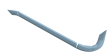
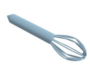
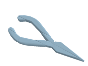
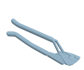
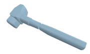
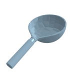
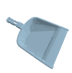
Appendix

A1 Tool Information

Table A1 shows additional information about the tools used in the paper, including a sample view of the object mesh, sample language descriptions, and the train-test split.

Tool	Sample view	Sample Language description	Used for testing
axe		An axe typically has a long, cylindrical handle with a flared end. The head of the axe typically has a slightly curved blade. An axe is often used for chopping wood.	No
chisel		Chisel can make clean, precise cuts using the beleveled edge. A chisel is a hand tool with a blade attached to a handle.	No
crowbar		A crowbar is used as a lever to pry things open. One end of a crowbar is usually curved or hooked so that it can be jammed under an object to apply leverage.	Yes
shovel		A shovel has a long, cylindrical handle and a scoop-shaped blade. The shovel typically has a curved part for digging into and scooping up materials.	No
hacksaw		A hacksaw is a hand saw with a thin blade attached to a handle, used for cutting various such as metal, plastic, or wood. A hacksaw is a saw with a thin, toothed blade on a rigid frame, used for cutting wood or metal.	No
paintroller		One common use of a paint roller is to apply paint evenly to a surface such as walls or ceilings. A paint roller consists of a long, cylindrical body with a handle on one end.	Yes
tweezer		A tweezer is a hand-held tool with two arms that meet at a point. A tweezer is a small hand-held tool with two pointed jaws that are used to pick up small objects or to remove unwanted hair or debris from the body.	No
whisk		A whisk typically has a long, thin handle with a series of loops at the end. The loops are usually made of metal and are arranged in a spiral pattern. A whisk is a common kitchen utensil that is used to mix ingredients together or to incorporate air into a mixture.	No
needlenose		A needlenose plier has a long, tapered nose with a small jaw, and is used for gripping and bending wire. Needlenose pliers are a type of plier that has a long, slender nose and is used for gripping small objects and for working in tight spaces.	NO
plier		A plier is a hand tool used for gripping objects. It consists of a pair of metal jaws with teeth that open and close when the handles are moved. Plier typically has a long, narrow neck and a tapered head that becomes progressively thinner as it extends from the neck to the tip.	Yes
gooseneck		A gooseneck plier is a type of plier that has a long, narrow neck and a slightly curved head. The neck allows the plier to reach into tight spaces, and the curved head provides extra leverage. A gooseneck plier is commonly used to grip and bend small objects.	No
plier-open		The shape of a plier is typically long and skinny with a grip at the end. Plier is a hand tool used for various purposes such as gripping, bending and cutting.	No

mallet		A mallet is a tool that is used to strike another object. A mallet is a type of hammer that usually has a large head and a long handle.	No
hammer		The purpose of a hammer is to strike or hit another object. A hammer typically has a long, cylindrical handle and a heavy head.	Yes
banana		The shape of a banana is generally long and curved, with a thin skin and fleshy inside. A banana is a curved, yellow fruit with a thick peel.	No
fork		A fork is long and thin, with three tines (prongs) at the end. A fork is a utensil that consists of a handle with several narrow tines on one end. The tines are used for piercing food and then lifting it to the mouth.	No
spoon		The purpose of a spoon is to transfer a liquid or semi-solid food from a container to the mouth. A typical spoon consists of a bowl-shaped container with a handle extending from one side. The bowl is generally oval or round, and the handle generally tapers towards the end.	Yes
knife		A knife typically has a sharp, narrow blade with a pointed tip. A knife is a common kitchen utensil used for cutting and slicing food.	No
spatula		A spatula is a kitchen utensil that is used to turn or lift food that is being cooked. It has a flat, usually slightly convex, blade that is attached to a handle. A spatula is commonly used to mix, spread, and flip food items.	No
scissors		A pair of scissors is a cutting tool that consists of two metal blades that are connected at a pivot point. A pair of scissors typically has two blades that are joined at a pivot point.	Yes
wrench		A wrench is a tool that is used to apply torque to an object in order to loosen or tighten it. A wrench is typically long and slender with a small, metal handle.	No
screwdriver		The geometry of a screwdriver can be described as a cylindrical shape with a pointed end. A screwdriver is a tool that is used to insert and remove screws.	No
clamp		A clamp is a mechanical device that is used to temporarily hold two or more objects together. The geometry of a clamp is typically that of a rectangular or U-shaped object with two handles.	No
wok		The shape of a wok is a deep, round bowl with sloping sides. A wok is a concave-shaped cooking utensil that is most commonly used in Chinese cuisine.	No
pickaxe		A pickaxe is used to break up rocks and other materials. A pickaxe is a tool that has a handle attached to a head.	No
faucet		A faucet is typically a small, thin, spout-like fixture that protrudes from a wall or sink. A faucet is a valve used to release water from a plumbing fixture, such as a sink or bathtub.	Yes
dustpan		A dustpan is a tool used for sweeping up dust and small debris from floors and other surfaces. It consists of a small, shallow pan with a handle attached to one side. A dustpan is a concave scoop with a flat bottom and flared sides.	No

