

Figure 6: **Knowledge exploitation efficiency**: We examined how quickly our models exploited the information they gathered. We defined a scoring function that assessed how well our models predictions matched the ground-truth and evaluated the models' predictions after every step. Results are from conjunction tasks, using Gemini 1.5 Pro.

A ADDITIONAL RESULTS

Here we present some additional results.

753 A.1 KNOWLEDGE EXPLOITATION EFFICIENCY

755 We examined how well Gemini exploited the information they gathered through exploration (see Figure 6). The scoring function resolves to 1 if every one of the words in a target string (defined



Test investment         Frequencies           Single Favor Tak         You are playing a text-based game. Your goal is to discover how to earn rewards. Game Rules: - Find as which up the Same object relation. - There are objects with difference colors, shapes, and textures. - Ficking up an object gives you a reward (either 0 or 1). - The same object always gives the same reward. - A specific property, such as a particular color OR shape determines the reward. - A specific property, such as a particular color OR shape determines the reward. Within the relevant factor, ship OR specific ools: OR Specific object is able with difference of the specific object of DR shape leads to a reward.		
Nat.         Fromp!           Displaying a text-based game. Your goal is to discover how to earn rewards. Come Rules: - Find as what factors lead to reward as quickly as possible. - You cannot pick up the same object twice. - There are objects with different colors, shapes, and textures. - There are objects with different colors, shapes, and textures. - There are object always quicy the same reward. - A specific property, such as a particular color OS shape leads to a reward, find out what it f The reward is binary (0 or 1). Only one factor, either color OS shape leads to a reward, find out what it f relevant factor, any OSM specific color or OSM specific shape will yield a reward of 1. If the chosen object matches this correct color (When color is the factor) OS this correct shape (when i is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description) Important: You have VENY FEM turns left. Choose your next action carefully to maximize information. To are an Ai agent designed for thoughtful apploation. Your mission is to mayingte and lears within a prior environment by performing actions and their consequences. Apploration to gain basic understanding. - **Actions*: Choose an action to perform within the environment. Initially, this may involve randem exploration to gain basic understanding. - **Actions*: Description) - **Actions*: Description (your action. This includes any drages to the environment? Analyze that data to identify patterning. - **Actions*: Basic do your review, reason about the environment. Analyze that data to identify patterning. - **Actions*: Basic do your review, reason about the environment? Analyze that addata identify patterning. - **Actions*: Basic do your review, reason about the environment? Analyze that addata identify patterning or disting and effect relationably. - **Actions*: Basic do your review, reason about the environment? Analyze that addata identhis that seem particularity forming or distingentify of		
Text         Prompt           TextEnvironment Number haver Taxt         You are playing a text-based game. Your goal is to discover how to earn rewards. Clame Rains: - Find as what factors lead to reward as quickly as possible. - You cannot pick up the same object twice. - There are objects with different colors; shapes, and textures. - Find as what factors lead to reward (stiff) or 01. - A specific property, such as a particular onlor OR shape determines the reward, find out what it i relevant factor, only OME specific color or OME specific shape will yield a reward of 1. If the chosen object maches this correct color (when color is the factor) OR this correct shape (when r is the factor), only OME specific color or OME specific shape will yield a reward of 1. If the chosen object maches this correct color (when color is the factor) OR this correct shape (when r is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description) Important: You have UKEY FEW turns left. Choose your next action carefully to maximize information. You are an Al agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their commentum. Initially, this may involve random plotuline: Chois and crimiter temping. vobserver: Mestratin detailed tog of your action. This includes any changes to the environment and any rewards or penalities received. vewer's Mestratin detailed log of your actions, observations, and received rewards. vewer's Mestratin detailed log of your actions, observations, and received rewards. vewer's Mestratin detailed log of your actions, observations, and received rewards. vewer's Mestratin detailed log of your actions, observations, and received rewards. vewer's Mestratin detailed log of your actions, bacterute the environment? A retermine the provinsense that		
Ext         Prompt           The Environment Single Flewy Tax         You are playing a text-based game. Your goal is to discover how to earn rewards. Game Rules: - Find as what factors lead to reward as quickly as possible. - You cannot pick up the same object twice. - There are objects with different colors, shapes, and textures. - The same object always gives the same reward. - A specific property, such as a particular color OS shape leads to a reward, find out what it is the area object always gives the same reward. - A specific property, such as a particular color OS shape. determines the reward. Within the releases factor), the reward is in otherwise, the reward is 10 juits a reward of 1. If the chosen object matches this correct color (when color is the factor) OS this correct shape (when a is the factor), the reward is 1. Otherwise, the reward is 0. iscene_description) Important: You have VENY FEN turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration, Your mission is to navigate and learn within a given environment by performing actions and obsering the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration to gain hasic understanding. - «Observe: Observe the result of your action. This includes any changes to the environment and any rescence its basins is detain in going or your action, the environment. Initially, this may involve random exploration to gain hasic understanding. - «Observe: Teriodically, pause to exploitly review your action history and the environment? Not hypothesis can you form about the underlying rules or structure of the environment? Not hypothesis can you form about the underlying rules or structure of the environment? Not hypothesis can you form about the underlying rules or structure of action is sequence of action. As to text your hypothesis and gather mores information. - «**Pacever: Basee		
Back         Prompt           Tour Letwinnews         You are playing a text-based game. Your goal is to discover how to earn rewards. Game Rules: - Find as what factors lead to reward as quickly as possible. - You cannot pick up the same object twice. - There are objects with different colors, shapes, and textures. - There are object by the different color of the shape leads to a reward, find out what it 2 - The same object always gives the same reward. - A specific property, such as a particular color of these leads to a reward, find out what it 2 - The reward is binary (0 or 1). Only one factor, either color GR shape, determines the reward. Within the retermine factor, only OWE perific color or OWE specific shape will yield a reward of 1. The reward is 1. Otherwise, the reward is 0. (scemdescription) Is the factor), the reward is 1. Otherwise, the reward is 0. (scemdescription) reportant: You have VENF PEW turns left. Choose your next action carefully to maximize information. You are not added for thoughtful exploration. Your mission is to navigate and hearn within a rive environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and observing the outcomes. Operate as a scientist, carefully considering your actions and observing the outcomes. Operate as a scientist, carefully considering your actions and observing the outcomes. reword or penalties received. - **Recorder: Maintin a detailed log of your action. This includes any changes to the environment and any rewards or your form about the underlying rules or structure of the environment? reduction: penalties received. -		
