

Roadmap of Appendix

The structure of the appendix is delineated as follows: Descriptions of the relevant experimental details are provided in the Section A. Subsequently, Section B encompasses a presentation of supplementary visualization results.

A. More Implementation Details

A.1. Data Details

Our cross-modal reasoning pipeline consists of: Data Curation and Filtering, Image Formal Description, Reasoning Process Generation and Quality Assurance. Here are the details of the formal description generation process for natural image, OCR, Charts & Diagrams, and image with text.

Natural Image

1. Caption(GPT-4o): “Please describe this image in detail.”
2. Category Description: “Please identify all the objects and things present in this image and organize your answer in a sentence to list all categories.”
3. Category Description → Grounding Dino: Bounding Box

Formal Description Result: Caption + Bounding Box

OCR

1. Easy OCR: OCR Result
2. Processing OCR Results to Obtain Original Text Format(GPT-4o): “The resolution of the image is width x height. Following is the OCR result, each element in this list contains the four xy coordinates of the bounding box, text content, and the confidence score. Please restore the complete text according to the original layout without omitting any content. Maintain the relative ordering and positioning of lines as they appear in the input. Output only the restored text, without any explanations or additional information: str(ocr result)”

Formal Description Result: Original Text Format

Charts & Diagrams

1. Caption(GPT-4o): “Please translate the plot into code or scripts that clearly explain the figure. For example: SPICE for circuit schematics, PlantUML or Mermaid.js for flowcharts, LilyPond for music scores, python for sudoku, HTML for UI layouts’ every components, CSV or JSON for tables, and Matplotlib for charts with a caption, annotated text, and a note.”

Formal Description Result:
SPICE/PlantUML/Mermaid.js...

Image with Text

1. Caption(GPT-4o): “Please describe this image in detail.”
2. Category Description(GPT-4o): “Please identify all the objects and things present in this image and organize your answer in a sentence to list all categories.”
3. Category Description → Grounding Dino: Bounding Box
4. Easy OCR: OCR Result
5. Processing OCR Results to Obtain Original Text Format(GPT-4o): “The resolution of the image is width x height. Following is the OCR result, each element in this list contains the four xy coordinates of the bounding box, text content, and the confidence score. Please restore the complete text according to the original layout without omitting any content. Maintain the relative ordering and positioning of lines as they appear in the input. Output only the restored text, without any explanations or additional information: str(ocr result)”

Formal Description Result: Caption + Bounding Box + Original Text Format

We adapt the following prompt in Reasoning Process Generation.

Answer the question and provide your reasoning process, including the following:

1. Simulate image reasoning: Treat the image caption as an image. Simulate reasoning by imagining you are looking at the image, and act as if you can see it. However, avoid visualization as a step in the reasoning process.
2. Direct visual language: Frame observations as if you are directly viewing the image (e.g., “The image shows...”). Avoid reasoning through image caption or description.
3. Forbidden phrases: Avoid phrases like “based on the caption”, “based on the description”, “visualizing the image”.

Question: {question}

Image Content: {caption}.

Then, we introduce “role play” to bridge the gap in real image understanding and then filter the data. The prompts are as follows:

Revise the provided Chain of Thought (CoT) to follow these guidelines:

1. **Style Shift:** Convert all references to image description-based reasoning into direct image-based reasoning. For example: Replace phrases like “based on the description” “based on the caption” with “the image shows” or “as seen in the image”.
2. **Remove image visualization step:** If the CoT contains an inference step for image visualization, remove it and rewrite the CoT to reflect reasoning directly on the image itself, rather than reasoning after visualization from the image description. Apply these changes rigorously to ensure that the final CoT reflects direct image interpretation, uninfluenced by description, caption, image visualization.

CoT: {cot}

Give your assistant’s response. This response is the reasoning steps for the assistant to solve the problem. Please follow the following rules to evaluate whether the assistant’s response is valid. Rules for judging as valid:

1. The assistant’s response has correct reasoning steps.
2. The assistant’s response has the final reasoning answer, and the final reasoning answer is consistent with the meaning of the standard answer.
3. The assistant’s response is based on the reasoning process of the image, not the image description or caption.
4. There are no steps in the assistant’s response that are irrelevant to the reasoning, and each reasoning step is closely related.

Standard answer: {gt}

Assistant’s response: {augmented answer}

Output:

A.2. Model Details

For model training, we utilized the llama-factory and adopted a full fine-tuning strategy to optimize the model’s performance. Following this, we further refined the model through Reinforcement Learning (RL). During the inference phase, the RL-tuned model was evaluated using a fixed prompt to ensure consistent and reliable results. The prompt is defined as: **“First output the thinking process in <think> </think> tags and then output the final answer in <answer> </answer> tags.”**.

