No LLM Aced Yu Tsumura's 554th Problem

Simon Frieder [1] **William Hart** [2] **Leonhard Hasler** [3]

Abstract

We present a case study about a group theory problem—Yu Tsumura's 554th problem—that has surprisingly diagnostic capabilities for LLMs. On the one hand, this shows, contrary to the optimism about LLMs' problem-solving abilities, fueled by the recent gold medals at the International Math Olympiad (IMO) that LLMs attained, that a group theory problem exists—Yu Tsumura's 554th problem—that a) is within the scope of an IMO problem in terms of proof sophistication, b) is not a combinatorics problem, which is known to cause issues for LLMs, c) has a publicly available solution (likely in the training data of LLMs), and d) cannot be stably solved by *any* existing off-the-shelf LLM (commercial or open-source). This problem is the end result of a discovery process involving 300+ group theory problems, few of which consistently produced comparable failures across a selection of LLMs, and none of which was at the same level of mathematical sophistication. We include an analysis of the output traces of 31 LLMs in the years 2025-2026, where our focus is on the top models and their lineages, tracking intermediate versions. Our analysis shows that LLMs reasoning performance across problems is spiky even for comparatively simple problems and implies that there may be fundamental issues with model training approaches. Additionally, we compare, in a case study, the generic LLM output to a new proof of Yu Tsumura's problem by a former IMO participant, to contrast human reasoning ability with LLM reasoning ability. The focus on Yu Tsumura's problem further reveals technical issues in the original publicly available proof, and we present a substantially better-motivated proof.

[1]University of Oxford, UK [2]University of Cambridge, UK [3]ETH Zürich, CH. Correspondence to: Simon Frieder <simon.frieder@cs.ox.ac.uk>, William Hart <wh369@cam.ac.uk>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

1. Introduction

While recent work tests the boundaries of LLMs' mathematical proof generation abilities on competition-level or research-level questions (Dekoninck et al., 2025; Glazer et al., 2024; Schmitt et al., 2025), studies evaluating the *robustness* of mathematical reasoning still often rely on answer-based benchmarks such as GSM8K (Cobbe et al., 2021) or MATH (Hendrycks et al., 2021), or FrontierMath (Glazer et al., 2024).

These robustness studies probe functional variations of proofs (Mirzadeh et al., 2024; Hao et al., 2025), perturbations such as typographical or numerical changes (Gan et al., 2024; Yang et al., 2025), incomplete-information settings (Yu et al., 2025), or by asking follow-up questions (Hou et al., 2025). These frameworks almost exclusively evaluate performance on benchmarks on the level of GSM8K and MATH, and exclusively rely on autograding.

Our paper is, to our knowledge, the first to study this form of robustness for a mathematical problem that requires deeper reasoning depths, using human assessment instead of automated assessments. Our approach also differs from prior work in several further respects:

- We focus on a single benchmark problem rather than a large benchmark dataset. Instead of evaluating many problems on only a few LLMs, we trace performance on one problem across several intermediate off-the-shelf LLM versions, both open and closed, which makes it possible to study the easy-hard phenomenon in finer detail. Our human evaluations cover 91 expertly rated proofs across 31 LLMs. Unlike other works, we pay particular focus on the strongest models, and their lineages, rather then a wide sample of LLMs, in order to assess the current frontier of LLMs reasoning abilities.

- Results such as from (Glazer et al., 2024) show that there exist LLMs that solve math problems that require intricate reasoning abilities, for which likely no solution was in the training data ahead of time, due to the recency of the problems. We show the converse: There exist mathematical problems that require intricate reasoning abilities that no current off-the-shelf LLMs have, for which a full solution exists online,

posted at a time that predates the advent of LLMs.

- We conduct a human case study to better understand how people reason about the same problem, and whether the problem-solving skills transfer in a similar way to LLMs' problem-solving skills. Yu Tsumura's problem, even if dressed up as group theory, does not use any group-theoretic concepts, and is solvable using only high-school-level mathematics.[1] This makes it possible to test whether reasoning is transitive: A human that solves an IMO-level problem should be able to solve this problem. LLMs are known to be able to solve IMO-level problems,[2] yet we find consistent evidence that they fail to solve this problem. We note that while this shows that the level of difficulty matches what a human that performs well at the IMO is able to solve, the proof of Yu Tsumura's problem, due to its heavy symbolic manipulations, is different from typical proofs of IMO-type math problems.