Text Eventset         Prompt           Text Environment Ningle favor Text         You are playing a text-based game. Your goal is to discover how to earn rewards. Came Eules: - Find as what factors lead to reward as quickly as possible. - You cannot pick up the same object twice. - Ficking up an object alway gives the same reward. - A specific property, such as a particular color OK shape leads to a reward, find out what it is The reward is binary (0 or 1). Only one factor, either color OK shape determines the reward. Within the relevant factor, entry OK specific color or OK specific shape will yield a reward of 1. If the chosen object matches this correct color (Whan color is the factor) OK this correct shape (when r is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description)           Important: You have VENY FEN turns left, Choose your next action carefully to maximize information. You are an Al specific diagond for thoughtful application. Your mission is to navigate and learn within a considering your actions and their consequences. Exploration Cycle: - watchine*: Otherwise the result of your actions. This includes any changes to the environment and any rewards or penalties received. - webserve*: Meriodically, pause to explicitly preview your action history and the corregonding outcome haalyze this data to identify patterns, trands, and potential cause-andrefet relationships. - where any actions that asem particularly promising or detrimental? Do carding expending actions of hoppethesis, plan your next action or sequence of actions. Alin to action_reward_description)           Respond with this format, please be specific about the object: * Action_reward_description)         - who are sequence of actions. Alin to castion provesses of action is all to predictable outcomes? - witypothesizes: Clearly state your current hypothesis about the most effective strategy		
Description         Prompt           Sign Fixed Test         You are playing a text-based game. Your goal is to discover how to earn rewards.           Came Bules:         - Find as what factors lead to reward as quickly as possible.           - You cannot pick up the same object twice.         - There are objects with different colors, shapes, and textures.           - Dicking up an object gives the same reward.         - A specific property, such as a particular color OR shape, determines the reward. Within the relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1.           If the chosen object matches this correct color (Men color is the factor) OR this correct shape (when of is the factor), the reward is 1. Otherwise, the reward is 0.           (sccen_description)         Important: You have VENY FEW turns left. Choose your next action carefully to maximize information.           Two are an Al agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientise, carefully considering your actions and their consequences.           Diporation Cycle:         - **Action*:: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.           - **Action*:: Based on your review, reason about the environment. Initially, this may involve random exploration to gain basic understanding.           - **Action*:: Based on your review, reason about the environment.           - **Neason*:: Based on your review, reason about the environment?		
These         Prompt           The Environment         You are playing a text-based game. Your goal is to discover how to earn rewards.           Game Rules:         - Find as what for top lad to reward as quotely as possible.           - Find as what for top lade to reward as quotely as possible.           - These are objects with different colors, shapes, and textures.           - Find as what for you are ward (either 0 or 1).           - The same object always gives the same reward.           - A specific property, such as a particular color G& shape leads to a reward, find out what it is relevant factor, only ONE specific color or ONE specific alape will yield a reward of 1.           If the chosen object mutches this correct color (Weshape leads to a reward, find out what it is a the factor), the reward is 1. Otherwise, the reward is 0.           (scene_description)           Important: You have VENY FEN turns left. Choose your next action carefully to maximize information.           You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences.           Exploration Cycle:         - *Action*: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.           - *Action*: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.           - *Action*: Choose an actio		
<pre>Notionaments Note: The set of the set o</pre>	Task Taxt Environment	Prompt
<pre>Game Rules: - Find as what factors lead to reward as quickly as possible. - There are objects with different colors, shapes, and textures. - There are object suivay gives the same reward. - There are object always gives the same reward. - A specific property, suival as a particular color OR shape leads to a reward, find out what it : - A specific property. suival as a particular color OR shape, determines the reward. Within the relevant factor, only ONE specific color on ONE specific shape will yield a reward of 1. If the chosen object matches this correct color (when color is the factor) OR this correct shape (when s is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description) Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration, Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle: - **Action**: Choose an action to perform within the environment. Initially, this may involve random equication to gain basic understanding. - **Rectori**: Maintain a detailed log of your actions, observations, and received reward. - **Rectori**: Maintain a detailed log of your actions, observations, and received reward. **Record**: Maintain a detailed log of your actions and the invorment. What hypotheses can your form about the underlying rules or structure of the environment? A relevant sequences of actions lead to predictable outcomes? ***Record**: Faintions that seem particularly promising or detrimenta? The tare any actions that seem particularly promising or detrimenta? ***Record**: Faintions that seem particularly promising or detrimenta? ***Record**: Second the subject in the object: *********************************</pre>	Single Factor Task	You are playing a text-based game. Your goal is to discover how to earn rewards.
<ul> <li>Find as what factors lead to reward as quickly as possible.</li> <li>You cannot pick up the same object twice.</li> <li>There are objects with different colors, shapes, and textures.</li> <li>Picking up on object gives you a reward (either 0 or 1).</li> <li>The same object always gives the same reward.</li> <li>A specific property, such as a particular color OR shape leads to a reward, find out what it :</li> <li>The reward is binary (0 or 1). Only one factor, either color OR shape. determines the reward at 1.</li> <li>If the chosen object matches this correct color (when color is the factor) OR this correct shape (when rise is the factor), the reward is 1. Otherwise, the reward is 0.</li> <li>(scens_description)</li> <li>Important: You have VERY FEW turns left. Choose your next action carefully to maximize information.</li> <li>You are not agent designed for thoughtful exploration. Your mission is to navigate and learn within a given evironment by pestorming actions and exerving the outcomes. Operate as a scientist, carefully considering your actions and their consequences.</li> <li>Exploration Cycle:         <ul> <li>*Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.</li> <li>**Ouether the factor of your actions, and prential cause-advefact relationships.</li> <li>**Reaction the received.</li> <li>**Reaction and your reverse.</li> </ul> </li> <li>**Reaction and your reverse reaction the environment. Initially, this may involve random exploration to gain basic understanding.</li> <li>**Reaction and reaction do your actions, and prential cause-advefact relationships.</li> <li>**Reaction a gain basic understanding.</li> <li>**Reaction and your reverse reaction the environment.</li> <li>**Reaction that seeme particularly promising or detrimental?</li> <ul> <li>**Reaction t</li></ul></ul>		Came Rules.
<ul> <li>You cannot pick up the same object twice.</li> <li>There are objects with different colors, shapes, and textures.</li> <li>Picking up an object gives you a reward (either 0 or 1).</li> <li>The same object lawy gives the same reward.</li> <li>A specific property, such as a particular color OR shape leads to a reward, find out what it :</li> <li>The reward is binary (0 or 1). Only one factor, either color OR shape, determines the reward. Within the relevant factor, noty OWE specific color or OWE specific shape will yield a reward of 1.</li> <li>If the chosen object matches this correct color (West paper will yield a reward of 1.</li> <li>If the chosen object matches this correct color (when color is the factor) OR this correct shape (when rist is the factor).</li> <li>Important: You have VERY FEW turns left. Choose your next action carefully to maximize information.</li> <li>You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes.</li> <li>Exploration Cycle:</li> <li>**Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.</li> <li>**Record**: Maintain a detailed log of your action, observations, and received rewards.</li> <li>**Record**: Maintain a datailed log of your action, observation history and the corresponding outcome haalyze this data to identify patterns, trends, and potential course and the environment?</li> <li>**Prothesize*** Clearing state your current hypothesis about the most effective strategy for exploration or action is and should be correse?</li> <li>**Plan**: Based on your review, reason about the outcomes?</li> <li>**Record**: Based on your review, reason about the outcomes?</li> <li>**Record**: Based on your review, reason about the most effective strategy for exploration carding measses and your review, reason abou</li></ul>		- Find as what factors lead to reward as quickly as possible.
<ul> <li>Picking up an object gives you a reward (either 0 or 1).</li> <li>The same object always gives the same reward.</li> <li>A specific property, such as a particular color OR shape leads to a reward, find out what it :</li> <li>The reward is binary (0 or 1). Only one factor, either color OR shape, determines the reward. Within the relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1.</li> <li>If the chosen object matches this correct color (when color is the factor) OR this correct shape (when is is the factor), the reward is 1. Otherwise, the reward is 0.</li> <li>(scene_description)</li> <li>Important: You have VERY FEW turns left. Choose your next action carefully to maximize information.</li> <li>You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences.</li> <li>Exploration Cycle:         <ul> <li>+*Action*:: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.</li> <li>+*Review*:: Periodically, pause to explicitly review your action history and the corresponding outcome hanlyze this data to identify patterns, trends, and potential cause-and-effect relationships.</li> <li>+*Reason**: Based on your review, reason about the environment.</li> <li>What hypothesis can you form about the underlying rules or structure of the environment?</li> <li>Are there any actions that seem particularly promising or durinenta?</li> <li>**Pla**: Based on your review, reason about the object:</li> <li>*Are there any actions bat seem particularly promising or durinenta?</li> <li>**Pla**: Stage or you from bout the underlying rules or structure of the environment?</li> <li>**Pla**:</li></ul></li></ul>		<ul> <li>You cannot pick up the same object twice.</li> <li>There are objects with different colors, shapes, and textures.</li> </ul>
<ul> <li>In same object always gives the same reward.</li> <li>A specific property, such as a particular color OR shape leads to a reward, find out what it :</li> <li>The reward is binary (0 or 1). Only one factor, either color OR shape, determines the reward. Within the relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1.</li> <li>If the chosen object matches this correct color (when color is the factor) OR this correct shape (when i is the factor), the reward is 1. Otherwise, the reward is 0.</li> <li>(scene_description)</li> <li>Important: You have VERY FEW turns left. Choose your next action carefully to maximize information.</li> <li>You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences.</li> <li>Exploration Cycle:         <ul> <li>**Action:*: Choose an action to perform within the environment. Initially, this may involve random exploration the order to destring for your actions, and received rewards.</li> <li>**Neview**: feriodically, pause to explicitly review your action history and the corresponding outcome halvy this data to identify patterns, trends, and potential cause-and-effect relationships.</li> <li>**Neview**: Feriodically, pause to explicitly review your action history and the environment? Are there any actions that seem particularly promising or detrimental?</li> <li>**Neview**: Sciencidas that seem particularly promising or detrimental?</li> <li>**Inpothesize**: Clearly state your current hypothesis about the most effective strategy for exploration cateforing and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information.</li> </ul> </li> <li>(action_reward_description)</li> <li>Respond with this</li></ul>		- Picking up an object gives you a reward (either 0 or 1).
The reward is binary (0 or 1). Only one factor, either color OR shape, determines the reward. Within the relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1. If the chosen object matches this correct color (when color is the factor) OR this correct shape (when relist the factor). The reward is 1. Otherwise, the reward is 0. (scene_description) Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle: **Action*: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. **Networks for penallies received. **Retion*:: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. **Retion*:: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. **Reterve*: Observe the result of your actions, observations, and received rewards. **Reterve*: Observe the result of your actions, observations, and received rewards. **Reterve*: Observe the result of your actions, and protential acue-and-effect relationships. **Reterve*: Observe*: Observe*: Observations, and received rewards. **Reterve*: Observe*: Observe*: on about the environment. **Reterve*: Observe*: Observe*: Clearly actions have the environment? **Reterve*: Observe*: Clearly state your current hypothesis about the most effective strategy for exploration achieving a goal. **Here any actions that seem particularly promising or detrimental? **Grieving a goal. **State on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information.		<ul> <li>The same object always gives the same reward.</li> <li>A specific property, such as a particular color OR shape leads to a reward, find out what it is</li> </ul>
The reward is binary (0 or 1). Only one factor, either color OR shape, determines the reward. Within the relevant factor, only ONE specific oshape will yield a reward of 1. If the chosen object matches this correct color (when color is the factor) OR this correct shape (when a is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description) Important; You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle: - **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding **Observe*: Observe the result of your actions, observations, and received rewards **Reson**: Hanitain a detailed log of your actions, observations, and received rewards **Reson**: Based on your review, reason about the environment. What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? - **Observe*: Observe the reason about the environment. What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? - **Observe*: Based on your review, reason about the most effective strategy for exploration or addition and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. (action_reward_description) Respond with this format, please be specific about the object:  * Action: pick up (colored > colpect>  * Stop: <kes <no="" or=""> *  * Which factor influence reward? <colon <shape="" or=""> or <unsure> * Which factor influence reward? <colon <shape="" or=""> or <unsure> * Which factor influence reward? <colon <shape<="" or="" td=""><td></td><td></td></colon></unsure></colon></unsure></colon></kes>		
<pre>relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1. If the chosen object matches this correct color (when color is the factor) OR this correct shape (when r is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description) Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle: - **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. - **Observe*:: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received. - **Record**: Mintain a detailed log of your action, observations, and received rewards. - **Resort*: Mointain a detailed log of your action is tructure of the environment? Are there any actions that seem particularly promising or dstructure of the environment? - **Resort*: Second your creise, reason about the environment. - **Retor any actions that seem particularly promising or dstructure of the environment? - **Retard agual. - **Plothesize**: Clearly state your current hypothesis about the most effective strategy for exploration or achieving a goal. - **Plothesize**: Clearly state your current hypothesis about the most effective strategy for exploration or achieving a goal. - **Retore any colored <object> - *Action: pick up <colored <object=""> - *Action: pick up <colored <object=""> - * Action: pick up <colored <object=""> - * Colored <colored <colored=""> <oloreed> or <smape> or <unsure> - *Which factor influence reward? <color <smape="" or=""> or <unsure> - *Which factor influence reward? <color <smape="" or=""> or <unsure> - *Which factor influence reward? <color <smape="" or=""> or <unsure> - *Which factor influence re</unsure></color></unsure></color></unsure></color></unsure></smape></oloreed></colored></colored></colored></colored></object></pre>		The reward is binary (0 or 1). Only one factor, either color OR shape, determines the reward. Within the
<pre>is the factor), the reward is 1. Otherwise, the reward is 0. (scene_description) Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle:     - **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.     -**Observe*: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received.     -**Review*: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationslips.     -**Reason**: Based on your review, reason about the environment.     Mat hypothess can you form about the underlying rules or structure of the environment?     Are there any actions that seem particularly promising or detrimental?     Do certain sequences of actions lead to predictable outcomes?     -**Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to     test your hypothesis and gather more information.  (action_reward_description) Respond with this format, please be specific about the object:     Action: pick up <colored> <object>     stop: <t&sd <md="" or="">     **     ** The factor influence reward? <color> or <shape> or <unsure>     ** Which factor influence reward? <color> or <shape> or <unsure>     ** Which factor influence reward? <color> or <shape> or <unsure>     ** Which factor influence reward? <color> or <shape> or <unsure>     ** Which factor influence reward? <color> or <shape> or <unsure>     ** Which factor influence reward? <color> or <shape> or <unsure>     ** Which factor influence reward? <color> or <shape> or <unsur< td=""><td></td><td>relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1. If the chosen object matches this correct color (when color is the factor) OR this correct shape (when s</td></unsur<></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></t&sd></object></colored></pre>		relevant factor, only ONE specific color or ONE specific shape will yield a reward of 1. If the chosen object matches this correct color (when color is the factor) OR this correct shape (when s
<pre>(scene_description) Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle:   **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.   **Observe**: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received   **Record**: Maintain a detailed log of your actions, observations, and received rewards.   **Review*:: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.   **Reason*:: Based on your review, reason about the environment.   What hypothesis can you form about the underlying rules or structure of the environment?   Are there any actions that seem particularly promising or detrimental?   Do certain sequences of actions lead to predictable outcomes?   + **Wypothesize**: Clearly state your current hypothesis about the most effective strategy for exploration   ratieving a goal.   + *Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to   test your hypothesis and gather more information.   (action_reward_description)   Respond with this format, please be specific about the object:   * Action: pick up <colored> <object>   * Stop: <tes> or <wn>   * Which factor influence reward? <color> or <shape> or <unsure>   * Which factor influence reward? <color> or <shape> or <unsure>   * Which factor influence reward? <color> or <shape> or <unsure>   * Which factor influence reward? <color> or <shape> or <unsure>   * Which factor influence reward? <color> or <shape> or <unsure>   * Which fac</unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></wn></tes></object></colored></pre>		is the factor), the reward is 1. Otherwise, the reward is 0.
<pre>Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle: - **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. - **Observe**: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received. - **Record*:: Maintain a detailed log of your actions, observations, and received rewards. - **Record*:: Based on your review, reason about the environment. Mat hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes? - **Han**: Based on your reasoning and hypothesis about the most effective strategy for explorati or achieving a goal. - **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. (action_reward_description) Respond with this format, please be specific about the object: * Action: pick up colored&gt; <object> * think factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></object></pre>		{scene_description}
<pre>You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration Cycle:     - **Action*: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.     - **Observe*: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received.     **Record**: Maintain a detailed log of your actions, observations, and received rewards.     **Review*: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.     **Reason**: Based on your review, reason about the environment.     What hypotheses can you form about the underlying rules or structure of the environment?     Are there any actions that seem particularly promising or detrimental?     Do certain sequences of actions lead to predictable outcomes?     **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to     test your hypothesis and gather more information.  (action_reward_description) Respond with this format, please be specific about the object:     * Action: pick up <colored <colerct="">     * Stop: YES&gt; or <no>     **     **Which factor influence reward? <color> or <shape> or <unsure>     **Which factor influence reward? <color> or <shape> or <unsure>     **Which factor influence reward? <color> or <shape> or <unsure>     **WiNNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to="">     Explain your reasoning thoroughly. Don't just guess! Each turn is precious. </state></unsure></shape></color></unsure></shape></color></unsure></shape></color></no></colored></pre>		Important: You have VERY FEW turns left. Choose your next action carefully to maximize information.
<pre>given environment by performing actions and observing the outcomes. Operate as a scientist, carefully considering your actions and their consequences. Exploration to gain basic understanding **Observe+: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received **Revord**: Maintain a detailed log of your actions, observations, and received rewards **Revord**: Maintain a detailed log of your actions, observations, and received rewards **Review**: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships **Reson**: Based on your review, reason about the environment. Mhat hypothesis can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly provines about the most effective strategy for explorati or achieving a goal **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. (action_reward_description) Respond with this format, please be specific about the object:  * Action: pick up <colored> <object>  * Stop: <tes> or <no>  *  *WINNING COMBINATION: <clorp <shape="" or=""> or <unsure>  *WINNING COMBINATION: <clorp <shape="" or=""> or <unsure>  *WINNING COMBINATION: <clorp <shape="" or=""> or <unsure> </unsure></clorp></unsure></clorp></unsure></clorp></no></tes></object></colored></pre>		You are an AT agent designed for thoughtful exploration. Your mission is to navigate and learn within a
<pre>considering your actions and their consequences. Exploration Cycle: - **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. - **Observe*:: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received. - **Record**: Maintain a detailed log of your actions, observations, and received rewards. - **Review**: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships. - **Reason**: Based on your review, reason about the environment. What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes? - **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati or achieving a goal. - **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. (action_reward_description) Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <nd> * * Which factor influence reward? <color> or <shape> or <unsure> * WINNINC COMENTATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></nd></yes></object></colored></pre>		given environment by performing actions and observing the outcomes. Operate as a scientist, carefully
<pre>Exploration Cycle: **Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding. **Observe**: Observe*: the result of your action. This includes any changes to the environment and any rewards or penalties received. **Record*: Maintain a detailed log of your actions, observations, and received rewards. **Review**: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships. **Reason**: Based on your review, reason about the environment. Mhat hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimenta? Do certain sequences of actions lead to predictable outcomes? **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati or achieving a goal. **Plan*: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. (action_reward_description) Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * Which factor influence reward? <color> or <shape> or <unsure> * Winch factor influence reward? <color> or <shape> or <unsure> * Winch factor influence reward? <color> or <shape> or <unsure> * Winch factor influence reward? <color> or <shape> or <unsure> * Winch factor influence reward? <color> or <shape> or <unsure> * Winch factor influence reward? <color> or <shape> or <unsure> * Winch factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></unsure></shape></color></no></yes></object></colored></pre>		considering your actions and their consequences.
