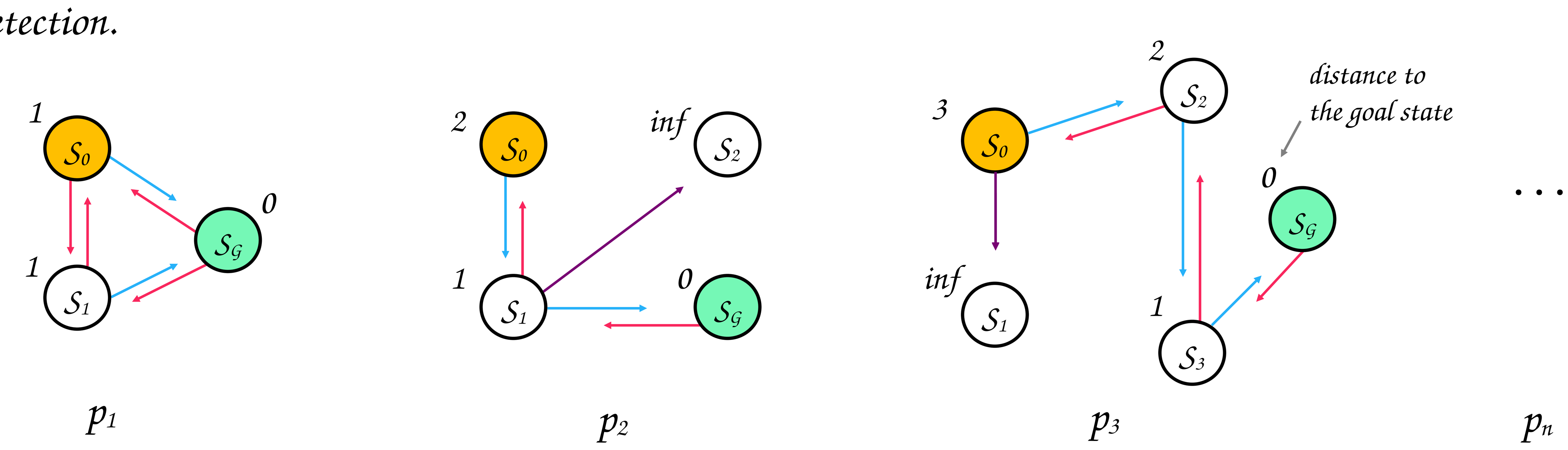
Combining Policies and Heuristics for Classical Planning

In this project, we aim to leverage the strengths of both learning and planning paradigms while mitigating their weaknesses. We follow one of the earliest works (Yoon et al. 2008) in this line of research and propose a combination of policy-guided heuristic search with unsolvability detection. Since our policies are prone to misclassification, the use of heuristics together with policies allows to mitigate the imperfections of the statistical learning. To decrease the impact of inaccurate heuristic estimates, we use the policy for unsolvable transitions as an unsolvability detection.

1 *Use a set of instances $\{p_1, p_2, p_3, \dots, p_n\}$ to learn a policy for 3 types of transitions:*



Good transition decreases the distance to the goal state.

Bad transition increases or does not change the distance to the goal state.