- Our benchmarking effort also produced a modest but revealing mathematical byproduct. Comparing LLM-generated reasoning with the existing solution of an elementary exercise, led to re-examining the original argument and identifying errors in the proof. Prior benchmarking efforts have likewise uncovered errors in benchmark problems and reference solutions (e.g.,(Ballon et al., 2026)), but such corrections are treated primarily as dataset maintenance. Here, by contrast, the correction feeds back into the mathematical source material itself: the exercise is not merely relabeled or removed, but supplied with a clearer alternative proof in a more motivated form (Pólya, 1949; Morris, 2020). In this respect, the episode echoes the way Lean formalization can expose and repair gaps in mathematical arguments, while differing in that the pressure comes from comparison with natural-language LLM reasoning rather than from full formalization. To our knowledge, this is among the first recorded instances where benchmarking efforts feed back to mathematics by prompting longer proof discussions.

2. Yu Tsumura's Problem

Yu Tsumura's problem was introduced in the year 2017 (we use an immutable `archive.org` link to guard against changes to the website[3]), predating LLMs. This link contains a copy of the statement and the proof. There are earlier listings online, the earliest that we could find was the problem being listed in a collection of exercises.[4]

It is a group theory problem, but we emphasize that no specialized knowledge of group theory is needed. We reproduce it below:

> *Let x, y be generators of a group G with relations*
>
> $$xy^2 = y^3x,$$
>
> $$yx^2 = x^3y.$$
>
> *Prove that G is the trivial group.*

Independent of group theory. Yu Tsumura's proof, which is also provided in the link above, is short and requires nothing more than clever symbolic manipulation–a task that LLMs solving Olympiad-level math problems need to possess. In fact, this problem, if one unpacks the group theory content, is similar to an IMO problem, such as a functional equation problem or inequality, where there are established proof techniques. This should be the easiest case for an LLM in terms of problem difficulty. While the proof, provided in the link above, utilizes the concepts of conjugacy and the order of a group element, these are only mentioned once and can also be unpacked. This makes the proof independent of any specific knowledge about group theory.

3. LLM Eval Results

All our evaluations were performed zero-shot, i.e., a single attempt was made to obtain the answer, over three runs. Our assessment is made from the point of view *of an end user at the present point in time*, thus we count apparent context overflows as failed attempts. Thus, we are assessing whether the model can answer Yu Tsumura's 554th problem *robustly*, which means that the model has to produce the correct answer most of the time, so a zero-shot evaluation should be sufficient. Repeated evaluations might produce correct proofs, but if it takes a best-of-n approach, majority voting, or other techniques to elicit them, from the perspective of the end user, this would be a different model that is evaluated (namely, one where the tested LLM incorporates an output refinement strategy on top, that mirrors the repeating-evaluation framework).

The list of models that we queried is given in Tables 1 and 2. These models arguably represent the state of the art among publicly available options in 2025, and 2026, respectively.

[1] The original proof uses conjugation as a group-theoretic concept, though our new proof, highlighted in Section B does not.

[2] Their exact ability has not yet been fully elucidated, as the research points to a nuanced picture, see (Petrov et al., 2025) vs. (Huang & Yang, 2025).

[3] The original link can be reconstructed from the archive link as `archive.org` contains the original one in its URL.

[4] `https://www.scribd.com/document/937648373/200a-f16-hw2\protect\protect\leavevmode@ifvmode\kern+.1667em\relaxexercise9.`

Although this list is not exhaustive, these models likely outperform most others and are the most highly rated ones on websites such as lmarena.ai. Therefore, we reason that if these models are unable to solve the problem, it is unlikely that other comparable or less capable models will succeed either.

The failures in each case in Table 1 are fatal to the proof. In all cases, the model relied on the error we listed to complete its output. None of the models makes really significant progress before such an error derails the model, or in the case of the "argument incomplete" annotation, the model appears to give up and declares success before much meaningful progress is made, or ceases reasoning without returning a proof. For Table 1, because, except for Gemini 2.5 DeepThink, we evaluated all models three times, and all failed, we do not list pass rates. Gemini 2.5 DeepThink was evaluated only once, and failed. We list the failure type for the first run only, and the time of the first evaluation. In Table 2, we can observe that the situation has improved approximately 9 months later, but still no model is able to solve the problem flawlessly in all three attempts.