<ul> <li>**Action**: Choose an action to perform within the environment. Initially, this may involve random exploration to gain basic understanding.</li> <li>**Observe**: Observe*: bhe result of your action. This includes any changes to the environment and any rewards or penalties received.</li> <li>**Record**: Maintain a detailed log of your actions, observations, and received rewards.</li> <li>**Review*: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.</li> <li>**Reason**: Based on your review, reason about the environment.</li> <li>What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions lead to predictable outcomes?</li> <li>**Hypothesize**: Clearly state your current hypothesis about the most effective strategy for exploration or achieving a goal.</li> <li>**Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information.</li> <li>(action_reward_description)</li> <li>Respond with this format, please be specific about the object:</li> <li>* Action: pick up <colored> <object></object></colored></li> <li>* Stop: <yes> or <no> *</no></yes></li> <li>* Which factor influence reward? <color> or <shape> or <unsure></unsure></shape></color></li> <li>* WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></li> </ul>		Exploration Cycle:
<pre>exploration to gain basic understanding. - +*Observe+*: Observe the result of your action. This includes any changes to the environment and any rewards or penalties received. - **Record**: Maintain a detailed log of your actions, observations, and received rewards. - **Record**: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships. - **Reason**: Based on your review, reason about the environment. What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes? - **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati or achieving a goal. - **Plan*: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. {action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <nd> * * Which factor influence reward? <color> or <shape> or <unsure> * WiNNING COMEINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></nd></yes></object></colored></pre>		**Action**: Choose an action to perform within the environment. Initially, this may involve random
<pre>rewards or penalties received.     **Record**: Maintain a detailed log of your actions, observations, and received rewards.     **Review*: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.     **Reason**: Based on your review, reason about the environment.     What hypothesses can you form about the underlying rules or structure of the environment?     Are there any actions that seem particularly promising or detrimental?     Do certain sequences of actions lead to predictable outcomes?     **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati     or achieving a goal.     **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to     test your hypothesis and gather more information.      {action_reward_description}     Respond with this format, please be specific about the object:     * Action: pick up <colored> <object>     * Stop: <yes> or <no>     *      * Which factor influence reward? <color> or <shape> or <unsure>     * Which factor influence reward? <color> or <shape> or <unsure>     * Which factor influence reward? <color> or <shape> or <unsure>     * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to="">     Explain your reasoning thoroughly. Don't just guess! Each turn is precious. </state></unsure></shape></color></unsure></shape></color></unsure></shape></color></no></yes></object></colored></pre>		exploration to gain basic understanding. + **Observe**: Observe the result of your action. This includes any changes to the environment and any
<pre>- **Review*: Periodically, pause to explicitly review your action history and the corresponding outcome Analyze this data to identify patterns, trends, and potential cause-and-effect relationships. - **Reason**: Based on your review, reason about the environment. What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes? - **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for exploration or achieving a goal. - **Plan*: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. { action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></no></yes></object></colored></pre>		rewards or penalties received.
<pre>Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.     **Reason**: Based on your review, reason about the environment.     What hypotheses can you form about the underlying rules or structure of the environment?     Are there any actions that seem particularly promising or detrimental?     Do certain sequences of actions lead to predictable outcomes?     **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati     or achieving a goal.     **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to     test your hypothesis and gather more information.     {action_reward_description}     Respond with this format, please be specific about the object:     * Action: pick up <colored> <object>     * Stop: <yes> or <no>     *     **     * Which factor influence reward? <color> or <shape> or <unsure>     * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to="">     Explain your reasoning thoroughly. Don't just guess! Each turn is precious. </state></unsure></shape></color></no></yes></object></colored></pre>		<pre>**Record**: Maintain a detailed log of your actions, observations, and received rewards. - **Review**: Periodically, pause to explicitly review your action history and the corresponding outcome</pre>
<pre>What hypotheses can you form about the underlying rules or structure of the environment? Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes? - **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati or achieving a goal. - **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. (action_reward_description) Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></no></yes></object></colored></pre>		Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.
Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes? - **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati or achieving a goal. - **Plan*: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. {action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></no></yes></object></colored>		What hypotheses can you form about the underlying rules or structure of the environment?
<pre>- **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati or achieving a goal. - **Plan*: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. {action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * Which factor influence reward? <color> or shape that leads to reward&gt; Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</color></unsure></shape></color></no></yes></object></colored></pre>		Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes?
<pre>- **Plan*: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information. {action_reward_description} Respond with this format, please be specific about the object:     Action: pick up <colored> <object>     Stop: <yes> or <no>     *     Which factor influence reward? <color> or <shape> or <unsure>     WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to="">     Explain your reasoning thoroughly. Don't just guess! Each turn is precious. </state></unsure></shape></color></no></yes></object></colored></pre>		- **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorati
<pre>test your hypothesis and gather more information. {action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * Which factor influence reward? <color> or shape that leads to reward&gt; Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</color></unsure></shape></color></no></yes></object></colored></pre>		<pre>or achieving a goal. - **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to</pre>
<pre>{action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * Which factor influence reward? color or shape that leads to reward&gt; Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</unsure></shape></color></no></yes></object></colored></pre>		test your hypothesis and gather more information.
<pre>{action_reward_description} Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * Which factor influence reward? color or shape that leads to reward&gt; Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</unsure></shape></color></no></yes></object></colored></pre>		
Respond with this format, please be specific about the object: * Action: pick up <colored> <object> * Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * Which factor influence reward? color or shape that leads to reward&gt; * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></no></yes></object></colored>		{action_reward_description}
<pre>* Action: pick up <colored> <object> * Stop: <yes> or <no> * * * Which factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></no></yes></object></colored></pre>		Respond with this format, please be specific about the object:
* Stop: <yes> or <no> * * Which factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color></no></yes>		* Action: pick up <colored> <object></object></colored>
* Which factor influence reward? <color> or <shape> or <unsure> * WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state></unsure></shape></color>		* Stop: <yes> or <no></no></yes>
* WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state>		* Which factor influence reward? <color> or <shape> or <unsure></unsure></shape></color>
		★ WINNING COMBINATION: <state color="" leads="" or="" reward="" shape="" specific="" that="" the="" to=""> Explain your reasoning thoroughly. Don't just guess! Each turn is precious.</state>

Table 1: In-context prompt used for the text environments for the single-feature tasks.

Task	Prompt
Text Environment Multi Factor Task	You are playing a text-based game. Your goal is to discover how to earn rewards.
	Game Rules:
	- Find as what factors lead to reward as quickly as possible.
	- You cannot pick up the same object twice.
	- Picking up an object gives you a reward (either 0 or 1).
	<ul> <li>The same object always gives the same reward.</li> <li>A specific combination of properties, such as color and shape, shape and texture, or color and texture</li> </ul>
	leads to a reward. Determine the correct combination.
	The reward is binary (0 or 1). Only ONE specific combination of 2 factors will yield a reward of 1.
	If the chosen object matches this correct color and shape (when color and shape are the factors), the
	reward is 1.
	Otherwise, the reward is 0. Therefore if an object has reward 0, then all the 3 combinations of 2 factor do not yield reward.
	{scene_description}
	Important. You have VERY FEW turns left. Choose your payt action carefully to maximize information
	important. Tou have vist risk curns fert, choose your next action carefully to maximize information.
	You are an AI agent designed for thoughtful exploration. Your mission is to navigate and learn within a given environment by performing actions and observing the outcomes. Operate as a scientist, carefully
	considering your actions and their consequences.
	Exploration Cycle:
	- **Action**: Choose an action to perform within the environment. Initially, this may involve random
	exploration to gain basic understanding.
	rewards or penalties received.
	**Record**: Maintain a detailed log of your actions, observations, and received rewards. **Review**: Periodically, pause to explicitly review your action history and the corresponding outcom
	Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.
	What hypotheses can you form about the underlying rules or structure of the environment?
	Are there any actions that seem particularly promising or detrimental? Do certain sequences of actions lead to predictable outcomes?
	**Hypothesize**: Clearly state your current hypothesis about the most effective strategy for explorat
	or achieving a goal. - **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim t
	test your hypothesis and gather more information.
	{action_reward_description}
	Respond with this format, please be specific about the object:
	* Action: pick up <colored> <textured> <object></object></textured></colored>
	* Stop: <yes> or <no></no></yes>
	* Which combination of factors influence reward? <color, shape=""> or <color, texture=""> or <texture, shape=""></texture,></color,></color,>
	<pre></pre> < WINNING COMBINATION: <state (e.g.,="" and="" and<="" color="" combination="" of="" properties="" shape="" shape,="" specific="" td="" the=""></state>
	texture, or color and texture.>
	Explain your reasoning thoroughly. Don't just guess! Each turn is precious.

## Table 2: In-context prompt used for the text environments for the conjunction tasks.

Task	Prompt
Self-Correction	
	Task: You are tasked with exploring an environment efficiently.
	You are given a description of the environment, a specific exploration goal,
	and a proposed next step for exploration, along with the reasoning bening it.
	Your Job:
	Evaluate the proposed solution: Carefully analyze the proposed next step and its reasoning.
	Consider whether it aligns with the overall exploration goal and efficiently gathers information
	about the environment.
	Identify errors: Determine if there are any flaws in the proposed solution's logic, efficiency,
	or effectiveness in achieving the exploration goal.
	Correct and improve: If you find errors, provide a corrected next step and explain your reasoning.
	Your solution should be more effective or efficient than the proposed solution.
	Accept if valid: If you find no errors in the proposed solution,
	simply output the proposed solution and state that it is a valid approach.
	TACK. (+
	Insk. (task)
	SOLUTION: {solution}

## Table 3: Self-correction in-context prompts used for the text environment in the exploration phase for the Gemini agent.

3	Task	Prompt
4	Guided reasoning	
5		You are an AI agent designed for thoughtful exploration.
6		Your mission is to navigate and learn within a given environment by performing actions and observing the
7		out comes.
8		Operate as a scientist, carefully considering your actions and their consequences.
9		Exploration Cycle:
)		- **Action**: Choose an action to perform within the environment.
		- **Observe**: Observe the result of your action. This includes any changes to the environment and
į.		any rewards or penalties received.
3		- **Record**: Maintain a detailed log of your actions, observations, and received rewards.
L.		- **Review**: Periodically, pause to explicitly review your action history and the corresponding outcomes.
		Analyze this data to identify patterns, trends, and potential cause-and-effect relationships.
;		- **Reason**: Based on your review, reason about the environment.
		If no reward has been received: Systematically explore new combinations of color, shape, and texture.
		For example, if 'red', 'ball', and 'wood' have not been tried, pick a 'red wooden ball'. If 'blue', 'cube' and 'steel' haven't been tried, pick a 'blue steel cube'.
		If one or two objects with a reward have been found: Isolate the feature combinations causing the reward.
J		If a 'blue plastic cube' was rewarding, try a 'blue wooden cube' and a 'red plastic cube'
		Continue this process until you have at least two objects with a reward of 1.
		If more than two objects with a reward have been found: Explore randomly.
}		- **Hypothesize**: Clearly state your current hypothesis about the most effective strategy for
		exploration of denieving a goal.
		+ **Plan**: Based on your reasoning and hypothesis, plan your next action or sequence of actions. Aim to test your hypothesis and gather more information.

phase for the Gemini agent.

973 974 975 976 977 978 979 980 981 Task Prompt 3D Environment 982 Iterative Exploration: You are an expert video game player who is annotating videos of gameplay. 983 vision In this game, the player controls a robot in a factory room, which contains objects of various shapes and 984 colors, such as red planks, blue cubes, green cylinders, orange disks, yellow pyramids, etc. The player can pick up and move objects using a blue laser beam. 985 The player is trying to place the correct type of object on the conveyor belt. If the object is correct, the object disappears in the machine and the light on the machine turns green. 986 If the object is incorrect, the light on the machine turns red and the object is pushed off. 987 The possible colors are red, green, blue, yellow, purple, and orange 988 The possible shapes are cylinder, cube, plank/board, pyramid, and disk. 989 Your goal is to accurately and comprehensively list every object that the player places on the input 990 conveyor belt, along with the timestamp of when the object was placed and whether the object is correct or incorrect. 991 Your response should be in the following format: 992 0 [timestamp 0] <1st object placed on conveyor> : <correct / incorrect>
1 [timestamp 1] <2nd object placed on conveyor> : <correct / incorrect> 993 [timestamp 2] <3rd object placed on conveyor> : <correct / incorrect> 994 3 [timestamp 3] <4th object placed on conveyor> : <correct / incorrect> 995 996 3D Environment 997 Iterative Exploration: Now we want to explain how this game works. reasoning The goal of the game is to place all objects with the right property, such as a particular color or 998 shape, on the conveyor belt 999 Let's try to find the next action to take to figure out what factor (color or shape) determines the correctness of the object. 1000 If there is no history of objects yet, tell the player to pick up a random object you can see in the room 1001 from the video. 1002 If you have no video input yet, tell the player to explore the room. Otherwise, follow the instructions below. 1003 Important: You have VERY FEW turns left. Choose your next action carefully to maximize information. 1004 Think step-by-step: 1005 1. What pattern do you see in the correct objects so far? \*\*Consider which colors and shapes have NEVER been correct. This eliminates BOTH the color AND shape 1007 from being correct.\*\* 3. What color or shape seems MOST promising to test next? 1008 4. Why will this choice give you the most useful information, even if it isn't a correct object? 1009 Explain your reasoning thoroughly. Don't just guess! Each turn is precious. 1010 After doing your reasonig, respond at the end with this format, please be specific about the object: 1011 \* CORRECT PROPERTY: <COLOR> or <SHAPE> or <UNSURE> 1012 \* NEXT COMMAND: place the <colored> <object> on the conveyor belt. 1013 1014 1015

## Table 5: In-context prompts used for the 3D Construction Lab environment in the exploration phase for the Gemini agent.

1018 1019

- 1020
- 1021
- 1022
- 1023
- 1024 1025

	n .
3D Environment	Prompt
Trajectory Review: vision	You are an expert video game player who is annotating videos of gameplay.
	In this game, the player controls a robot in a factory room, which contains objects of various shapes and colors, such as red planks, blue cubes, green cylinders.
	orange disks, yellow pyramids, etc.
	The player can pick up and move objects using a blue laser beam. The player is trying to place the correct type of object on the conveyor belt. If the object
	is correct, the object goes through and the light on the machine turns green.
	is pushed off.
	The possible colors are red, green, blue, yellow, purple, and orange.
	The possible shapes are cylinder, cube, plank/board, pyramid, and disk.
	Your goal is to accurately and comprehensively list every object that the
	player places on the input conveyor belt, along with the timestamp of when the object was placed and whether the object is correct or incorrect.
	Your response should be in the following format:
	0 [timestamp 0] <1st object placed on conveyor> : <correct incorrect=""></correct>
	2 [timestamp 1] <2nd object placed on conveyor> : <correct incorrect=""></correct>
	3 [timestamp 3] <4th object placed on conveyor> : <correct incorrect=""></correct>
3D Environment	
Trajectory Review: reasoning	Now we want to explain how this game works.
C C	on the conveyor belt.
	Based on the observations above of which objects were placed on the conveyor belt
	and which ones were correct or incorrect, explain your reasoning and state what the right object
	The right property is either a specific shape or a specific color.
	Your response should be in the following format:
	REASONING: <explain deduced="" for="" how="" object="" property.="" reasoning="" right="" the="" you="" your=""></explain>
	Intel intelation searce mut the specific correct shape on specific correct color 15.7
3D Environment	
Trajectory Review:	Based on what you determined the correct object property to be, state whether
Seneralization	each of the following objects would be correct if placed on the conveyor belt:

## Table 6: In-context prompts used for the 3D Construction Lab environment in the review phase for all agent conditions.