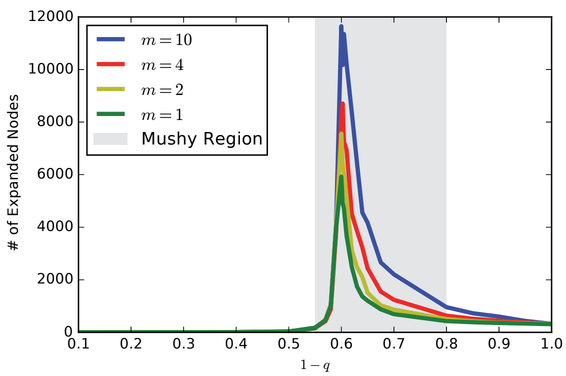

Figure 2: 500 × 500 Grid Navigation: Median effort vs. probability of an unblocked cell.

As Figure 3 clearly shows, the increase in search effort, associated with the large operator ratio, is centered in the region of phase transition. Outside that region, the effort ratio gradually diminishes towards a ratio of one.

Figure 4 shows the median search effort ratio for the

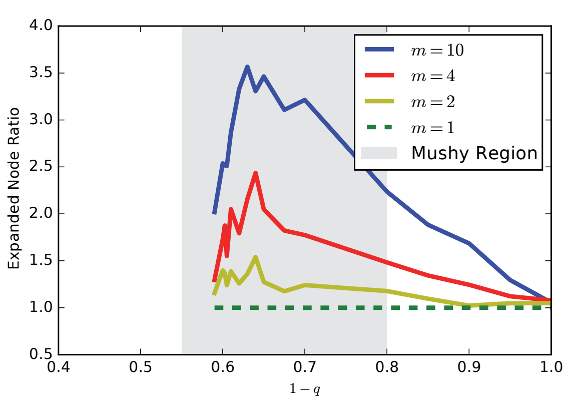

Figure 3: 500 × 500 Grid Navigation: Median effort ratio of soluble instance vs. probability of an unblocked cell.

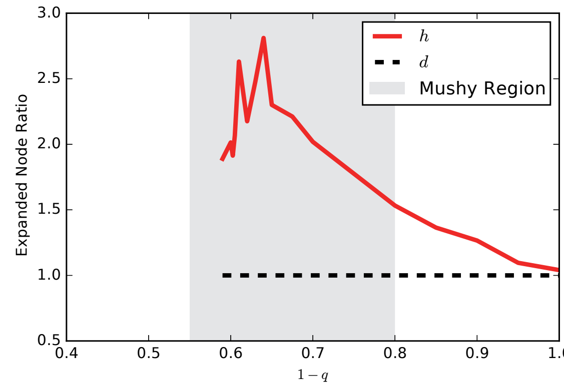

Figure 4: 500×500 Grid Navigation: Median effort ratio between the distance heuristic and a cost heuristic with m=4.

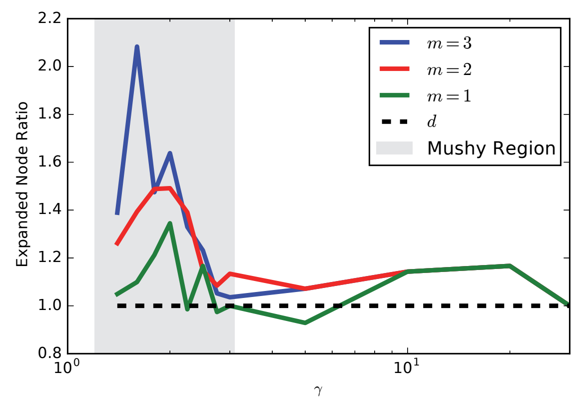

Figure 5: 8-Pancake Problem: Median effort ratio of soluble instance vs. γ.

500×500 Grid Navigation Problem (Model 2) with a large operator ratio (m=4) between the cost heuristic and the distance heuristic d. The improvement due to the distance heuristic is also concentrated in the phase transition region. In fact, the distance heuristic seems to behave similarly to a cost-based heuristic with a lower operator cost ratio (which, of course, it is).

Pancake Problem. For the 8-Pancake Problem, based on Model 1, our cost function is defined based on the bottom pancake in the sub-pile that is about to be flipped. Although somewhat artificial, this cost function is easily incorporated into the *gap heuristic* (Helmert 2010) and allows us to investigate the effect of the operator cost ratio on the search effort. Given z, the size of the lowest pancake in the flipped sub-pile, we define the cost of the flip to be z^m. Again, we use the parameter m to control the operator cost ratio, which is 8^m for the 8-Pancake Problem.