All LLM outputs, including those performed in 2026, are available in Markdown at https://github.com/wbhart/yutz.

We have included the failures for the frontier models GPT-5.5 Thinking and Gemini 3.1 Pro, which were able to solve the problem, in Section D.

Our analysis spans a total of 31 LLMs, including 91 proof attempts.

4. Human Comparison and New Proof

Yu Tsumura's 554th problem belongs to the domain of group theory, which is not a domain that is present in problems from the IMO.

Inspection of the original proof reveals that no specific group theory knowledge or group theoretic proof strategies are needed to solve Yu Tsumura's problem, beyond the definition of a group and a generator. Nonetheless, it is unclear whether the difficulty of the problem is within the reach of an IMO-level competitor.

To clarify this, we carried out an $n = 1$ study with a former IMO25 participant, who was not yet exposed to any group theory.

Precise instructions were provided to him that he was not to look up information about group theory, and to receive any information he needed to understand the foundational group theoretic definitions solely from interactions with ChatGPT. We share the full, unredacted transcript of interactions with ChatGPT, as well as his write-up of the proof.[5]

The fact that an IMO participant was able to solve this problem demonstrates, perhaps unsurprisingly, that Yu Tsumura's 554th problem is well within the reach of IMO-level students.

More interesting is to observe that the student engaged, without being explicitly prompted to do so in any way, to devise a proof that comes close to what is known as a "motivated proof" (Pólya, 1949; Morris, 2020). These are proofs where each step made in the proof is made more transparent by providing clear motivation. It was observed that LLMs struggle with devising motivated proofs (Frieder et al., 2024), and the current paper highlights more strongly the distinction in human and LLM proof quality. This pertains both to the thoughtful outline he shared that represents his thinking process, as well as the final proof that resulted from this process.

What is noticeable about the solution by the IMO participant is that after spotting the identity $xy^{2n}x^{-1} = y^{3n}$, which some LLMs are also able to devise, the participant exploits it by picking special values of n.

In particular, the participant noticed that we can replace a factor of 2 with a factor of 3, at the expense of wrapping things in another x and x^{-1}. But then by just focusing on the power of 3 dividing n, specifically by assuming n is divisible by a sufficiently high power of 3, the participant can control the use of the identity. The main thing that is different here is that he keeps focusing on powers, whereas both Yu Tsumura's original solution and other known correct proofs prove some identity involving y^{27} but then go back to using the original identities to knock out the rest of the solution (or, in case of LLMs, just "fiddling" until the whole thing collapses).

The IMO participant's solution is much nicer because it shows why 27 is important in the proof and where it comes from - a motivated proof step. We have included the proof in Appendix B and invited him as co-author.

This shows that whereas the LLMs seem to need to spend a lot of their time just trying random algebraic manipulations with little clear direction, the IMO participant clearly motivated different proof strategies, which highlights a completely different approach to problem-solving that LLMs lack.

Lastly, we note that 60 members of the public have self-reported on Yu Tsumura's website that they solved the problem.

[5] https://github.com/friederrr/yutsumuras_554th_problem

LLM	Access	Eval Date	Failure
o3-pro (D.1)	OpenRouter	summer 2025	D
o3 (D.2)	online GUI	summer 2025	C, T
o4-mini-high (D.3)	online GUI	summer 2025	T
GPT-4o (D.4)	online GUI	summer 2025	I
Gemini 2.5 Pro (D.5)	online GUI	summer 2025	A
DeepSeek R1 (D.6)	online GUI	summer 2025	U
Claude Sonnet 4 (Ext. Thinking) (D.7)	online GUI	2025	U
Claude Opus 4 (Ext. Thinking) (D.8)	OpenRouter	2025	T
Grok 4 0709 (D.9)	LMArena	summer 2025	U
Kimi K2 (D.10)	OpenRouter	summer 2025	A, I
Qwen3 235B A22B Thinking 2507 (D.11)	OpenRouter	2025	U
GLM-4.5 (D.12)	OpenRouter	2025	A
Gemini 2.5 Deep Think (D.13)	online GUI	2025	A
Llama 4 Maverick (D.14)	LMArena	2025	U
DeepSeek v3 0324 (D.15)	OpenRouter	2025	I
QwQ 32B (D.16)	LMArena	2025	A, U
GPT-OSS-120B (D.17)	OpenRouter	2025	U
GPT-5 Thinking (D.18)	online GUI	2025	D

