

Level 3 - Independent Count-Object-Color

Question: You are presented with a visual reasoning puzzle consisting of a 3×3 grid of images, where one image is missing in one of the nine positions.

Below the grid are four candidate options labeled A, B, C, and D, arranged horizontally from left to right. Exactly one option correctly completes the grid.

Task: 1. The grid follows a row-wise logical rule, where multiple visual attributes change systematically across each row; 2. The same rule or transformation pattern applies consistently across all rows in the grid; 3. In this puzzle, focus on the following attributes: color, number of elements (count), and object type, which may vary independently or in relation to one another; 4. Carefully analyze how these attributes vary across each row, identify the common rule of change, and determine how it applies to the row with the missing image; 5. Choose the option that best satisfies the row rule and completes the grid in a logically consistent way.

In your reply, follow this format exactly:

Reasoning: Describe the main rule or pattern and explain why your answer fits it.

Answer: [A, B, C, or D]



Ground Truth: B ✓

GPT-5-Thinking

Reasoning: Across each row, two things change in a consistent way: the images move from a lighter color on the left toward a darker color on the right, and the number of items tends to decrease from left to right.

Row 1 goes from white (stacked trays) → red/white (train set) → red (single caboose).

Row 2 goes from lighter metal/wood (three hammers) → metal/wood (single hammer) → dark black (two vases).

In Row 3 we already have a light image on the left (two white planes) and a dark image on the right (black toaster). The missing center image should therefore be mid-tone/red and contain multiple (but not many) elements—between the left's two items and the right's single item. Option C (the red two-bin container) fits: it's red (mid-tone) and shows a small multiple, matching the row's progression in both color and count.

Model Answer: C ✗

Compositional rule failure: the model misses the true row-wise rules for count (subtraction), color and object identity (two-similar, one-different), and instead relies on vague heuristics (“lighter-to-darker,” “decreasing count”), failing to integrate the attributes.