Figure 5 shows the median effort ratio between cost-based search for $m \in \{1, 2, 3\}$ and a distance-based search using d. The increase in search effort ratio that is associated with larger operator cost ratio is significant only inside the phase transition region and diminishes outside. The distance heuristic, as before, behaves similarly to a cost-based heuristic with a lower operator cost ratio.

Sliding Tiles. We consider the 3×3 Sliding Tiles Problem based on Model 1 for which the state space consists of two large disconnected components (Wilson 1974). In our investigation we generate problems in which the initial state and the goal state are in the same component, and the added (resp., removed) edges of the relaxed (resp., restricted) instances are within this component. This allows us to observe the full phase transition on one component and to avoid the sudden connection of two large components.

Wilt and Ruml (2011) incorporate costs for the Sliding Tiles problem by assigning different costs for moving each tile. In a later work, they showed that using an inverse cost structure, in which the cost of moving a tile is in inverse correlation to the face of the tile, has a larger expected local minimum (Wilt and Ruml 2014). We therefore use a parametrized version of Wilt and Ruml's cost function in which the cost of moving a tile with a face-value of z is $\frac{1}{z^m}$. The parameter m controls the operator cost ratio, which is 8^m, for this domain. The heuristic function is based on the standard Manhattan Distance, weighted by the cost associated with each tile.

Figure 6 shows the median search effort ratio between a cost-based search for $m \in \{1, 1.5, 2\}$ and a search based on the distance heuristic d. The impact on search effort ratio that is associated with a larger operator cost ratio is concentrated in the phase transition, and diminishes, though is still apparent, as we move away from the phase transition.

TopSpin. Wilt and Ruml (2014) found that in some cases, using a cost-based heuristic requires less search effort than using the distance heuristic that has a lower operator cost ratio, due to smaller local minima in the cost-based heuristic. We observe such behavior for the TopSpin domain.

We consider a 10-disk TopSpin domain with a 4-disk turnstile. Our cost function is based on the sum of faces of the disks in the turnstile:

$$C_m(s,a) = (\sum_{z \in T_a} z)^m$$

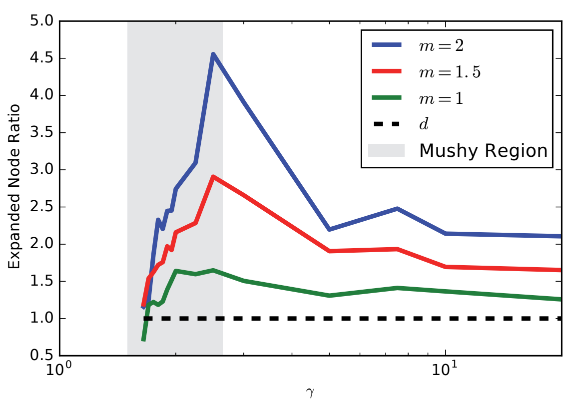

Figure 6: 3×3 Sliding Tiles: Median effort ratio of soluble instance vs. γ.

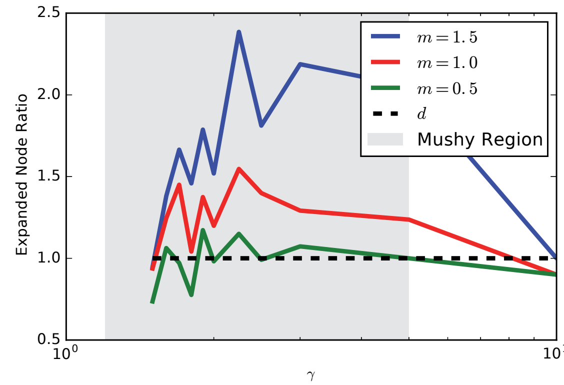

Figure 7: 10-disk Top: Median effort ratio of soluble instances vs. γ.

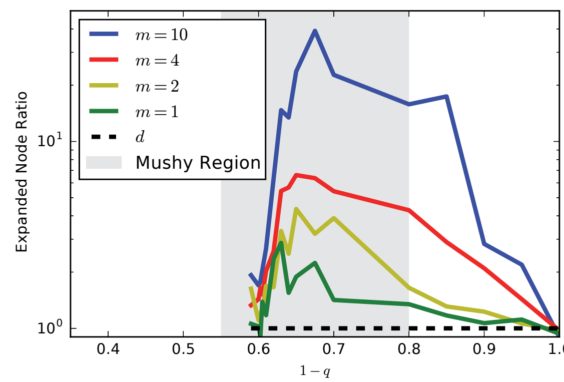

Figure 8: 500×500 Grid Navigation: 99.9%-percentile effort ratio vs. probability of an unblocked cell.

where T_a is the set of faces of the disks in the turnstile, and m is a parameter controlling the operator cost ratio.