Table 1. A table of 18 evaluated LLMs on Yu Tsumura's 554th problem, evaluated in a time span of less than three weeks in summer 2025 (28 July to 16 Aug) for the first evaluation, with links to the Appendix containing full outputs and detailed failure mode descriptions. For both Claude models, the "Extended Thinking" option was turned on. Some models are missing size specifications; for example, the GUI used to access DeepSeek R1 did not reveal this information about the underlying model. See Appendix A for more information. All models were evaluated with zero-shot, on three runs, except Gemini 2.5 Deep Think which was evaluated only once. None of the listed models were able to solve it flawlessly, as outlined by the failure modes (see key below), which report the failures on the first run. We refer to Section D for full output traces and a detailed explanation of the (potentially multiple) critical failure modes, and of the lines of the proof on which they occur.

Key: A = algebra error, C = missed case, D = incompatible definition,
I = argument incomplete, T = inapplicable theorem,
U = unwarranted assumption/claim

LLM	Access	Eval Date	Pass rate
Gemini 3.1 Pro		spring 2026	2/3
GPT-5		spring 2026	0/3
GPT-5.1		spring 2026	0/3
GPT-5.2		spring 2026	0/3
GPT-5.4		spring 2026	0/3
GPT-5.5		spring 2026	2/3
GPT Pro 5.2		spring 2026	2/3
GPT Pro 5.4		spring 2026	1/3
Claude Opus 4.0		spring 2026	0/3
Claude Opus 4.1		spring 2026	0/3
Claude Opus 4.5		spring 2026	0/3
Claude Opus 4.6		spring 2026	1/3
Claude Opus 4.7		spring 2026	0/3
Claude Sonnet 4.5		spring 2026	0/3
Claude Sonnet 4.6		spring 2026	0/3

Table 2. Historical pass rates for model lineages evaluated on Yu Tsumura's 554th problem in spring 2026. Pass rates are reported as passes over available zero-shot evaluations; apparent context overflows are counted as non-passes.

5. Discussion and Original Proof

We speculate about two reasons why such a large cohort of models may have failed to solve a problem that is publicly available.

- **Step-by-step reliability** One possible explanation is that the failure is not primarily due to the intrinsic difficulty of the theorem, but to the models' tendency to make a local mistake before making any substantial progress. In earlier LLM generations, these mistakes varied from proof to proof and typically did not coincide with the specific error in the published solution: For example, pass@n results, for each value of $n \in \{1 \dots, 8\}$, in the *best-of-n* data split of the OpenProof Corpus (Dekoninck et al., 2025), suggest that out of the total of 8 proofs that o4-mini generated for 60 competition-level problems, a significant number of proofs were wrong.[6] As model reliability improved, however, this explanation becomes less convincing. Anecdotal evidence suggests these systems went from solving a fraction of early graduate-level problems to solving nearly all of them, which suggests that simple step-by-step unreliability is no longer a sufficient explanation for their failure on this problem, and which may explain why current best LLMs occasionally manage to solve this problem.

[6]While there exist several further paper that study the performance of LLMs on multiple generations of the problem, few make in-depth analysis of correctness, e.g. (Sun et al., 2025) and most results in this domain are anecdotal.

- **Need for deeper, verifier-guided reasoning pipelines.** Another possible explanation is that the model is not failing because the theorem is intrinsically out of reach, but because it does not search deeply enough through the relevant algebraic transformations. Problems of this kind are, at least in principle, well suited to verifier-guided training: many intermediate algebraic steps can be checked automatically, for instance using symbolic algebra systems or automated theorem provers such as Prover9. This connects our example to recent work showing that reinforcement learning from verifiable rewards can improve LLM performance on mathematical reasoning tasks (Guo et al., 2025), as well as work on process supervision and step-level verification (Lightman et al., 2024; Zelikman et al., 2022). The fact that such a large number of models have issues with Yu Tsumura's 554th problem may therefore indicate not a fundamental limitation of the models, but indicate that few models are trained on reasoning pipelines that provide sufficiently targeted feedback for the kind of deep symbolic search required here.