A.3. Evaluation Details

To evaluate the reasoning capabilities of the models, we adopted a unified evaluation framework for both public benchmarks and our benchmark R1-Onevision bench. Following [2, 3], we utilized GPT-4o-mini [1] to assess model performance. For extracting answers from Chain-of-Thought (CoT) responses, we employed the prompt **“Below is a thought process and an answer that includes the final choices. Please extract only the final choices (A, B, C, D, etc.) from the text and do not return any other words. If the final choice is not explicitly stated in the text, output NONE. No reasoning is required; just extract the answer.”** for multiple choice questions and **“The following is a thought process and a free-form answer. Please extract the numerical value or text of the final answer, excluding any explanation. If the final answer cannot be extracted from the given text, output NONE. No reasoning is required; just extract the answer.”** for free form questions. Then we score the extracted answer with the groundtruth using **“Compare ‘final answer’ with ‘groundtruth’. If final answer matches ‘groundtruth’, output YES; otherwise, output NO. Do not return any extra words. For numerical answers with units, if ‘final answer’ contains the unit but its numeric value matches ‘groundtruth’, consider it a match.”**

In addition to evaluating on our R1-Onevision benchmark, we also assessed the model’s reasoning capabilities on the following reasoning datasets. (1) **MathVista**: a math benchmark designed to combine challenges from diverse mathematical and visual tasks. We use the Test Mini split, around 1000 samples. (2) **MathVision**: a meticulously curated collection of 3,040 high-quality mathematical problems with visual contexts sourced from real math competitions. We use the Test Mini split, consisting of 304 samples. (3) **MathVerse**: an all-around visual math benchmark designed for an equitable and in-depth evaluation of MLLMs. We apply the full dataset and also report the ‘Vision-Only’ result which unveils the great challenge by rendering the whole question within the diagram. (4) **WeMath**: a benchmark designed to explore the problem-solving principles beyond the end-to-end performance. We adopt the Test Mini split of WeMath, around 1740 samples, we report the average accuracy as the main metric.

B. Visualization

B.1. Image Formal Description

In this section, we demonstrate some examples of image formal description.

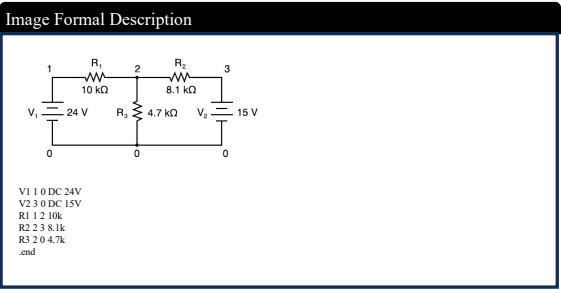


Figure 1. Formal Description of Circuits.

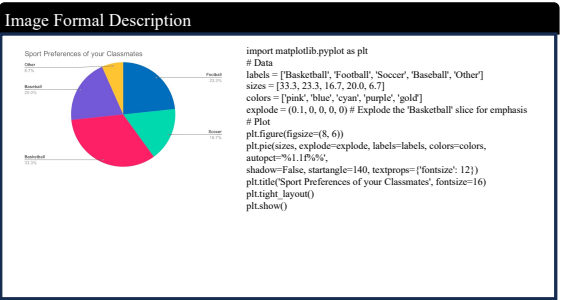


Figure 4. Formal Description of Table.

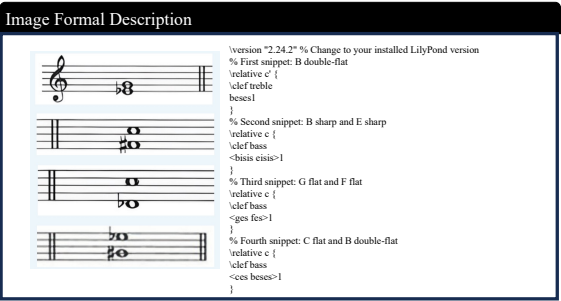


Figure 2. Formal Description of Music Sheet.



Figure 5. Formal Description of print, handwriting.

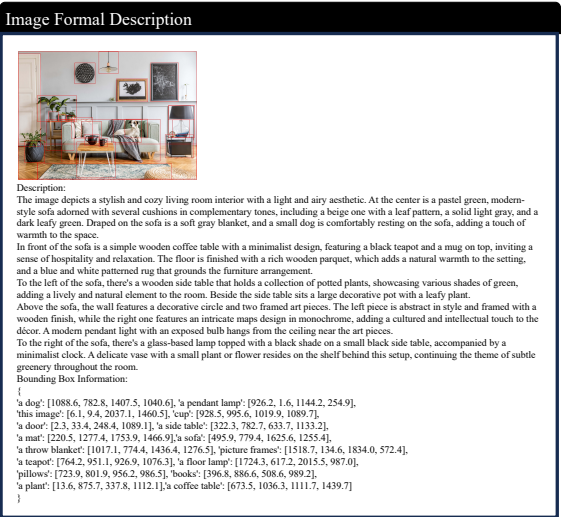


Figure 3. Formal Description of Natural Scene.

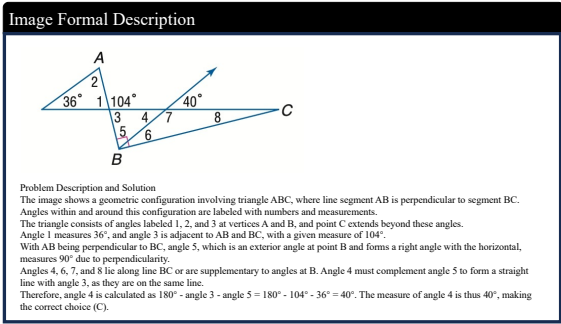


Figure 6. Formal Description of Math.

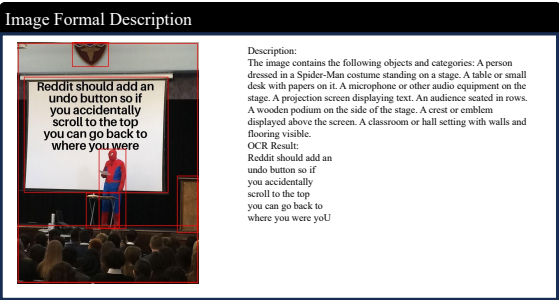


Figure 7. Formal Description of MEME.

B.2. Model Reasoning Capabilities and Qualitative Examples

In this section, we present some practical examples of our R1-Onevision.

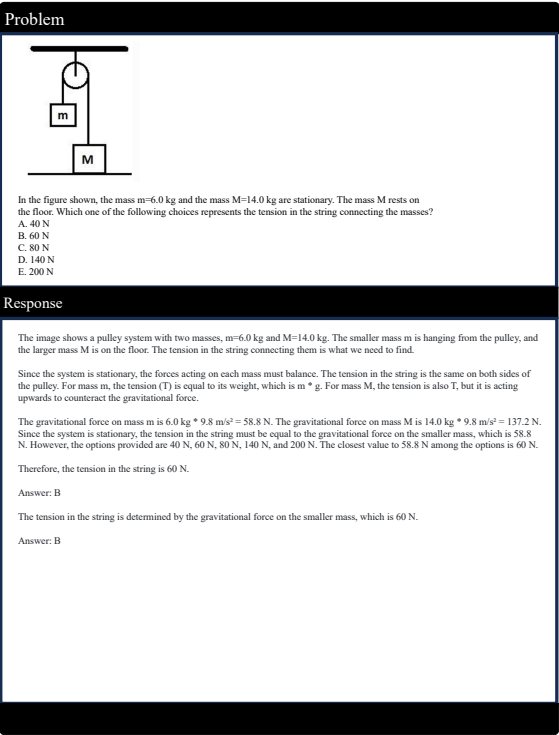


Figure 9. The model displays an adeptness in processing physics.

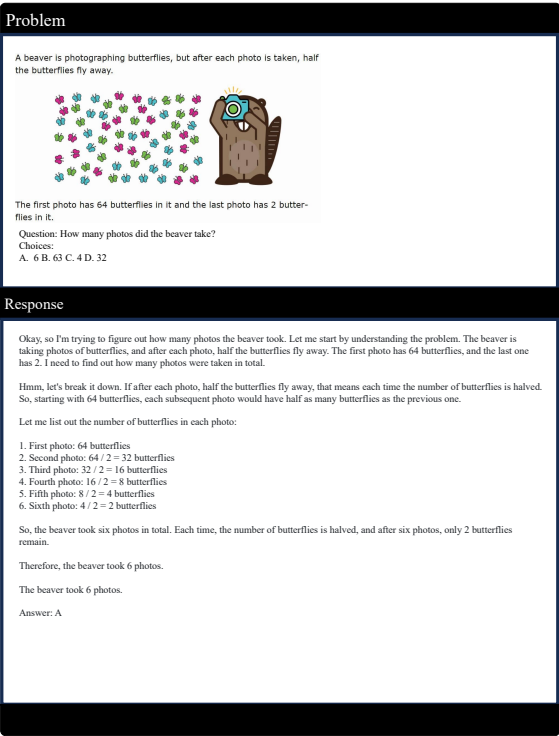
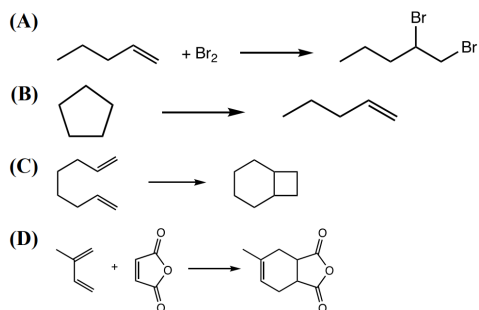


Figure 8. The model displays an adeptness in processing math.

Problem



Question: Which reaction has $\Delta S_{\text{rxn}}^{\circ} > 0$? All reactants and products are liquids.

Choices:

- A. A
- B. B
- C. C
- D. D

Response

Okay, so I need to figure out which reaction among the four options has a positive standard entropy change, denoted as $\Delta S_{\text{rxn}}^{\circ} > 0$. The hint says all reactants and products are liquids, which is important because the state of matter affects the entropy change.

Looking at each option:

Option A: The image shows a reaction where a pentene is brominated. The reactant is a liquid, and the product is a liquid. Bromination typically involves adding bromine atoms, which might form a new compound. Since both are liquids, the change in state isn't a factor here. The reaction is likely exothermic, but without knowing the exact products, it's hard to say if ΔS° is positive or negative.

Option B: The image shows a cyclopentene being converted into a hexene. Cyclopentene is a liquid, and hexene is also a liquid. The reaction involves breaking a ring and forming a double bond. This usually increases the number of degrees of freedom, leading to a higher entropy. So ΔS° should be positive here.

Option C: The image shows a reaction where a diene forms a cyclobutane. Both reactants and products are liquids. Diene to cyclobutane is a ring closure, which reduces the number of degrees of freedom. This would decrease the entropy, making ΔS° negative.

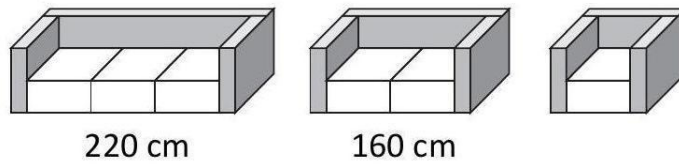
Option D: The image shows a Diels-Alder reaction between a diene and a dienophile. Both are liquids. The products are a cyclohexene and a ketone. The formation of a six-membered ring might increase the entropy, but the overall change depends on the specific reactants and products. However, since both are liquids, the state change isn't a factor.

Considering the state of matter, all reactions involve liquids, so the main factor is the type of reaction. Option B involves breaking a ring and forming a double bond, which increases the number of degrees of freedom, leading to a positive ΔS° . Options A and D might have some increase in complexity, but not necessarily a significant increase in degrees of freedom. Option C reduces degrees of freedom, making ΔS° negative.

Therefore, the correct answer is B.

Figure 10. The model displays an adeptness in processing chemistry.

Problem



Question: A furniture shop sells 3-seater, 2-seater and 1-seater sofas that each have an equally wide armrest on the left and the right hand side. Each seat is equally wide (see picture). Together with the armrests the 3-seater sofa is 220 cm wide and the 2-seater sofa 160 cm wide. How wide is the 1-seater sofa?

Choices:

- (A) 60 cm
- (B) 80 cm
- (C) 90 cm
- (D) 100 cm
- (E) 120 cm

Response

Okay, so I'm trying to figure out the width of the 1-seater sofa based on the information given. Let me start by understanding the problem.

We have three types of sofas: 3-seater, 2-seater, and 1-seater. Each has equally wide armrests on both sides, and each seat is equally wide. The total widths are given as 220 cm for the 3-seater and 160 cm for the 2-seater. We need to find the width of the 1-seater.

First, let's denote the width of each seat as (S) and the width of each armrest as (A) . Since there are two armrests per sofa, the total width of the sofa would be $(2A + 3S)$ for the 3-seater, $(2A + 2S)$ for the 2-seater, and $(2A + S)$ for the 1-seater.

From the image, we can see:

- The 3-seater sofa is 220 cm wide.
- The 2-seater sofa is 160 cm wide.

Using the image, we can set up the following equations:

1. For the 3-seater: $(2A + 3S = 220)$
2. For the 2-seater: $(2A + 2S = 160)$

We can solve these equations to find (A) and (S) .

Subtract the second equation from the first:

$$\begin{aligned} (2A + 3S) - (2A + 2S) &= 220 - 160 \\ S &= 60 \end{aligned}$$

Now, substitute $(S = 60)$ back into the second equation:

$$\begin{aligned} 2A + 2(60) &= 160 \\ 2A + 120 &= 160 \\ 2A &= 40 \\ A &= 20 \end{aligned}$$

Now, we can find the width of the 1-seater sofa:

$$2A + S = 2(20) + 60 = 40 + 60 = 100 \text{ cm}$$

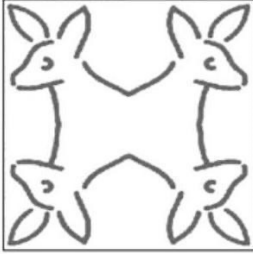
Therefore, the width of the 1-seater sofa is 100 cm.

The width of the 1-seater sofa is determined by calculating the individual seat and armrest widths using the given total widths of the 3-seater and 2-seater sofas. Solving the equations yields a seat width of 60 cm and an armrest width of 20 cm. Applying these to the 1-seater sofa gives a total width of 100 cm.

Answer: D

Figure 11. The model displays an adeptness in processing math.

Problem



Question: How many lines of symmetry does this figure have?

Response

The image shows a symmetrical design with two rabbits facing each other. Each rabbit has two lines of symmetry: one vertical and one horizontal. Since the rabbits are identical and arranged symmetrically, the entire figure has two lines of symmetry.

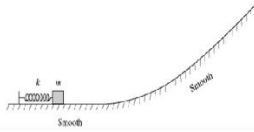
Therefore, the figure has 2 lines of symmetry.

The figure has 2 lines of symmetry.

Answer: 2

Figure 12. The model displays an adeptness in processing logic.

Problem



Question: A box of mass m is pressed against (but is not attached to) an ideal spring of force constant k and negligible mass. The spring is compressed a distance x . After it is released, the box slides up a frictionless incline as shown in the diagram at right and eventually stops. If we repeat this experiment with a box of mass $2m$

Choices:

- A. the lighter box will go twice as high up the incline as the heavier box.
- B. just as it moves free of the spring, the lighter box will be moving twice as fast as the heavier box.
- C. both boxes will have the same speed just as they move free of the spring.
- D. both boxes will reach the same maximum height on the incline.
- E. just as it moves free of the spring, the heavier box will have twice as much kinetic energy as the lighter box.

Response

Okay, so I'm trying to figure out this physics problem about a box being pushed against a spring and then released up an incline. Let me break it down step by step.

First, the setup: there's a box of mass (m) pressed against an ideal spring with force constant (k) . The spring is compressed by a distance (x) . When released, the box slides up a frictionless incline and stops. Then, the experiment is repeated with a box of mass $(2m)$. The question is asking which statement is correct regarding their behavior.

I remember that when a spring is compressed, it stores potential energy, which is converted into kinetic energy when released. The formula for the potential energy stored in a spring is $(\frac{1}{2}kx^2)$. So, when the spring is released, this energy becomes kinetic energy for the box.

Now, the box moves up an incline, which is frictionless, so all the kinetic energy turns into gravitational potential energy. The height (h) it reaches is given by $(mgh = \frac{1}{2}mv^2)$, where (v) is the initial velocity after the spring is released.

For the lighter box (m) , the initial kinetic energy is $(\frac{1}{2}kx^2)$. So, $(\frac{1}{2}kx^2 = \frac{1}{2}mv^2)$, which simplifies to $(v = \sqrt{\frac{kx^2}{m}})$. The height (h) it reaches is $(h = \frac{v^2}{2g} = \frac{\frac{kx^2}{m}}{2g} = \frac{kx^2}{2mg})$.

For the heavier box $(2m)$, the initial kinetic energy is $(\frac{1}{2}kx^2)$. So, $(\frac{1}{2}kx^2 = \frac{1}{2}(2m)v^2)$, which simplifies to $(v = \sqrt{\frac{kx^2}{2m}})$. The height (h) it reaches is $(h = \frac{v^2}{2g} = \frac{\frac{kx^2}{2m}}{2g} = \frac{kx^2}{4mg})$.

Comparing the heights, the lighter box reaches $(\frac{kx^2}{2mg})$ and the heavier box reaches $(\frac{kx^2}{4mg})$. So, the lighter box goes twice as high as the heavier box. That means option A is correct.

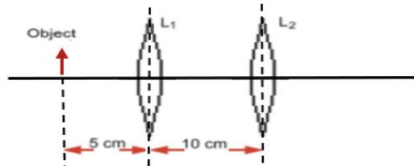
Alternatively, looking at the options:

- A. The lighter box goes twice as high as the heavier box. This matches my calculation.
- B. The lighter box moves twice as fast as the heavier box when released. The velocities are $(\sqrt{\frac{kx^2}{m}})$ and $(\sqrt{\frac{kx^2}{2m}})$, so the lighter box is moving faster, but not twice as fast.
- C. Both boxes have the same speed when released. No, their speeds are different.
- D. Both reach the same height. No, the heights are different.
- E. The heavier box has twice the kinetic energy. The kinetic energy is $(\frac{1}{2}mv^2)$, so for the lighter box, it's $(\frac{1}{2}m(\sqrt{\frac{kx^2}{m}})^2 = \frac{1}{2}kx^2)$, and for the heavier box, it's $(\frac{1}{2}(2m)(\sqrt{\frac{kx^2}{2m}})^2 = \frac{1}{2}kx^2)$. So both have the same kinetic energy, not twice as much.

So, the correct answer is A.

Figure 13. The model displays an adeptness in processing physical mechanics.

Problem



Question: A combination of two thin convex lenses is placed as shown at right. An object is placed 5 cm in front of L1 which has a focal length of 10 cm. L2 is 10 cm behind L1 and has a focal length of 12 cm. How far from L2 is the final image for this lens combination?

Choices:

- A. 8 cm
- B. 15 cm
- C. 22 cm
- D. 24 cm
- E. 30 cm

Response

Okay, so I'm trying to figure out how far the final image is from lens L2 in this setup. Let's start by analyzing the image itself. The image shows two convex lenses, L1 and L2, placed 10 cm apart. The object is 5 cm in front of L1, and L1 has a focal length of 10 cm. L2 is 10 cm behind L1 and has a focal length of 12 cm.

First, I need to determine the image distance for the first lens, L1. The object distance for L1 is 5 cm, and its focal length is 10 cm. Using the lens formula $\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$:

$$\frac{1}{10} = \frac{1}{5} + \frac{1}{d_i}$$

Solving for (d_i) :

$$\frac{1}{d_i} = \frac{1}{10} - \frac{1}{5} = \frac{1}{10} - \frac{2}{10} = -\frac{1}{10} \\ d_i = -10 \text{ cm}$$

The negative sign indicates that the image is virtual and located 10 cm behind L1.

Next, the image from L1 is 10 cm from L1, and the object for L2 is 10 cm from L1, making the object distance for L2 20 cm. Using the lens formula for L2:

$$\frac{1}{12} = \frac{1}{20} + \frac{1}{d_i}$$

Solving for (d_i) :

$$\frac{1}{d_i} = \frac{1}{12} - \frac{1}{20} = \frac{5}{60} - \frac{3}{60} = \frac{2}{60} = \frac{1}{30} \\ d_i = 30 \text{ cm}$$

The positive sign indicates that the image is real and located 30 cm from L2.

Therefore, the final image is 30 cm from L2.

Answer: E

Figure 14. The model displays an adeptness in processing physical optics.

Problem

Substance	CH ₄ (g)	CH ₂ Cl ₂ (g)	HCl(g)
ΔH_f° , kJ·mol ⁻¹	-74.6	-95.4	-92.3

Question: $\text{CH}_4(\text{g}) + \text{Cl}_2(\text{g}) \rightarrow \text{CH}_2\text{Cl}_2(\text{g}) + 2\text{HCl}(\text{g})$. Use the data in the table below to calculate the standard enthalpy, ΔH° , for the reaction above.

Choices:

- A. -205 kJ·mol⁻¹
- B. -113 kJ·mol⁻¹
- C. 113 kJ·mol⁻¹
- D. 205 kJ·mol⁻¹

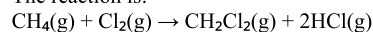
Response

Okay, so I need to calculate the standard enthalpy change, ΔH° , for the reaction $\text{CH}_4(\text{g}) + \text{Cl}_2(\text{g}) \rightarrow \text{CH}_2\text{Cl}_2(\text{g}) + 2\text{HCl}(\text{g})$. The table provides the standard enthalpies of formation (ΔH_f°) for each substance involved.