Figure 7 shows the median search effort for the TopSpin problem with a PDB heuristic for $m \in \{0.5, 1, 1.5\}$, compared to a distance-based heuristic. As expected, we observe an increased effort as we increase the operator cost ratio. In this case the distance heuristic is not strictly better than the cost-based heuristic for $m = 0.5$. This is consistent with Wilt and Ruml's observation. The differences diminish as we move away from the phase transition.

4.2 The Hardest Instances

In the previous section we investigated the effect of operator cost ratio on the median search effort. Here we examine the hardest instances across the constrainedness range.

Figure 8 shows the 99.9%-percentile effort ratio for soluble instances of the Grid Navigation Problem (note the log scale on the y-axis for better readability). The differences are significantly larger (i.e., the effort ratio peaks at 40) inside the phase transition region, compared to the median case. However, high-percentile ratios too, gradually diminish towards one as we move away from the phase transition. Interestingly, the peak of the ratio curve slightly shifts to the

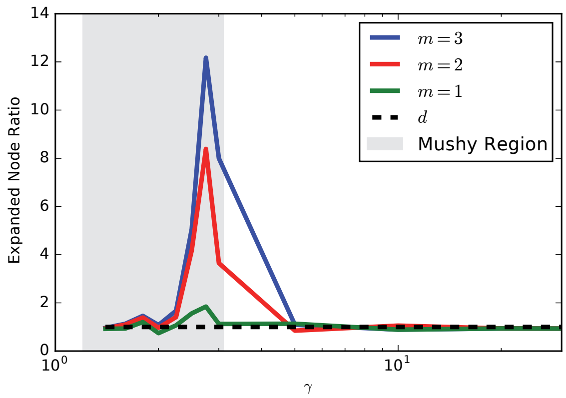

Figure 9: 8-Pancake: 99.9%-percentile effort ratio vs. γ.

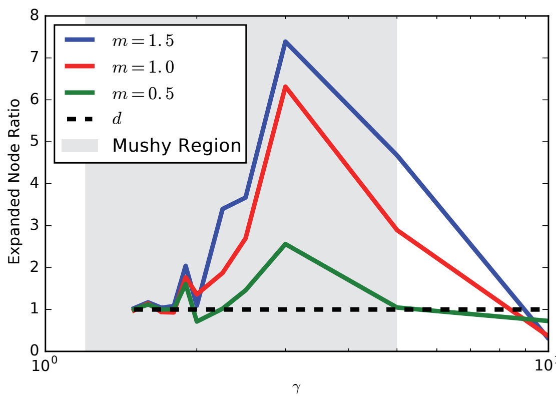

Figure 10: TopSpin: 99.9%-percentile effort ratio vs. γ.

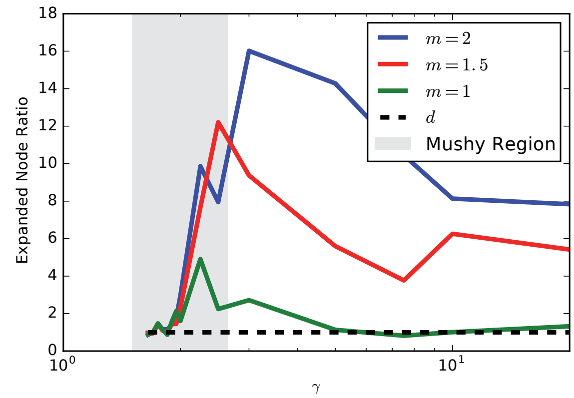

Figure 11: Sliding Tiles: 99.9%-percentile effort ratio vs. γ.