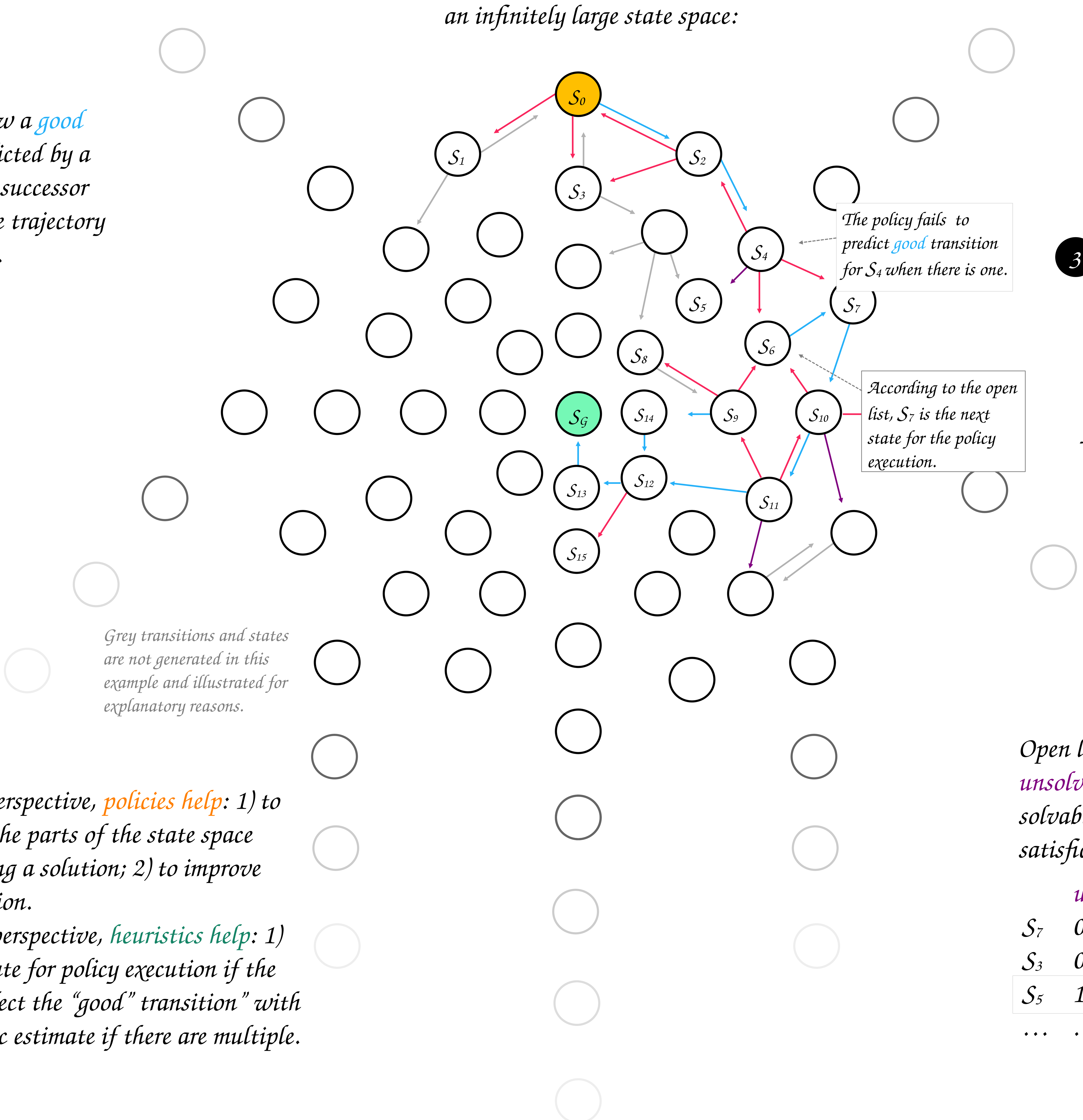
Unsolvable transition leads to the state from which the goal state can never (inf) be reached.

S_0 *initial state*

S_G *goal state*

An example of applying policy-guided search to an instance with an infinitely large state space:

2 *Always follow a good transition predicted by a policy. Add all successor states along the trajectory to the open list.*



3 *If for a given state there is no good transition, terminate the policy. Choose the next state from the open list and execute the policy from that state.*

Open list is sorted by unsolvability (uns) (0 – solvable, 1 – unsolvable) and satisficing heuristic h^{FF}:

	uns	h^{FF}
S_7	0	8
S_3	0	10
S_5	1	7
…	…	…

Unsolvability detection helps to avoid expanding S_5 before S_7.

From the learning perspective, policies help: 1) to guide the search to the parts of the state space favourable for finding a solution; 2) to improve unsolvability detection.

From the planning perspective, heuristics help: 1) to select the next state for policy execution if the policy fails; 2) to select the "good" transition" with the smallest heuristic estimate if there are multiple.

Our preliminary results support the main motivation that the fields of learning and planning can complement each other rather than one superseding the other.

LINKÖPING UNIVERSITY

Machine Reasoning Lab

Farid Musayev
Supervisors: Jendrik Seipp, Daniel Gnad

Acknowledgement: Dominik Drexler

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM