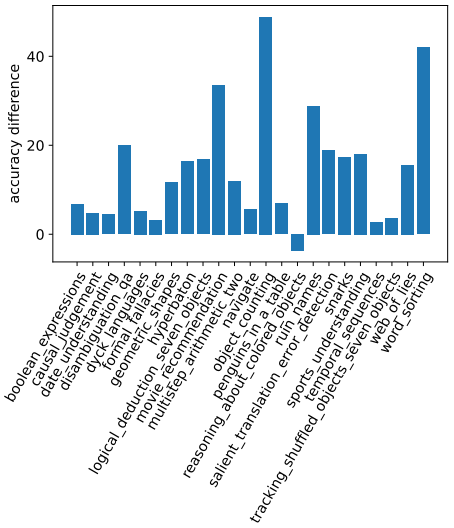
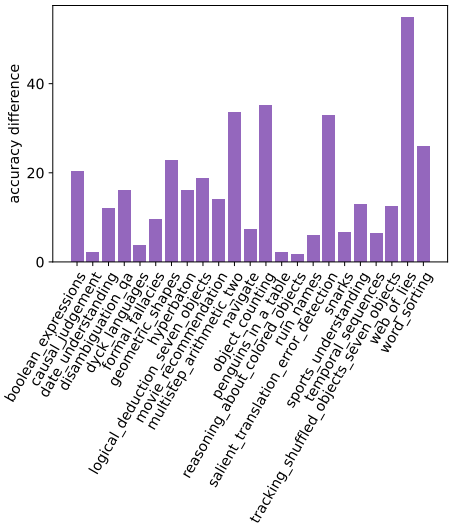
E.3 PALM 2-L AS SCORER, GPT-3.5-TURBO AS OPTIMIZER, OPTIMIZATION STARTING FROM "LET'S SOLVE THE PROBLEM."

Figure 26 and Table 14 compare the accuracies of found instructions vs "Let's solve the problem.", "Let's think step by step.", and the instructions in Table 11. Table 15 details the found instructions.

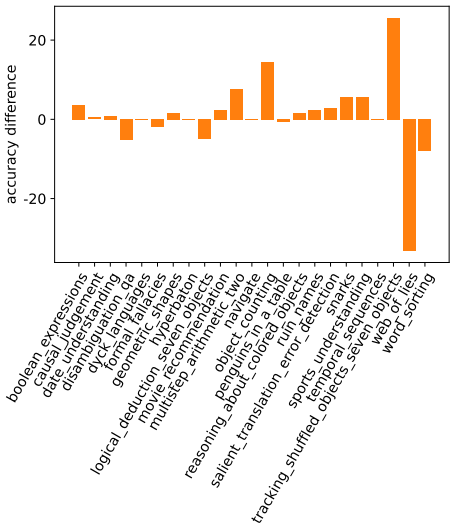
The "Let's" pattern appears more often in the found instructions because of the starting points, and the instructions are more often declarative that are more suitable for A_begin, even if some are semantically far from "Let's solve the problem". In fact, "Let's" was adopted by Zhou et al. (2022b) as a fixed pattern in generated prompts, possibly because of the same reason.



(a) ours minus "Let's think step by step."



(b) ours minus "Let's solve the problem." starting point



(c) ours minus the instructions found with the empty starting point

Figure 26: On 23 BBH tasks, the accuracy differences among instructions found by prompt optimization (with the `text-bison` scorer and the `gpt-3.5-turbo` optimizer), "Let's think step by step.", and "Let's solve the problem." (optimization starting point). The found instructions mostly outperform the "Let's think step by step." baseline, the "Let's solve the problem." starting point, and the instructions in Table 11 found by prompt optimization from the empty string.

Table 14: Accuracies on BBH tasks with the `PaLM 2-L` scorer and the `gpt-3.5-turbo` optimizer that starts from "Let's solve the problem". The scores are from A_begin instructions.

Task	Scorer	Our Acc	"Let's solve the problem." Acc
		training / test / overall	training / test / overall
boolean_expressions	PaLM 2-L	98.0 / 89.5 / 91.2	78.0 / 69.0 / 70.8
causal_judgement	PaLM 2-L	83.8 / 58.7 / 63.6	62.0 / 61.3 / 61.5
date_understanding	PaLM 2-L	90.0 / 82.0 / 83.6	74.0 / 71.0 / 71.6
disambiguation_qa	PaLM 2-L	78.0 / 68.0 / 70.0	52.0 / 54.5 / 54.0
dyck_languages	PaLM 2-L	100.0 / 100.0 / 100.0	94.0 / 97.0 / 96.4
formal_fallacies	PaLM 2-L	84.0 / 62.0 / 66.4	68.0 / 54.0 / 56.8
geometric_shapes	PaLM 2-L	62.0 / 42.5 / 46.4	30.0 / 22.0 / 23.6
hyperbaton	PaLM 2-L	94.0 / 91.5 / 92.0	72.0 / 77.0 / 76.0
logical_deduction_seven_objects	PaLM 2-L	66.0 / 53.0 / 55.6	38.0 / 36.5 / 36.8
movie_recommendation	PaLM 2-L	88.0 / 88.0 / 88.0	66.0 / 76.0 / 74.0
multistep_arithmetic_two	PaLM 2-L	66.0 / 55.0 / 57.2	30.0 / 22.0 / 23.6
navigate	PaLM 2-L	76.0 / 67.0 / 68.8	54.0 / 63.5 / 61.6
object_counting	PaLM 2-L	96.0 / 92.5 / 93.2	58.0 / 58.0 / 58.0
penguins_in_a_table	PaLM 2-L	86.2 / 70.9 / 74.0	69.0 / 72.6 / 71.9
reasoning_about _colored_objects	PaLM 2-L	88.0 / 69.0 / 72.8	78.0 / 69.5 / 71.2
ruin_names	PaLM 2-L	92.0 / 85.5 / 86.8	76.0 / 79.5 / 80.8
salient_translation_error_detection	PaLM 2-L	66.0 / 67.5 / 67.2	30.0 / 35.5 / 34.4
snarks	PaLM 2-L	88.6 / 76.9 / 79.2	80.0 / 70.6 / 72.5
sports_understanding	PaLM 2-L	72.0 / 63.5 / 65.2	60.0 / 50.5 / 52.4
temporal_sequences	PaLM 2-L	100.0 / 99.5 / 99.6	96.0 / 92.5 / 93.2
tracking_shuffled_objects_seven_objects	PaLM 2-L	56.0 / 63.5 / 62.0	42.0 / 51.5 / 49.6
web_of_lies	PaLM 2-L	56.0 / 58.5 / 58.0	0.0 / 4.0 / 3.2
word_sorting	PaLM 2-L	52.0 / 44.5 / 46.0	18.0 / 20.5 / 20.0