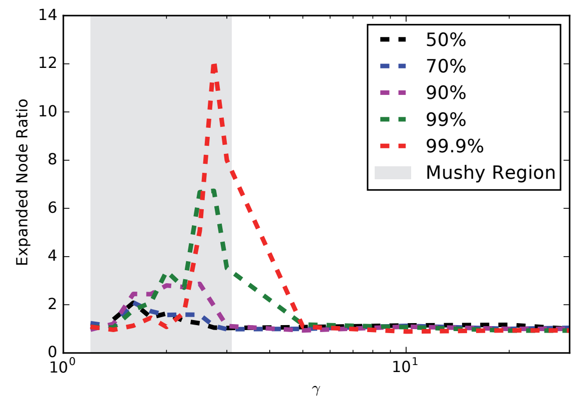

Figure 12: 8-Pancake: 99.9%-percentile effort ratio vs. γ.

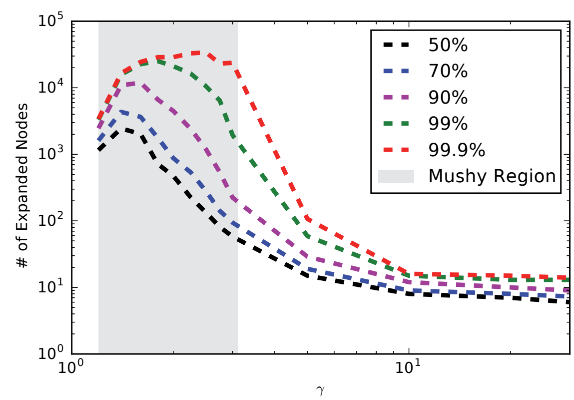

Figure 13: 8-Pancake: 99.9%-percentile effort vs. γ.

more relaxed areas, compared to the median case.

Figure 9, Figure 10 and Figure 11 show the 99.9%-percentile effort ratio for soluble instances of the Pancake Problem, TopSpin and Sliding Tiles Problem respectively. Again, we see that the ratio of the number of expanded nodes inside the phase transition is significantly larger compared to the median case and it diminishes as we move away from the phase transition. Also, we can see a much stronger shift in peak compared to the median case. For the Sliding Tiles, we observe that the peak is located beyond the phase transition region, however as we move away from the peak we see the expected decline.

The shift in peak suggests that the largest ratio for the hardest problems is found in a more relaxed region of the phase transition. Figure 12 and 13 show the relative effort (compared to the distance heuristic) and absolute effort in the higher percentiles of the Pancake problem with $m = 3$. We can see that the peak of the highest percentiles of the absolute effort and the peak of the effort ratio both shift to the more relaxed regions in the phase transition. However, the results suggest that this phenomenon is limited to the highest percentiles ($\geq 99\%$). Similar results have been observed for the other domains.

4.3 Discussion

The behavior of the median search effort ratio and the hardest instances suggests that the effect of the operator cost ratio is concentrated inside the phase transition. Our results provide a deeper understanding of Wilt and Ruml's observations on the operator cost ratio (Wilt and Ruml 2011; 2014), empirically demonstrating that they depend on the constrainedness of the problem.

The anomaly of finding the hardest instances in the "easy" regions of the phase transition has been observed for other types of computational problems (Gent and Walsh 1994a; 1994b; Hogg and Williams 1994). Such problem instances have been termed *exceptionally hard problems (ehps)*. The *ehps* are not simply outliers but rather outliers in an unexpected region of the phase transition and hence have been the subject of a number of investigations (Gent and Walsh 1994a; Smith and Grant 1995).

We previously showed that, for unit-cost problems, the 100%-percentile peaks in the "easy" region (Cohen and Beck 2017). However, that curve is dominated by insoluble instances. This result is not surprising, since heuristic search with a heuristic function that returns a finite value has to exhaust the accessible state space to prove infeasibility. However, the existence of soluble *ehps* in heuristic search is a new result.

The deviation observed for the effort ratio of the hardest problems of the Sliding Tiles domain, although small, is an anomaly that may be due to the interaction of the phase transition phenomenon and the reasons for ehps that requires further study to understand.

5 Node Re-Expansions

In A*, $f(n) = g(n) + h(n)$ and re-expansions of previously visited nodes only occur when using an inconsistent heuristic. In suboptimal search algorithms such as GBFS and Weighted A*, as $g(n)$ is not considered or is weighted less than $h(n)$, we are not guaranteed to avoid re-expansions even when using an admissible and consistent heuristic.

Although no theoretical or empirical analysis of which we are aware has specifically addressed re-expansions in GBFS, the authors of several recent works chose to configure GBFS to not re-open closed nodes, even if a shorter path is found (Valenzano et al. 2014; Xie, Müller, and Holte 2014).

For Weighted A*, re-expansions can have significant negative effect on the search effort. Valenzano, Sturtevant and Schaeffer (2014) presented empirical analysis of re-expansions for Weighted A* on pathfinding problems. As they increased the weight on $h(n)$, the proportional search effort spent on re-expansions of visited nodes increased. For $w = 10$, they observe that 91% of the total node expansions were re-expansions. Sepetnitsky, Felner and Stern (2016) performed an empirical analysis for Weighted A* and showed that in more than 90% of the cases, a policy of node re-opening leads to a search effort that is at least as high as no-reopening, reaching 99.9% for higher weights.

5.1 GBFS

In this section we present an empirical analysis of the effect of node re-expansions across the constrainedness range. All the experiments in this section use GBFS. Naturally, we configure the search to re-open closed nodes if a cheaper path is found. As before, we randomly break h-value ties. We limit the analysis to unit-cost problems and generated 1,000 random problem instances for each of the 25 sampled γ values in $[0, |nodes| - 1]$.

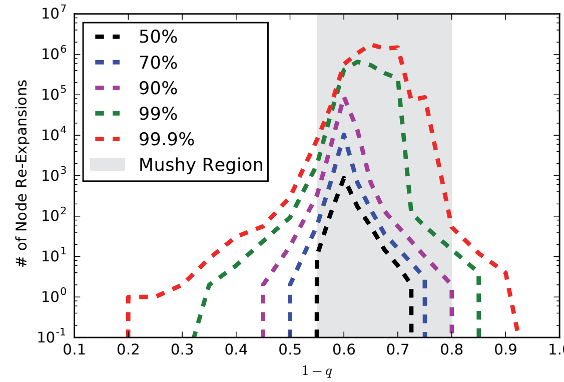

Figure 14: 250 × 250 Grid Navigation: Median and higher percentiles of node re-expansions.

Grid Navigation. Figure 14 shows the median and higher percentiles of re-expansions across the constrainedness range. On average, re-expansions only occur within the phase transition: the median number of re-expansions out-

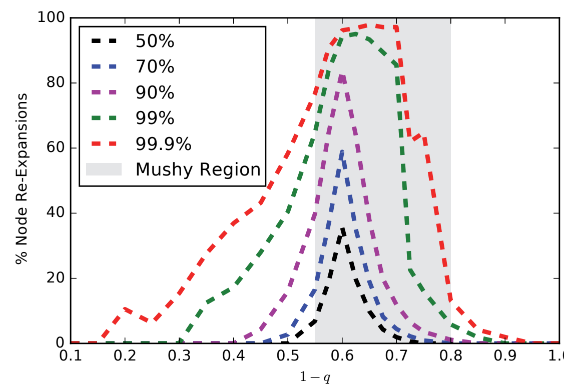

Figure 15: 250 × 250 Grid Navigation: Median and higher percentiles of node re-expansions.

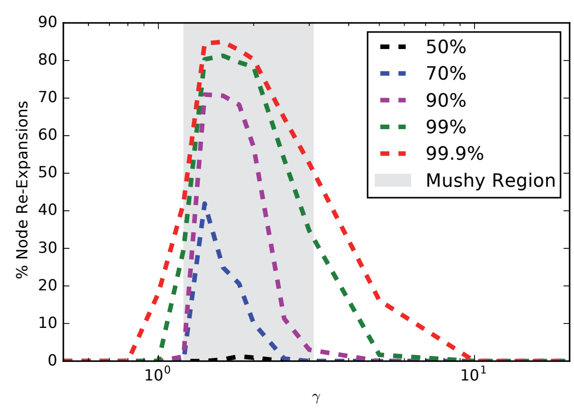

Figure 16: 8-Pancake Problem: Median and maximal node re-expansions vs. γ.

side of the phase transition region is zero. Even when considering the higher percentiles, we see that outside the phase transition the number is much smaller than inside and declines further from the phase transition region. The re-expansions seem to follow a low-high-low pattern, consistent with the easy-hard-easy pattern in search effort. Also, for percentiles $\geq 99\%$, we can see the peak stretches toward areas where the median is declining, similar to the ehps.

Figure 15 shows the number of re-expansions relative to total expansions (Valenzano, Sturtevant, and Schaeffer 2014) and presents a similar pattern.

Pancake Problem. Figure 16 shows the median and higher percentiles of re-expansions for the 8-Pancake problem. The median is very low, which can be attributed to the quality of the *gap heuristic*. However, the trends are clear and similar to the Grid Navigation problem. In the median case, we see zero re-expansions outside the phase transition region, and the higher percentiles of re-expansions decrease as we move away from the phase transition. Again, the peak is wider for the higher percentiles. Similar trends are observed for the relative number of re-expansions.

