

1 Evaluating Hypothesis $H3$

Next to evaluate hypothesis $H3$, we ran experiments in the eight complexity levels listed in Section ??, with and without including zooming. Table ?? and Table ?? summarize the corresponding results. All trials included \mathcal{ATI} for coarse-resolution reasoning with the adapted theory of intentions, and fine-resolution domain representation for fine-resolution reasoning. As before, the goal in each experimental trial was to find and move a target object to a target location.

Domain			$L1$	$L2$	$L3$
Average Ratios	Total Reasoning Time	Zooming	1.00 (6.24±1.56)	1.00 (8.82±2.95)	1.00 (11.59±3.88)
		Non-Z	1.01 (6.28±1.57)	1.20 (10.74±4.14)	20.07 (225.49±177.64)
	Fine-Resolution Reasoning Time	Zooming	1.00 (1.57±0.51)	1.00 (2.53±0.92)	1.00 (4.19±1.48)
		Non-Z	1.02 (1.61±0.54)	1.71 (4.46±2.29)	55.23 (218.1±176.97)
	Coarse-Resolution Reasoning Time	Zooming	1.00 (4.67±1.05)	1.00 (6.29±2.06)	1.00 (7.4±2.57)
		Non-Z	1.00 (4.67±1.03)	1.00 (6.28±2.06)	1.00 (7.39±2.58)
	Average Time per Refined Plan	Zooming	1.00 (0.37±0.02)	1.00 (0.41±0.02)	1.00 (0.56±0.05)
		Non-Z	1.03 (0.39±0.01)	2.26 (0.93±0.32)	83.49 (45.98±43.05)
Completed Runs		Zooming	100%	100%	100%
		Non-Z	100%	100%	65%

Table 1: Comparing performance of runs with and without zooming. Values of performance measures of reasoning times for non-zooming are expressed as a fraction of the values with zooming. In the averages of complexity level $L3$, we only consider the ratios of those trials in which both runs (with zooming and without zooming) are completed (i.e. we do not get errors due to complexity).

Domain	$L4$	$L5$	$L6$	$L7$	$L8$
Average Total Reasoning Time	18.63±5.49	22.73±7.46	29.8±10.55	41.93±17.26	42.99±18.04
Fine-Resolution Reasoning Time	8.89±2.99	10.78±4.08	17.77±7.55	30.07±15.48	29.36±15.76
Coarse-Resolution Reasoning Time	9.74±3.26	11.95±4.17	12.03±4.33	11.86±4.12	13.63±5.07
Average Time per Refined Plan	0.95±0.26	0.98±0.22	1.64±0.68	2.3±1.02	2.3±1.09

We make the following observations from the results obtained:

- When \mathcal{ATI} is used with zooming, all runs are terminated successfully at the eight different complexity levels (see Table
- The coarse-resolution reasoning time (the total time spent computing plans and diagnosing observations at coarse-resolution level) increases very steadily with the increase in the complexity level for runs with zooming and without zooming, as expected. The fine-resolution reasoning time (i.e. the total time spent computing plans at fine resolution and inferring fine-resolution observations to coarse-resolution observations) also increases with the complexity level for the runs with zooming, but at a reasonable steady pace. However, the increase in fine-resolution

reasoning time is much more pronounced with the absence of zooming taking, in average, over 55% longer to run at complexity level $L3$ than when using zooming.

- The average time spent in each refined plan also scales well with the level of complexity when zooming is used. When zooming is not used the average time spent in each refined plan increases dramatically and only by level $L3$ is, in average, each plan takes 85% longer than those runs using zooming.
- Looking at Table ??, we can see how the reasoning time at coarse-resolution and at fine-resolution increases at a reasonable pace when using zooming, giving evidence of the capability of the system to scale well at higher levels of complexity.
- Looking more closely at the time differences between levels $L7$ and $L8$ we observe that only the coarse-resolution reasoning time is affected. The differences between the domains in these two complexities is simply double the number of coarse-resolution objects, and we kept the refinement of each object consistent (i.e. at most two fine-resolution parts per coarse-resolution object) and the refinement of each room consistent (at 16 grid cells per room). This indicates that the potential effect of an exponential growth in time when adding of objects in the domain is dissipated with zooming.
-

These results provide evidence in support of hypothesis $H3$.