First, I'll identify the relevant ΔH_f° values from the table:

- CH₄(g) has $\Delta H_f^\circ = -74.6 \text{ kJ}\cdot\text{mol}^{-1}$
- Cl₂(g) has $\Delta H_f^\circ = 0 \text{ kJ}\cdot\text{mol}^{-1}$ (since it's a diatomic element in its standard state)
- CH₂Cl₂(g) has $\Delta H_f^\circ = -95.4 \text{ kJ}\cdot\text{mol}^{-1}$
- HCl(g) has $\Delta H_f^\circ = -92.3 \text{ kJ}\cdot\text{mol}^{-1}$

The reaction is:



To find ΔH° for this reaction, I need to calculate the difference between the sum of the products' ΔH_f° and the sum of the reactants' ΔH_f° .

So, the calculation is:

$$\Delta H^\circ = [\Delta H_f^\circ(\text{CH}_2\text{Cl}_2) + 2 \times \Delta H_f^\circ(\text{HCl})] - [\Delta H_f^\circ(\text{CH}_4) + \Delta H_f^\circ(\text{Cl}_2)]$$

Plugging in the values:

$$\Delta H^\circ = [-95.4 + 2 \times (-92.3)] - [-74.6 + 0]$$

$$\Delta H^\circ = [-95.4 - 184.6] - [-74.6]$$

$$\Delta H^\circ = -280 - (-74.6)$$

$$\Delta H^\circ = -280 + 74.6$$

$$\Delta H^\circ = -205.4 \text{ kJ}\cdot\text{mol}^{-1}$$

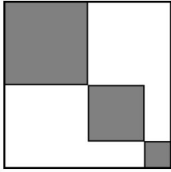
Rounding to one decimal place, $\Delta H^\circ \approx -205 \text{ kJ}\cdot\text{mol}^{-1}$.

Looking at the options, A is $-205 \text{ kJ}\cdot\text{mol}^{-1}$, which matches my calculation.

Answer: A

Figure 15. The model displays an adeptness in processing chemical reaction.

Problem



Question: In the diagram, a square with side length 6 is partially shaded. The largest shaded region is a square with side length 3. The other two shaded regions are squares with side lengths 2 and 1. What is the total area of the unshaded region?

Choices:

- A. 12
- B. 18
- C. 22
- D. 24
- E. 30

Response

The image shows a large square with a side length of 6 units, which has a total area of 36 square units. Inside this large square, there are three smaller shaded squares with side lengths of 3, 2, and 1 units. The areas of these shaded squares are 9, 4, and 1 square units respectively.

To find the total area of the unshaded region, we need to subtract the areas of the shaded squares from the total area of the large square. The total area of the shaded squares is $9 + 4 + 1 = 14$ square units. Therefore, the unshaded area is $36 - 14 = 22$ square units.

The total area of the unshaded region is 22 square units.

Answer: C

Figure 16. The model displays an adeptness in processing math geometry.

Problem

Volume of CO ₂ (number of bubbles)	Time taken for rennin to curdle milk samples (seconds)				
	Trial 1	Trial 2	Trial 3	Trial 4	Average
100	253	257	250	260	255
150	238	232	241	229	235
200	216	214	219	211	215
250	208	202	212	198	205
300	210	200	199	311	203

Question: Chymosin, known also as rennin, is a proteolytic enzyme related to pepsin that is synthesised by chief cells in the stomach of some animals. Its role in digestion is to curdle or coagulate milk in the stomach, a process of considerable importance in the very young animal. If milk were not coagulated, it would rapidly flow through the stomach and miss the opportunity for initial digestion of its proteins. Students investigated rennin activity. They bubbled different volumes of carbon dioxide (CO₂) gas into milk samples. Each sample was 50 mL and was kept at a constant temperature. The students then added rennin to each milk sample and recorded the time taken for the milk to curdle. What can the students conclude from the results of their rennin investigation?

Choices:

- A. Rennin does not curdle milk.
- B. The results are too varied to make any conclusion.
- C. The volume of carbon dioxide bubbled through the milk samples increases the time taken to curdle the milk.
- D. The volume of carbon dioxide bubbled through the milk samples decreases the time taken to curdle the milk.

Response

Okay, so I need to figure out what the students can conclude from their rennin investigation. Let me start by understanding the setup. They used different volumes of CO₂ gas in milk samples, each 50 mL, kept at a constant temperature. Then they added rennin and measured how long it took for the milk to curdle.

Looking at the data, the CO₂ volumes are 100, 150, 200, 250, and 300 bubbles. The times taken are in seconds, and the averages are given. The times decrease as the CO₂ volume increases. For example, 100 bubbles take about 255 seconds, 150 bubbles take 235, 200 bubbles take 215, 250 bubbles take 205, and 300 bubbles take 203 seconds.

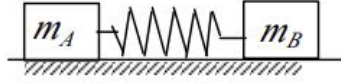
So, as the CO₂ volume increases, the time taken to curdle decreases. This suggests that more CO₂ leads to faster curdling. Since CO₂ is a gas, it might be affecting the milk's properties, maybe making it more acidic or changing its structure, which could speed up the rennin's action.

Therefore, the conclusion should be that the volume of CO₂ decreases the time taken to curdle the milk.

Answer: D

Figure 17. The model displays an adeptness in processing biology.