Sliding Tiles. Figure 17 shows the results for the 3×3 Sliding Tiles problem. The re-expansions peak inside the phase transition. The median reaches zero shortly after leaving the phase transition region. The higher percentiles do not reach zero in the sampled γ range, however we can see the decline as we move away from the phase transition. As expected, the peak of the higher percentiles is wider for higher percentiles.

TopSpin results are similar to those of the Grid Navigation and the Pancake Problem and are omitted due to space.

5.2 Weighted A*

With the interest in node re-expansions in Weighted A* and the similarity between GBFS and Weighted A* with large weight on $h(n)$, it is interesting to see if the GBFS behavior we have observed can also been seen in Weighted A*.

We use the Grid Navigation problem (Model 2), for which the heuristic function remains admissible and consistent for all values of q. Figures 18 and 19 show the relative and absolute number of expanded nodes for Weighted A* with different weights on $h(n)$.

When $w = 1$, the number of node re-expansions is zero everywhere, as expected. As we increase w, the number of node re-expansions increases in the phase transition. Consistent with our observation for GBFS, as we move away from the phase transition, the number of re-expansions declines for all weights to the point it reaches zero.

We also analyzed the higher percentiles of node re-expansions for each w value. Figure 20 shows the 99.9%-percentile of node re-expansions for $w \in \{1, 1.5, 2, 5, 10\}$. The results show that the 99.9%-percentile of relative node re-expansions is also higher as we increase the weight, although it is already very high for $w = 1.5$. Again, the peak of the higher percentiles is wider.

6 Discussion

Our results show that the phase transition phenomenon plays an important role in 1) the relation between search effort for cost-based heuristics and operator cost ratio and 2) the relation between search effort and allowing node re-expansions. Consistent with these observations, we also show that the distance heuristic, that can mitigate the effect of a large operator cost ratio, provides significant improvement only inside the phase transition.

It is important to clarify that while the phase transition is useful in predicting the location of the hardest problems (both the median hardest and single hardest instances), it does not predict the required search effort. Such a prediction requires taking into account other factors such as the size of the state space and the strength of the heuristic. However, our results show that the phase transition is a key factor in search effort.

The Phase Transition and Local Minima. Previous work has shown that the negative effect of increasing the operator cost ratio is due to the deepening of the local minima, while the distance heuristic tends to have smaller local minima (Wilt and Ruml 2011; 2014; Cushing, Benton, and Kamb-

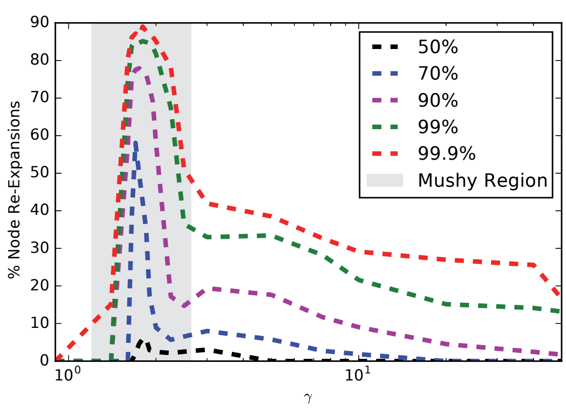

Figure 17: 3×3 Sliding Tiles: Median and maximal node re-expansions vs. γ.

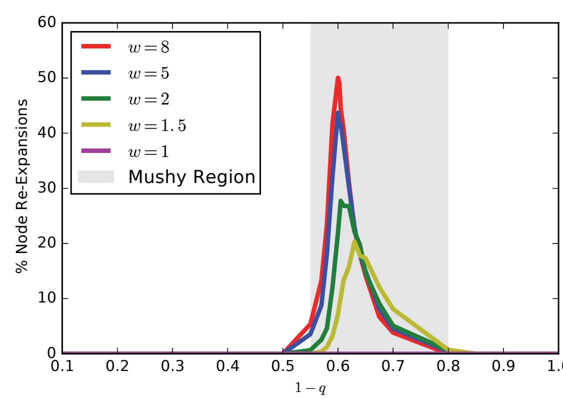

Figure 18: 500×500 Grid Navigation: Median percent of node re-expansions vs. probability of an unblocked cell.

