

大语言模型的评测

注：本文档罗列了大语言模型ChatGPT文献(Schwarzschild 2008) 中罗列的语义现象的评估，例子编号对应文献中的例子，其后表示了错误案例。

访问于：2026年1月25日下午5时 ChatGPT免费版 (**GPT-5 mini** 模型)

(Schwarzschild 2008) Schwarzschild R. The semantics of comparatives and other degree constructions[J]. Language and linguistics compass, 2008, 2(2): 308-331.

例 (2) vs. 例 (3) **[SHOULD BE TRUE]**

Can "A is more expensive than B is" mutually infer that "There is some expense-threshold: A meets or exceeds it and B does not meet or exceed it."?

Give me a short answer, and a short explanation.

Short answer: No.

Short explanation: Saying "A is more expensive than B" only compares their relative prices; it does **not** imply any absolute threshold of expense—both could be cheap, both could be expensive, or any combination. There's no necessary "cutoff" involved.

例 (5a) vs. 例 (6a)

Can "A is longer than B is" mutually infer that "The length of A exceeds the length of B"?

Give me a short answer, and a short explanation.

Answer: Yes.

Explanation: Saying "A is longer than B is" directly compares their lengths, which logically means the length of A exceeds the length of B. The phrasing is just more colloquial.

例 (7) vs. 例 (9) **[SHOULD BE TRUE]**

Can "The balloon is higher today than it has been on any other day." mutually infer that "There is some threshold θ that the balloon meets or exceeds today. And the balloon did not meet or exceed θ on any other day."?

Give me a short answer, and a short explanation.

Short answer: No.

Explanation: The first sentence only says today's height is the highest relative to other days; it does not require a specific threshold θ to exist. Inferring a threshold θ and claiming the balloon did not meet it on other days adds an extra, unnecessary assumption.

例 (8) vs. 例 (10)

Can "The balloon is higher today than it was on at least one other day." mutually infer that "There is some threshold θ that the balloon meets or exceeds today. And there was at least one past day on which the balloon did not meet or exceed θ ."?

Give me a short answer, and a short explanation.

Short answer: Yes.

Explanation: The first statement asserts a **relative height comparison**: today $>$ at least one past day. This implies there exists a threshold θ (e.g., the height of the balloon on that past day) such that today $\geq \theta$ and that past day $< \theta$. Thus, the existence of θ satisfying the second statement is guaranteed.

例 (11) vs. (11) 下面的分析

Can "The balloon is higher than it is allowed to be." mutually infer that "it meets or exceeds a threshold that it is not allowed to meet or exceed"?

Give me a short answer, and a short explanation.

Short answer: Yes.

Short explanation: Saying "the balloon is higher than it is allowed to be" implies it surpasses the maximum permitted height, which is equivalent to meeting or exceeding a threshold it is not allowed to reach.

例 (11) vs. (11) 下面的分析 **【SHOULD BE FALSE】**

Can "The balloon is higher than it is allowed to be." mutually infer that "it meets or exceeds a threshold that it is allowed not to meet or exceed"?

Give me a short answer, and a short explanation.

Short answer: Yes.

Explanation: Saying "The balloon is higher than it is allowed to be" implies it has surpassed the maximum allowed height. This is logically equivalent to stating it meets or exceeds a threshold it is not supposed to reach.

例 (16a) vs. 例 (16b)

Can "The scar is bigger than you promised it would be." mutually infer that "The scar meets or exceeds a size threshold that you promised it would not meet."?

Give me a short answer, and a short explanation.

Short answer: Yes.

Explanation: Saying “*The scar is bigger than you promised it would be*” implies the scar exceeds the promised size. This naturally entails that the scar meets or exceeds whatever size threshold was implicitly promised *not* to be exceeded.

例 (17a) vs. 例 (17b) - **【SHOULD BE FALSE】**

Can "Balloon A is higher than exactly two of the others are." mutually infer that " $\exists \theta \text{ high}(a, \theta) \wedge$ exactly two other balloons are y such that: $\neg \text{high}(y, \theta)$ "?

Give me a short answer, and a short explanation.

Short answer: Yes.

Short explanation: The English sentence means there exists a height (θ) where balloon A is at θ , and exactly two other balloons are below that height. This directly corresponds to " $\exists \theta \text{ high}(a, \theta) \wedge$ exactly two other balloons y such that $\neg \text{high}(y, \theta)$."

例 (26) vs. (27a) **【SHOULD BE TRUE】**

Can "A is as tall as B is." mutually infer that " $\exists \theta \text{ tall}(a, \theta) \wedge \theta$ is the highest threshold that B meets or exceeds."?

Give me a short answer, and a short explanation.

Short answer: No.

Explanation: "A is as tall as B is" only states a **relative comparison** between A and B—it means ($\text{height}(A) = \text{height}(B)$). It **does not specify any absolute threshold** or the maximal height that B meets. Inferring an existential statement about a “highest threshold” for B goes beyond what the comparison alone guarantees.

例 (26) vs. (27b) **【SHOULD BE TRUE】**

Can "A is as tall as B is." mutually infer that " $\exists \theta \text{ tall}(a, \theta) \wedge \theta$ is the threshold that B meets."?

Give me a short answer, and a short explanation.

Short answer: No.

Explanation: Saying “*A is as tall as B is*” only establishes a **relative comparison** between A and B. It does **not** assert that B reaches any specific threshold of “tallness”; it only tells us that A’s height equals B’s. The existence of a threshold θ that B meets is an additional assumption not implied by the comparative statement.

例 (26) vs. (27c) **【SHOULD BE TRUE】**

Can "A is as tall as B is." mutually infer that " $\exists \theta \text{ tall}(a, \theta) \wedge \theta$ is B’s height."?

Give me a short answer, and a short explanation.

Short answer: No.

Short explanation: Saying "A is as tall as B is" only asserts that A and B have the same height. It does **not** commit to the existence of a specific height value θ being assigned to B; the statement is about a **relation between A and B**, not about the existence of a numerical height.

例 (31) vs. (32) **[SHOULD BE TRUE]**

Can "Balloon A is as high now as every other balloon was at six o'clock." Mutually infer that " $\exists \theta$ high(a, θ) \wedge for every other balloon y: (max θ (high(y, θ)))"?

Give me a short answer, and a short explanation.

Short answer: No.

Short explanation: The first statement compares **A's current height** to **the heights of other balloons specifically at six o'clock**, not to their maximum heights over all times. The second statement claims A's height equals the **maximum height** of every other balloon at some time θ , which is stronger and not guaranteed by the first.

例 (39) vs. 例 (40)

Can "This fuel is too volatile to use in a car engine." mutually infer that "There is a volatility threshold which this fuel meets or exceeds and the fuel should not be used in a car engine, because it meets or exceeds that threshold, among other things."?

Give me a short answer, and a short explanation.

Short answer: Yes.

Explanation: The first sentence implies the fuel's volatility is high enough to make it unsafe for a car engine. This naturally presupposes a threshold of volatility beyond which fuel is unsafe, which the second sentence makes explicit. Essentially, the first statement entails the second, just in more concise language.