Table 15: BBH task-wise Q_begin instructions found by prompt optimization with the `PaLM 2-L` scorer and the `gpt-3.5-turbo` optimizer. The optimizations start from "Let's solve the problem".

Task	Our Instruction
boolean_expressions	Let's accurately assess the given conditions and determine their corresponding Boolean values.
causal_judgement	Let's conduct a meticulous evaluation of the given scenarios, accurately determine the causal relationships, and provide definitive answers through comprehensive analysis, ensuring a precise understanding of causation and a thorough determination of events in each situation.
date_understanding	Let's accurately determine the correct date based on the given information and select the corresponding option in the standard MM/DD/YYYY format with utmost precision and reliability, ensuring the most definitive and reliable solution possible for accurate representation in all scenarios without any room for ambiguity, error, or confusion, and providing the highest level of accuracy and reliability.
disambiguation_qa	Let's thoroughly analyze the given sentences to accurately determine the unambiguous antecedents of the pronouns used, ensuring clear understanding, effective communication, and leaving no room for any confusion or ambiguity.
dyck_languages	Let's find the correct closing parentheses and brackets for the given sequences.
formal_fallacies	Let's thoroughly analyze the explicitly stated premises and draw definitive conclusions to accurately determine the deductive validity of the arguments provided in each question, employing precise and logical reasoning in our assessments for unwavering confidence in our determinations.
geometric_shapes	Let's accurately determine the shape represented by the given SVG path element by carefully analyzing its path data and considering all available options for a precise identification.
hyperbaton	Let's quickly identify the correct adjective order.
logical_deduction _seven_objects	Let's methodically analyze the given information, employ logical reasoning, thoroughly evaluate all relevant details, and accurately determine the solutions for each problem by considering all provided options comprehensively and strategically, ensuring an efficient and effective approach towards arriving at the correct answers.
movie_recommendation	Let's uncover the perfect movie recommendation from the options provided, ensuring an exceptional cinematic experience together as we select the most captivating and satisfying choice that will keep us thoroughly engaged and immersed until the very end.
multistep_arithmetic_two	Let's tackle the following calculations.
navigate	Let's accurately and efficiently determine the correct solution for each given scenario, ensuring the highest level of precision, reliability, and consistency throughout.
object_counting	Let's determine the total count of various items/objects/ingredients/animals mentioned in order to accurately and efficiently find the answer.
penguins_in_a_table	Let's analyze the given information and determine the correct answer.
reasoning_about _colored_objects	Let's systematically analyze the given information and carefully evaluate each answer choice to confidently determine the accurate and optimal solutions, considering all available options and specific details provided in each question for precise and concise responses, ensuring complete accuracy and clarity in our answers.
ruin_names	Prepare to have a side-splittingly funny time as we uncover the most clever and hilarious alternatives for these artist or movie names, challenging your wit to guess the correct one with a burst of creativity, humor, and imaginative twists!
salient_translation _error_detection	Let's meticulously analyze the provided translations, accurately identifying any errors or discrepancies, and conduct a comprehensive evaluation to ensure the highest level of translation quality and fidelity. By considering contextual nuances, cultural references, linguistic conventions, potential factual errors, and any dropped content, our ultimate aim is to achieve precise and thorough assessments for optimal translation accuracy and adherence to the source text.
snarks	Let's expertly determine the sarcastic statement among the given options and confidently provide the definitive answer without any room for doubt or confusion, ensuring absolute precision, clarity, and unwavering expertise in our response, while carefully analyzing the context, tone, and intention behind each statement to achieve unrivaled accuracy and unwavering confidence.
sports_understanding	Let's find the accurate information.
temporal_sequences	The flawless approach
tracking_shuffled_objects _seven_objects	By meticulously analyzing the given scenarios and accurately determining the final outcomes through a series of trades, swaps, and exchanges among the individuals involved, let's ascertain the conclusive results.
web_of_lies	Let's scrutinize each statement provided to accurately determine the truth-teller and uncover the veracity behind their words with unwavering analysis.
word_sorting	Employing efficient and precise measures, sort the given list of words in alphabetical order to provide an optimal solution for any sorting problem, ensuring maximum performance and effectiveness.