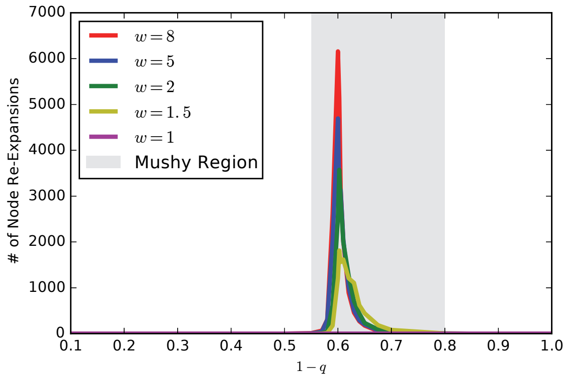

Figure 19: 500 × 500 Grid Navigation: Median absolute node re-expansions vs. probability of an unblocked cell.

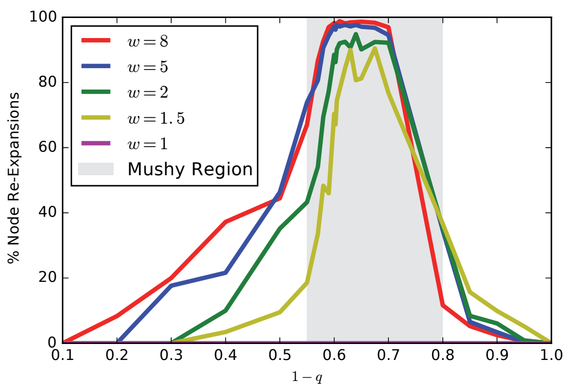

Figure 20: 500 × 500 Grid Navigation: 99.9%-percentile of node re-expansions vs. probability of an unblocked cell.

hampati 2010; 2011). Our results show that the effect of increasing the operator ratio, and the benefit often gained by using the distance heuristic, is significant in the phase transition region, and decreases as we move away.

These results suggest a connection between the constrainedness of a problem and the existence of local minima. A reasonable hypothesis is that the likelihood and/or extent of local minima is much larger in the phase transition and insignificant outside. A prime area for our future work will be to investigate this hypothesis.

The hypothesis is supported by the discovery of soluble ehps in heuristic search. Such instances in other types of computational problems are associated with large insoluble subproblems that the search has to exhaust if it enters (Smith and Grant 1997; 1995; Gent and Walsh 1994a). Wilt and Ruml (2014) defined local minima in heuristic search as a region that does not contain the goal but that the search will have to exhaust if it enters. The similarity between these definitions, as well as their similar location in the phase transition, suggests that they are analogous phenomena. As the large insoluble subproblems are directly associated with the constrainedness of the problem, we conjecture a similar relationship exists for the local minima in satisficing heuristic search.

Several methods have been suggested to mitigate the effect of local minima, including the use of randomization (Valenzano et al. 2014) and local exploration (Xie, Müller, and Holte 2014; 2015). Investigating these methods using the framework of phase transition may yield interesting new insights.

Wilt and Ruml (2015) suggested a quantitative metric to evaluate and compare heuristics for greedy best first search, called the Global Distance Rank Correlation (GDRC), similar to the Fitness-Distance Correlation (FDC) used for other computational problems (e.g., Heckman and Beck, 2008). They note that, in general, domains with large local minima have poor GDRC. Investigating the implications of the phase transition to the GDRC/FDC metric is also an interesting open question.

7 Conclusion

We performed an empirical analysis of problem instances generated across the phase transition on two aspects of heuristic search algorithms: using cost-based search with varying operator cost ratio and using node re-expansions. We showed that the effect on search effort associated with a larger operator ratio is concentrated in the phase transition. We also demonstrated that the number of node re-expansions for both GBFS and Weighted A* peaks in the phase transition and decreases sharply outside.

Our results suggest that many of the phenomena that are associated with larger search effort are effected by the phase transition and, therefore, that they should be studied at different level of constrainedness. We hypothesize that the existence of local minima in heuristic search problems is closely related to the phase transition phenomenon.

Acknowledgments We thank the anonymous reviewers whose valuable feedback helped improve the final paper. The authors gratefully acknowledge funding from the Natural Sciences and Engineering Research Council of Canada.

References

Cheeseman, P.; Kanefsky, B.; and Taylor, W. M. 1991. Where the really hard problems are. In *The Twelfth International Joint Conference on Artificial Intelligence (IJCAI)*, 331–337.