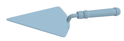
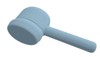
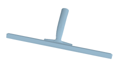
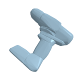
trowel		A trowel is generally a small hand tool with a pointed, scoop-shaped blade on one end and a flat surface on the other. A trowel is a small, hand-held gardening tool with a curved, pointed blade that is used for digging, planting, and transferring small amounts of soil or other materials.	Yes
ladle		A ladle is a tool used to transfer liquids from one container to another. A ladle typically has a long, curved handle and a large, deep, spoon-like bowl.	No
tongs		A pair of tongs has a thin, curved metal shaft with two flat metal paddles at the end. A pair of tongs is a device used to grip and hold objects.	No
gavel		A gavel is a small hammer that is used to strike a sound block, typically made of wood. A gavel is a mallet used to strike a block of wood, typically used by a presiding officer or auctioneer to maintain order or to signal the start and end of an auction.	No
squeegee		The purpose of a squeegee is to remove water or other liquid from a surface. A squeegee is a rod-shaped tool with a flat, blunt edge, and a small handle.	No
powerdrill		A powerdrill is typically cylindrical in shape, with a handle attached to one side and a chuck on the other side for holding drill bits. A power drill is a tool that is used to create holes in various materials, or to fasten screws or bolts.	No
wineglass		A wineglass is a glass with a small bowl and a long stem. They are used to serve wine and are often used in restaurants. A wineglass is typically shaped with a long, thin stem and a bowl that is larger at the bottom than the top.	Yes
shoehorn		A shoehorn is a curved, rod-shaped object used to assist in putting on shoes. A shoehorn is a curved or stepped tool designed to help slide a shoe onto the foot.	No
horseshoe		A horseshoe is a U-shaped metal bar that is nailed to the hooves of a horse. A horseshoe is typically U-shaped, with two large curves and two smaller curves at either end.	No

Table A1: Sample views, sample language descriptions, and the train-test split of the 36 tools considered in the paper.

Fig. 3 shows the t-SNE analysis of the BERT embeddings of all the tools. First we use PCA to project the 768-dimensional embeddings to 50-dimensional, and then perform t-SNE to project them to 2-dimensional for visualization.

A2 Task Information

Table. A2 shows the episode length, reward function, and action space of the tasks. We find the policy can explore well in pushing and lifting tasks with relatively simple reward functions; in sweeping and hammering task, we tune the reward function carefully to guide the arm towards the cylinder/nail. Fig. A1 shows the camera observations for the four tasks. We use a single view for the pushing task as it is sufficient for the task, and dual views for other tasks. A wrist view is used in the lifting task. Fig. A2 visualizes the workspace of the tasks including the initial position of the tools and the target.

For the hammering task, we set the lateral and torsional friction coefficient of the nail to be high (1 and 0.1) in the simulator. We also make the gripper fingers longer to prevent the gripper hitting the block when attempting to hammer the nail.

Please see the included video for more visualization of the tasks.

Task	Episode length	Reward function	Action space
Pushing	25	$\max(0, 1 -$ distance-tool-target	$[-0.05, 0.15] m/s$ in x $[-0.1, 0.1] m/s$ in y $[-\pi/4, \pi/4]$rad$/s$ in yaw
Lifting	25	$0.1 * \max(0, 1 -$ distance-EE-tool$)+$ $0.5 * \max(0, 1 -$ distance-tool-target$)$	$[-0.1, 0.1] m/s$ in x, y, and z $[-\pi/4, \pi/4]$rad$/s$ in yaw
Sweeping	40	$0.1 * \max(0, 1 -$ distance-EE-tool$)+$ $0.1 * \max(0, 1 -$ distance-tool-cylinder$)+$ $0.5 * \max(0, 1 -$ distance-cylinder-target$)$	$[-0.2, 0.2] m/s$ in x, y, and z $[-\pi/4, \pi/4]$rad$/s$ in yaw
Hammering	40	$0.1 * \max(0, 1 -$ distance-EE-tool$)+$ $0.1 * \max(0, 1 -$ distance-tool-nail$)+$ $0.5 * \max(0, 1 -$ distance-nail-hole_end$)$	$[-0.2, 0.2] m/s$ in x, y, and z $[-\pi/4, \pi/4]$rad$/s$ in yaw

Table A2: Episode length, reward function, and action space for the four tasks. Distance-{A}-{B} denotes distance from A to B, normalized by the initial distance. EE denotes end-effector of the arm. See Fig. A2 for visualization of the task space and target.

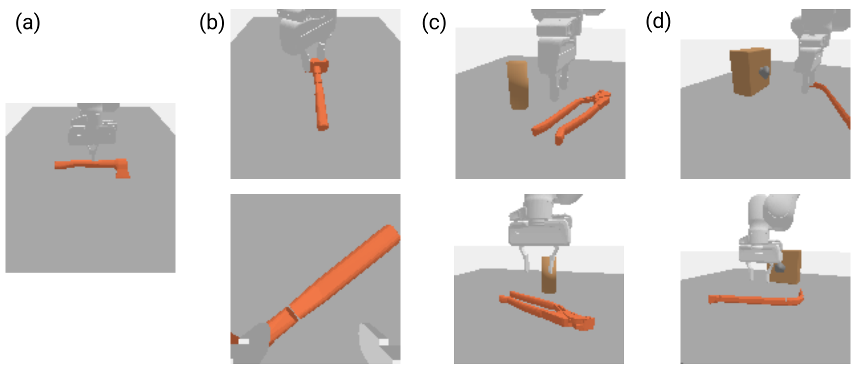


Figure A1: Camera observations of the tasks: (a) pushing (single view only); (b) lifting (including a wrist view); (c) sweeping; (d) hammering.

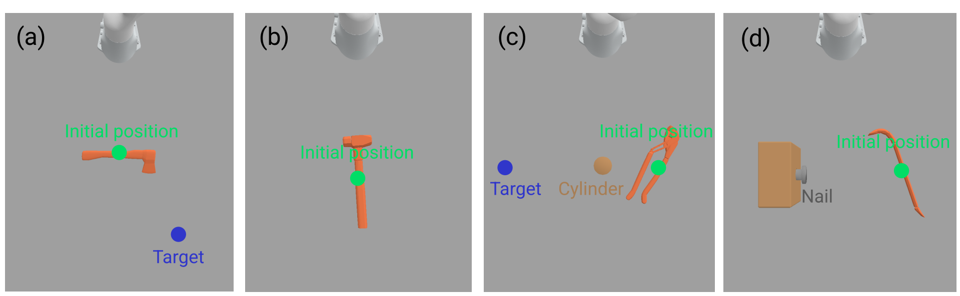


Figure A2: Top-down visualization of the workspace of the tasks: (a) pushing; (b) lifting; (c) sweeping; (d) hammering.

A3 Model Architecture

For all policies, the image encoder π_o contains three convolutional layers for either view of the image input; the three layers have kernel size $[7 \times 7, 5 \times 5, 3 \times 3]$, stride size $[4, 3, 2]$, no padding, and channel size $[4, 8, 16]$. The language head π_l contains a single fully-connected layers with 128-dimensional output. Both the actor head π_a and critic head π_c have two hidden layers of size 128 (except for AT-XL with hidden size 256). All convolutional layers and fully-connected layers are followed with a ReLU activation. The first layer in π_a and π_c are additionally normalized by Layernorm [46].

A4 Multi-environment Training Algorithm

Algorithm 3 below shows the procedures of the SAC and SAC-LA baselines. Multi-environment training does not involve any inner adaptation ($B = 0$). In the context of multi-environment training, N means the number of environments where, experiences that each gradient update uses, belong. $N = \infty$ means experiences can come from any environment.

Algorithm 3 SAC, SAC-LA

Require: $\mathcal{T} = \{\tau_i\}_{i=1}^{K}$: training set of tools; $\{L_i\}_{i=1}^{K}$: sets of language descriptions (for SAC-LA only); θ_l (for SAC-LA only), $\theta_o, \theta_a, \theta_c$: policy modules; β: replay buffer; N

while training **do**
    Sample τ_i from T; reset environment with τ_i
    Collect episode (with $l_{ij} \sim L_i$ for SAC-LA); add to β
    if $N = \infty$ **then**
        Sample experiences from β that are from any environment
    else
        Sample experiences from β that are from N different environments
    end if
    Update θ_o, θ_l (only for SAC-LA), θ_c, θ_a with sampled experiences
end while

A5 Training Hyper-parameters

The hyper-parameters used for meta-learning (shared among AT-LA, AT-TinyLA, AT, AT-XL) and multi-environment learning (shared between SAC-LA, SAC) are outlined in Table A3. We ensure meta-learning and multi-environment learning sample the same amount of transitions from environments.

<table>
<tr><th>Setting</th><th>Meta-learning</th><th>Multi-environment learning</th></tr>
<tr><td># training steps</td><td>1000 (iteration)</td><td>2.5e6 (pushing/lifting)
4e6 (sweeping/hammering)</td></tr>
<tr><td>Meta replay buffer size</td><td>30000</td><td>—</td></tr>
<tr><td>Base replay buffer size</td><td>∞</td><td>100000</td></tr>
<tr><td>Replay ratio</td><td colspan="2">16</td></tr>
<tr><td>N</td><td>1</td><td>∞</td></tr>
<tr><td>M</td><td>2</td><td>—</td></tr>
<tr><td>B</td><td>5</td><td>0</td></tr>
<tr><td>B_ν</td><td colspan="2">10</td></tr>
<tr><td>Optimization</td><td></td><td></td></tr>
<tr><td>Optimizer</td><td colspan="2">Adam</td></tr>
<tr><td>Batch size</td><td colspan="2">128</td></tr>
<tr><td>Discount factor</td><td colspan="2">0.99</td></tr>
<tr><td>SAC entropy coefficient</td><td colspan="2">0.01</td></tr>
<tr><td>SAC actor update period</td><td colspan="2">1</td></tr>
<tr><td>Base learning rate</td><td colspan="2">3e-4</td></tr>
<tr><td>Meta learning rate</td><td>1e-3</td><td>—</td></tr>
<tr><td>Hardware Resource</td><td></td><td></td></tr>
<tr><td># CPU threads</td><td colspan="2">20</td></tr>
<tr><td>GPU</td><td colspan="2">Nvidia RTX 2080Ti</td></tr>
<tr><td># hours for runtime</td><td colspan="2">6 (pushing/lifting), 10 (sweeping), 16 (hammering)</td></tr>
</table>

Table A3: Hyper-parameters used for meta-learning and multi-environment learning.

A6 Results and Additional Studies

Table A4 below shows the results in Fig. 4 in numbers.

Method	Crowbar	Paint Roller	Plier	Hammer	Spoon	Scissors	Faucet	Trowel	Wineglass
Pushing									
ATLA	**15.4 ± 0.6**	**19.9 ± 0.2**	19.9 ± 0.4	**11.1 ± 0.1**	10.0 ± 0.3	**18.2 ± 0.5**	**14.4 ± 0.5**	**16.9 ± 2.2**	4.5 ± 0.3
AT-TinyLA	12.3 ± 1.8	19.2 ± 0.9	**20.4 ± 0.2**	10.7 ± 0.3	10.5 ± 0.8	17.0 ± 0.8	8.8 ± 0.9	12.5 ± 1.1	7.4 ± 1.6
AT	7.8 ± 0.3	17.4 ± 1.0	18.4 ± 0.5	10.0 ± 0.4	9.4 ± 0.2	15.6 ± 0.4	9.6 ± 1.2	15.8 ± 1.5	6.1 ± 0.9
AT-XL	12.2 ± 2.5	19.0 ± 0.8	19.6 ± 0.5	10.8 ± 0.1	**11.3 ± 0.2**	16.3 ± 1.4	14.1 ± 1.8	15.4 ± 0.7	7.1 ± 0.6
SAC-LA	13.4 ± 0.8	16.8 ± 0.6	18.4 ± 0.5	10.7 ± 1.0	7.8 ± 1.9	14.2 ± 1.3	10.1 ± 1.6	12.7 ± 2.5	7.8 ± 1.1
SAC-LA-N=1	13.8 ± 0.9	15.9 ± 0.3	17.8 ± 0.7	10.5 ± 1.1	8.2 ± 2.0	13.2 ± 1.0	10.3 ± 1.8	14.2 ± 3.0	**7.9 ± 1.2**
SAC	12.9 ± 1.1	12.2 ± 4.7	16.9 ± 1.5	10.6 ± 0.6	9.2 ± 1.5	12.0 ± 5.6	10.4 ± 3.8	11.2 ± 3.3	6.4 ± 0.2
Lifting									
ATLA	**7.4 ± 0.8**	**6.5 ± 0.8**	**9.0 ± 0.9**	9.6 ± 1.3	**5.5 ± 1.0**	**2.1 ± 0.0**	**3.5 ± 0.5**	**5.0 ± 0.5**	2.3 ± 0.6
AT-TinyLA	2.6 ± 0.3	5.7 ± 0.8	9.0 ± 1.2	**10.4 ± 0.3**	5.0 ± 0.8	2.0 ± 0.4	2.3 ± 0.4	1.9 ± 0.2	2.4 ± 0.5
AT	5.0 ± 0.5	4.6 ± 0.8	7.7 ± 1.3	9.1 ± 2.3	3.7 ± 0.8	2.0 ± 0.3	1.8 ± 0.2	3.4 ± 0.9	**2.5 ± 0.6**
AT-XL	6.8 ± 1.0	5.7 ± 1.0	8.5 ± 0.3	9.0 ± 0.4	4.9 ± 0.8	1.8 ± 0.0	2.4 ± 0.2	4.2 ± 1.2	2.9 ± 0.7
SAC-LA	5.1 ± 0.1	4.4 ± 0.3	7.7 ± 0.2	8.3 ± 0.1	4.9 ± 0.1	1.8 ± 0.0	3.0 ± 0.1	3.8 ± 0.1	1.8 ± 0.1
SAC-LA-N=1	5.5 ± 0.3	4.2 ± 0.3	8.0 ± 0.5	8.1 ± 0.2	4.5 ± 0.1	1.8 ± 0.0	3.2 ± 0.1	3.6 ± 0.2	1.9 ± 0.2
SAC	6.6 ± 1.5	4.9 ± 1.1	6.9 ± 1.4	7.9 ± 1.5	4.2 ± 0.6	1.6 ± 0.6	3.0 ± 0.4	3.6 ± 0.6	2.2 ± 0.6
Sweeping									
ATLA	**25.5 ± 1.0**	5.2 ± 2.4	**18.4 ± 3.5**	**26.8 ± 2.6**	**22.5 ± 2.7**	**21.7 ± 2.0**	**20.9 ± 1.6**	**24.3 ± 2.2**	**14.4 ± 1.8**
AT-TinyLA	21.2 ± 0.7	4.6 ± 0.2	16.5 ± 0.6	22.3 ± 0.8	18.0 ± 0.1	19.2 ± 1.0	15.2 ± 0.8	21.2 ± 0.6	12.5 ± 0.4
AT	20.0 ± 0.8	7.7 ± 3.7	17.4 ± 3.5	19.3 ± 1.5	18.4 ± 1.3	18.0 ± 3.0	18.0 ± 3.2	14.9 ± 2.9	12.0 ± 2.2
AT-XL	19.0 ± 3.0	6.1 ± 1.4	18.2 ± 2.8	21.0 ± 2.8	17.1 ± 2.4	16.8 ± 3.7	16.3 ± 3.4	15.1 ± 8.4	7.6 ± 3.6
SAC-LA	18.4 ± 4.0	**9.0 ± 3.7**	15.5 ± 3.2	20.7 ± 2.2	17.4 ± 3.0	18.4 ± 3.7	16.7 ± 2.9	18.6 ± 1.8	12.2 ± 1.1
SAC-LA-N=1	16.8 ± 3.2	7.5 ± 2.2	16.1 ± 3.5	21.9 ± 1.7	17.1 ± 2.8	17.9 ± 4.0	17.1 ± 2.6	18.1 ± 1.9	12.6 ± 1.6
SAC	17.6 ± 2.5	7.7 ± 2.1	12.4 ± 2.4	12.3 ± 4.7	17.3 ± 5.4	17.8 ± 2.4	15.9 ± 2.8	17.1 ± 0.2	10.6 ± 3.3
Hammering									
ATLA	**12.9 ± 1.0**	**12.5 ± 1.3**	**10.9 ± 1.7**	14.4 ± 0.9	**12.4 ± 1.3**	3.9 ± 0.3	3.8 ± 0.6	**11.7 ± 1.2**	10.4 ± 1.5
AT-TinyLA	12.1 ± 1.9	12.0 ± 2.3	10.8 ± 1.5	**16.8 ± 1.0**	11.5 ± 1.4	**4.4 ± 0.2**	**4.1 ± 0.8**	10.3 ± 0.8	**11.2 ± 0.8**
AT	6.0 ± 0.9	9.7 ± 2.0	10.3 ± 2.6	12.4 ± 1.9	7.2 ± 1.7	3.7 ± 0.1	3.1 ± 0.1	9.0 ± 2.8	7.1 ± 2.0
AT-XL	6.5 ± 1.2	8.3 ± 1.5	11.3 ± 1.6	12.1 ± 1.7	7.2 ± 1.6	3.7 ± 0.1	3.1 ± 0.3	10.4 ± 3.4	8.2 ± 1.7
SAC-LA	4.3 ± 1.1	3.8 ± 1.3	9.0 ± 3.2	9.3 ± 2.0	7.8 ± 1.6	3.2 ± 0.5	3.3 ± 0.3	4.7 ± 1.3	3.0 ± 0.7
SAC-LA-N=1	5.3 ± 1.5	6.9 ± 1.1	10.0 ± 2.5	11.5 ± 1.5	9.8 ± 2.9	3.6 ± 0.7	3.8 ± 0.2	8.9 ± 1.6	6.8 ± 1.0
SAC	5.3 ± 3.1	5.1 ± 3.2	6.8 ± 4.2	7.4 ± 5.3	3.4 ± 1.1	3.0 ± 0.6	3.2 ± 0.6	3.5 ± 0.4	3.2 ± 1.0

Table A4: Post-adaptation reward in mean and standard deviation over 3 seeds across 4 tasks and 9 test tools.

Effect of B in meta-learning. The value of B determines the number of inner adaptation for each tool during meta-training. To evaluate the effect of B, we perform sensitivity analysis varying B in the sweeping task, and the results are shown in Table A5 (values for $B = 5$ are from Table A4). With $B = 1$, the post-adaptation performance is generally worse than $B > 1$, and it is close to the performance of multi-environment training (SAC-LA in Table A4). Higher value of B helps, but there is no significant performance gain with $B > 2$. This highlights the importance of performing a few steps of inner adaptation at training time for better adaptation to new tools.

B	Sweeping: Crowbar	Paint Roller	Plier	Hammer	Spoon	Scissors	Faucet	Trowel	Wineglass
1	18.6 ± 1.1	7.2 ± 2.5	16.1 ± 2.2	20.9 ± 1.8	19.3 ± 1.8	18.2 ± 2.9	16.2 ± 0.8	19.3 ± 2.2	12.7 ± 1.1
2	21.2 ± 0.9	**7.3 ± 2.0**	17.2 ± 2.5	22.8 ± 2.1	22.1 ± 2.1	**22.8 ± 2.3**	18.9 ± 1.8	21.2 ± 2.8	13.8 ± 1.7
4	22.6 ± 0.8	5.9 ± 1.8	17.6 ± 2.3	**27.8 ± 2.9**	20.1 ± 3.0	**20.2 ± 2.2**	**21.1 ± 1.9**	23.2 ± 2.9	**14.6 ± 1.9**
5	**25.5 ± 1.0**	5.2 ± 2.4	**18.4 ± 3.5**	26.8 ± 2.6	**22.5 ± 2.7**	21.7 ± 2.0	20.9 ± 1.6	**24.3 ± 2.2**	14.4 ± 1.8

Table A5: Effect of B in meta-learning. The values are post-adaptation reward in mean and standard deviation over 3 seeds across 9 test tools in the sweeping task.

Effect of N in meta-learning. The value of N determines how many environments the meta update gradient is averaged over in Eq. 1. To evaluate the effect of N, we perform sensitivity analysis varying N in the hammering task, and the results are shown in Table A6 (values for $N = 1$ are from Table A4). Although typical Reptile-style meta learning in supervised learning uses $N > 1$ [5], the results here do not show the improvement with $N > 1$. $N = 1$ also matches our setup of adapting to a single tool at test time. We also show the meta training reward curves in Fig. A3. The final training rewards are similar for $N = 1, 2, 5$, but $N = 1$ trains slightly faster. Surprisingly we also find $N > 1$ exhibits larger variances in reward at later iterations, which is contrary to the idea that $N > 1$ stabilizes training. Thus we use $N = 1$ in our main experiments.

	Hammering								
N	Crowbar	Paint Roller	Plier	Hammer	Spoon	Scissors	Faucet	Trowel	Wineglass
1	**12.9** $\pm$ **1.0**	12.5 $\pm$ 1.3	10.9 $\pm$ 1.7	**14.4** $\pm$ **0.9**	12.4 $\pm$ 1.3	3.9 $\pm$ 0.3	**3.8** $\pm$ **0.6**	11.7 $\pm$ 1.2	**10.4** $\pm$ **1.5**
2	12.2 $\pm$ 1.2	**12.9** $\pm$ **1.5**	9.7 $\pm$ 1.2	14.2 $\pm$ 1.1	12.9 $\pm$ 1.5	4.0 $\pm$ 0.2	3.6 $\pm$ 0.3	**12.0** $\pm$ **1.6**	10.1 $\pm$ 1.8
5	12.5 $\pm$ 1.0	12.5 $\pm$ 1.7	**11.2** $\pm$ **1.3**	11.5 $\pm$ 1.2	**13.2** $\pm$ **1.1**	**4.2** $\pm$ **0.5**	3.7 $\pm$ 0.6	**12.0** $\pm$ **1.5**	9.9 $\pm$ 1.3

Table A6: Effect of N in meta-learning. The values are post-adaptation reward in mean and standard deviation over 3 seeds across 9 test tools in the hammering task.

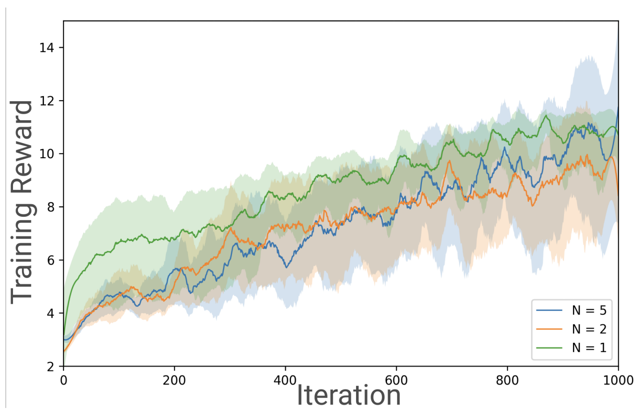


Figure A3: Meta training reward for different values of N in the hammering task running ATLA.

SAC-LA with $N = 1$. Typically multi-environment training uses gradient update averaged over a batch of experiences sampled from any environment ($N = \infty$); we perform the same setup in our main experiments. It is also worth investigating $N = 1$, meaning each gradient update uses experiences from only one environment, since we find $N = 1$ works for the meta learning setup as shown above. In Table A4 we show the results for all four tasks (SAC-LA-N=1). In Pushing, Lifting, and Sweeping tasks, the results are similar to SAC-LA with $N = \infty$. However, in the hammering task, the results are improved from SAC-LA across all tools, although they are still worse than meta-learning baselines. It is possible that $N = 1$ mitigates training instability of multi-environment training. We use $N = \infty$ in our main experiments since it is a common setup for multi-environment training.