Problem



Question: The masses of two wooden blocks A and B are m_A and m_B respectively, with $m_B = 2m_A$. They are connected by a light spring and rest on a smooth horizontal table as shown in the figure. If an external force compresses the spring by pushing the two blocks closer and then the force is removed, the ratio of the kinetic energies of the two blocks in motion E_{KA}/E_{KB} is

Choices:

- A. $\frac{1}{2}$
- B. $\frac{\sqrt{2}}{2}$
- C. $\sqrt{2}$
- D. 2

Response

Okay, so I'm trying to solve this physics problem about two blocks connected by a spring on a smooth table. Let me break it down step by step.

First, the image shows two blocks, A and B, with masses m_A and m_B respectively. The image also shows that m_B is twice the mass of m_A , so $m_B = 2m_A$. They are connected by a light spring, which means the spring is massless and doesn't add any extra mass. The setup is on a smooth horizontal table, so there's no friction involved.

The problem states that an external force compresses the spring, bringing the two blocks closer together. Then, when the force is removed, the blocks start moving. We need to find the ratio of their kinetic energies, E_{KA}/E_{KB} .

When the external force is removed, the spring will exert forces on both blocks, causing them to move. Since the spring is light, the acceleration of each block will depend on their respective masses. According to Newton's second law, $F = ma$, so the acceleration a is F/m .

For block A, the force is F , so $a_A = F/m_A$. For block B, the force is also F , so $a_B = F/m_B$. Since $m_B = 2m_A$, $a_B = F/(2m_A)$, which is half of a_A .

Now, kinetic energy is given by $KE = \frac{1}{2}mv^2$. Since both blocks start from rest and move under the same force, their velocities will depend on their accelerations. After some time, they will have velocities v_A and v_B .

The ratio of their kinetic energies is $\frac{E_{KA}}{E_{KB}} = \frac{\frac{1}{2}m_A v_A^2}{\frac{1}{2}m_B v_B^2} = \frac{m_A v_A^2}{m_B v_B^2}$.

Since $a_A = \frac{F}{m_A}$ and $a_B = \frac{F}{m_B}$, the ratio of their accelerations is $\frac{a_A}{a_B} = \frac{m_B}{m_A} = 2$. This means $v_A = 2v_B$.

Substituting $v_A = 2v_B$ into the kinetic energy ratio, we get:

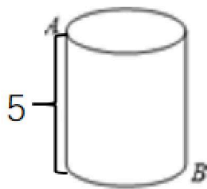
$$\frac{E_{KA}}{E_{KB}} = \frac{m_A (2v_B)^2}{m_B v_B^2} = \frac{m_A \cdot 4v_B^2}{2m_A v_B^2} = \frac{4}{2} = 2$$

Therefore, the ratio of the kinetic energies is 2.

Answer: D

Figure 18. The model displays an adeptness in processing physical mechanics.

Problem



Question: As shown in the figure, a cylinder with a bottom circumference of 24.0, the shortest route that an ant passes along the surface from point A to point B is ()

Choices:

A: 12m

B: 15m

C: 13m

D: 14m

Response

Okay, so I'm trying to solve this problem about the shortest path an ant takes on a cylinder. The image shows a cylinder with a bottom circumference of 24.0 meters. Points A and B are on the surface, and the ant needs to find the shortest path from A to B.

First, I remember that when dealing with shortest paths on curved surfaces, sometimes it's helpful to unfold the cylinder into a flat shape. As seen in the image, if I imagine cutting the cylinder along a vertical line and laying it flat, the bottom becomes a rectangle. The height of this rectangle would be the height of the cylinder, which is 5 meters, and the width would be half the circumference because the ant can travel around the cylinder.

So, the rectangle would have a height of 5 meters and a width of 12 meters (since 24 divided by 2 is 12). Now, the shortest path on the cylinder should correspond to the straight line on this rectangle. To find the length of this straight line, I can use the Pythagorean theorem.

The theorem states that in a right-angled triangle, the square of the hypotenuse (which would be the shortest path here) is equal to the sum of the squares of the other two sides. So, if I let the shortest path be 'd', then:

$$d^2 = 5^2 + 12^2$$

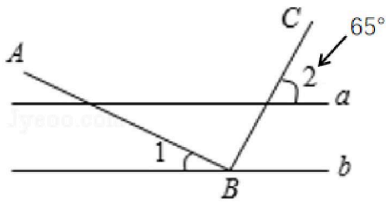
Calculating that, 5 squared is 25, and 12 squared is 144. Adding those together gives 169. Taking the square root of 169 gives me 13. So, the shortest path should be 13 meters.

Looking at the options provided, option C is 13 meters, so that must be the correct answer.

Answer: C

Figure 19. The model displays an adeptness in processing math.

Problem



Question: As shown in the figure, the straight line a parallel b , and AB perpendicular BC , then the degree of angle 1 is ()

Choices:

A: 65°

B: 25°

C: 35°

D: 45°

Response

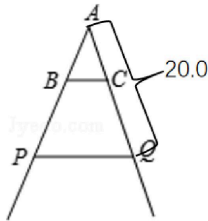
The image shows two parallel lines, a and b , intersected by a transversal. AB is perpendicular to BC , forming a right angle at B . Angle 2 is 65° , and angle 1 is adjacent to it. Since a and b are parallel, angle 2 and its corresponding angle on line b are equal. Thus, the angle adjacent to angle 1 is 65° . As AB is perpendicular to BC , angle 1 is $90^\circ - 65^\circ = 25^\circ$.

Therefore, the measure of angle 1 is 25° , which corresponds to option B.

Answer: B

Figure 20. The model displays an adeptness in processing math.

Problem



Question: It is known that BC parallel PQ, $AB:AP = 2.0:5.0$, then the length of CQ is ()

Choices:

A: 8cm

B: 12cm

C: 30cm

D: 50cm

Response

Okay, so I'm trying to solve this geometry problem where BC is parallel to PQ, and the ratio of AB to AP is 2:5. I need to find the length of CQ. The image shows a triangle with points A, B, C, P, and Q. BC is parallel to PQ, which means that triangles ABC and APQ are similar because of the parallel lines.

Since BC is parallel to PQ, the triangles ABC and APQ are similar by the Basic Proportionality Theorem (Thales' theorem).

This theorem states that if a line is drawn parallel to one side of a triangle, it divides the other two sides proportionally.

The ratio of AB to AP is given as 2:5. This means that AB is 2 parts and AP is 5 parts. Since the triangles are similar, the ratio of their corresponding sides should be the same. Therefore, the ratio of BC to PQ should also be 2:5.

The image shows that PQ is 20 cm. Using the ratio, we can set up the proportion:

$$\frac{BC}{PQ} = \frac{2}{5}$$

Substituting $PQ = 20$ cm into the proportion:

$$\frac{BC}{20} = \frac{2}{5}$$

Solving for BC:

$$BC = \frac{2}{5} \times 20 = 8 \text{ cm}$$

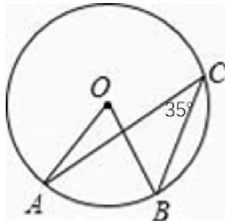
Since BC is parallel to PQ, the length of CQ is the remaining part of PQ after subtracting BC. Therefore, $CQ = PQ - BC = 20 - 8 = 12$ cm.

So, the length of CQ is 12 cm, which corresponds to option B.

Answer: B

Figure 21. The model displays an adeptness in processing math.

Problem



Question: As shown in the figure, O is the center of the circle, then the degree of angle AOB is ()

Choices:

A: 35°

B: 70°

C: 105°

D: 150°

Response

Okay, so I'm trying to solve this geometry problem where I need to find the measure of angle AOB in a circle. The image shows a circle with center O, and points A, B, and C on the circumference. The angle at point C, which is angle ACB, is given as 35 degrees. The question is asking for the measure of angle AOB.

Hmm, I remember something about angles in circles. There's a theorem that relates central angles to inscribed angles. Let me think... Oh right! The Inscribed Angle Theorem states that an inscribed angle is half the measure of the central angle that subtends the same arc. So, if angle ACB is an inscribed angle, then the central angle AOB should be twice that.

So, if angle ACB is 35 degrees, then angle AOB should be 2 times 35 degrees. Let me calculate that: 2 multiplied by 35 is 70 degrees. That makes sense because the central angle is always larger than the inscribed angle for the same arc.

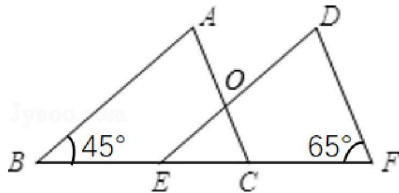
Looking at the choices provided: A is 35° , B is 70° , C is 105° , and D is 150° . So, 70° is option B. That must be the correct answer.

I don't think I made any mistakes here. The key was remembering the Inscribed Angle Theorem and applying it correctly. So, yeah, angle AOB is 70 degrees.

Answer: B

Figure 22. The model displays an adeptness in processing math geometry.

Problem



Question: As shown in the figure, triangle ABC congruent triangle DEF, then the degree of angle COE is ()

Choices:

A: 40°

B: 60°

C: 70°

D: 100°

Response

Okay, so I'm trying to solve this geometry problem where triangle ABC is congruent to triangle DEF. The question is asking for the measure of angle COE. The image shows that angle B is 45 degrees and angle F is 65 degrees. Since the triangles are congruent, their corresponding angles are equal. Therefore, angle A in triangle ABC corresponds to angle D in triangle DEF, angle B corresponds to angle E, and angle C corresponds to angle F.

Given that angle B is 45 degrees, angle E must also be 45 degrees. Similarly, since angle F is 65 degrees, angle C must also be 65 degrees. Now, angle COE is the angle at point O between points C and E. As seen in the image, angle COE is formed by the intersection of lines CO and EO, which are the extensions of sides AC and DF respectively.

Since angle C is 65 degrees and angle E is 45 degrees, and they are adjacent angles forming a linear pair, their sum is 180 degrees. Therefore, angle COE, which is the remaining angle, is $180 - 65 - 45 = 70$ degrees.

Thus, the measure of angle COE is 70 degrees.

Answer: C

Figure 23. The model displays an adeptness in processing math geometry.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. [2](#)
- [2] Pan Lu, Hritik Bansal, Tony Xia, Jiacheng Liu, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, Kai-Wei Chang, Michel Galley, and Jianfeng Gao. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. In *The Twelfth International Conference on Learning Representations*, 2024. [2](#)
- [3] Ke Wang, Junting Pan, Weikang Shi, Zimu Lu, Houxing Ren, Aojun Zhou, Mingjie Zhan, and Hongsheng Li. Measuring multimodal mathematical reasoning with math-vision dataset. *Advances in Neural Information Processing Systems*, 37:95095–95169, 2025. [2](#)