Cohen, E., and Beck, J. C. 2017. Problem difficulty and the phase transition in heuristic search. In *The Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, 780–786.

Crawford, J. M., and Auton, L. D. 1996. Experimental results on the crossover point in random 3-SAT. *Artificial Intelligence* 81(1):31–57.

Cushing, W.; Benton, J.; and Kambhampati, S. 2010. Cost based search considered harmful. In *The Third Annual Symposium on Combinatorial Search (SoCS)*, 140–141.

Cushing, W.; Benton, J.; and Kambhampati, S. 2011. Cost based satisficing search considered harmful. *arXiv preprint arXiv:1103.3687*.

Fan, G.; Müller, M.; and Holte, R. 2017. The two-edged nature of diverse action costs. In *The Twenty-Seventh International Conference on Automated Planning and Scheduling (ICAPS)*, in press.

Gent, I. P., and Walsh, T. 1994a. Easy problems are sometimes hard. *Artificial Intelligence* 70(1):335–345.

Gent, I. P., and Walsh, T. 1994b. The hardest random SAT problems. In *KI-94: Advances in Artificial Intelligence*, 355–366. Springer.

Heckman, I., and Beck, J. C. 2008. Fitness distance correlation and solution-guided constructive search for CSPs. In *Proceedings of the Fifth International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR08)*, 112–126.

Helmert, M. 2010. Landmark heuristics for the pancake problem. In *Third Annual Symposium on Combinatorial Search (SoCS)*, 109–110.

Hogg, T., and Williams, C. P. 1994. The hardest constraint problems: A double phase transition. *Artificial Intelligence* 69(1–2):359–377.

Mitchell, D.; Selman, B.; and Levesque, H. 1992. Hard and easy distributions of SAT problems. In *The Tenth National Conference on Artificial Intelligence (AAAI)*, 459–465.

Prosser, P. 1996. An empirical study of phase transitions in binary constraint satisfaction problems. *Artificial Intelligence* 81(1):81–109.

Sepetnitsky, V., and Felner, A. 2015. To reopen or not to reopen in the context of Weighted A*. Classifications of different trends. In *The Eighth Annual Symposium on Combinatorial Search (SoCS)*, 218–219.

Sepetnitsky, V.; Felner, A.; and Stern, R. 2016. Repair policies for not reopening nodes in different search settings. In *The Ninth Annual Symposium on Combinatorial Search (SoCS)*, 81–88.

Smith, B. M., and Dyer, M. E. 1996. Locating the phase transition in binary constraint satisfaction problems. *Artificial Intelligence* 81(1):155–181.

Smith, B. M., and Grant, S. A. 1995. Sparse constraint graphs and exceptionally hard problems. In *The Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 646–654.

Smith, B. M., and Grant, S. A. 1997. Modelling exceptionally hard constraint satisfaction problems. In *International Conference on Principles and Practice of Constraint Programming*, 182–195. Springer.

Valenzano, R. A.; Sturtevant, N. R.; Schaeffer, J.; and Xie, F. 2014. A comparison of knowledge-based GBFS enhancements and knowledge-free exploration. In *The Twenty-Fourth International Conference on Automated Planning and Scheduling (ICAPS)*, 375–379.

Valenzano, R. A.; Sturtevant, N. R.; and Schaeffer, J. 2014. Worst-case solution quality analysis when not re-expanding nodes in best-first search. In *The Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 885–892.

Wilson, R. M. 1974. Graph puzzles, homotopy, and the alternating group. *Journal of Combinatorial Theory (Series B)* 16(1):86–96.

Wilt, C. M., and Ruml, W. 2011. Cost-based heuristic search is sensitive to the ratio of operator costs. In *The Fourth Annual Symposium on Combinatorial Search (SoCS)*, 172–179.

Wilt, C. M., and Ruml, W. 2014. Speedy versus greedy search. In *The Seventh Annual Symposium on Combinatorial Search (SoCS)*, 184–192.

Wilt, C. M., and Ruml, W. 2015. Building a heuristic for greedy search. In *The Eighth Annual Symposium on Combinatorial Search (SoCS)*, 131–139.

Xie, F.; Müller, M.; and Holte, R. 2014. Adding local exploration to greedy best-first search in satisficing planning. In *The Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 2388–2394.

Xie, F.; Müller, M.; and Holte, R. 2015. Understanding and improving local exploration for GBFS. In *The Twenty-Fifth International Conference on Automated Planning and Scheduling (ICAPS)*, 244–248.