- **Erroneous training data** A third explanation is that the failure is due to the presence of a flawed proof in the training data, even though the proof strategy is correct. It was pointed out to us that the original proof of Yu Tsumura's problem, which was posted online in 2017, contains an algebraic error. This error may have been present in the training data for the LLMs, and could have led them to learn an incorrect approach to solving the problem. The fact that we reported the lack of LLM solutions of this problem in the first preprint may have led to companies' internal training pipelines fixing this proof, which also may explain the improved performance.

  To assess the nature of the algebraic error, we reproduce the original proof of Yu Tsumura's problem,

> *Let x, y be generators of a group G with relations*
>
> $$xy^2 = y^3x, \quad (1)$$
> $$yx^2 = x^3y. \quad (2)$$
>
> *Prove that G is the trivial group.*

  below:

  Proof. From the relation (1), we have

$$xy^2x^{-1} = y^3.$$

  Computing the power of n of this equality yields that

$$xy^{2n}x^{-1} = y^{3n} \quad (3)$$

for any $n \in \mathbb{N}$. In particular, we have

$$xy^4x^{-1} = y^6 \quad \text{and} \quad xy^6x^{-1} = y^9.$$

Substituting the former into the latter, we obtain

$$x^2y^4x^{-2} = y^9. \tag{4}$$

Cubing both sides gives

$$x^2y^{12}x^{-2} = y^{27}.$$

Using the relation (3) with $n = 4$, we have $xy^8x^{-1} = y^{12}$. Substituting this into equality (4) yields $x^3y^8x^{-1} = y^{27}$.

Now we have

$$y^{27} = x^3y^8x^{-1} = (x^3y)y^8(y^{-1}x^{-3}) = yx^2y^8x^{-2}y.$$

Squaring the relation (4), we have $x^2y^8x^{-2} = y^{18}$.

Substituting this into the previous, we obtain $y^{27} = y^{18}$, and hence

$$y^9 = e,$$

where e is the identity element of G.

Note that as we have $xy^2x^{-1} = y^3$, the elements y^2, y^3 are conjugate to each other. Thus, the orders must be the same. This observation together with $y^9 = e$ imply $y = e$.

It follows from the relation (2) that $x = e$ as well. Therefore, the group G is the trivial group. □

The line "Substituting this into equality (4) yields $x^3y^8x^{-1} = y^{27}$" is incorrect. While the displayed equality is true, (since $x = y = e$), it does not follow from the preceding step. The valid deduction would instead be

$$x^3y^8x^{-3} = y^{27}.$$

The written step therefore implicitly, and without justification, assumes $x^{-2} = e$, which is very close to assuming the desired conclusion. This mistake is then quietly reversed in the following line, where the exponent -1 becomes -3 again; that is another technically invalid deduction.

Later, the expression $yy^{18}y$ is also silently simplified to y^{18}, which is again incorrect. By that stage, however, the error is largely harmless, since the final conclusion is about as easy to derive from $y^{11} = e$ as from $y^9 = e$. As often happens, the proof contains several small mistakes that ultimately do not affect the truth of the result.

The theorem itself is true, and we have provided a novel proof for it in Appendix B, which does not rely on the original proof, and which is more motivated (cf. (Pólya, 1949; Morris, 2020)) than the original proof.

We note that **reduced human oversight** is the driving factor of all these issues, as the finer points of mathematical reasoning become apparent only through human evaluation, which then inform what training strategies have to be used. In contrast to final-answer benchmarks and evaluations, such as OlympiadBench (He et al., 2024), for which automatic assessment is possible, there are few benchmarks for assessing proof-based reasoning, due to the high human effort involved, e.g., exceptions are `matharena.ai`, (Petrov et al., 2025) and the earlier GHOSTS benchmark and evaluation (Frieder et al., 2023) problems. Hence, problems where human proof assessment is performed are necessary to establish the baseline of reasoning abilities, and the current failure shows that some gaps may still exist towards final-answer benchmarks that automatically assess the various facets of reasoning.

We conclude by observing that one could argue that another issue that might drive poor LLM performance is **confounding problem representation**: Yu Tsumura's 554th problem is presented as a group-theoretic exercise, but, as discussed in Section 4, its solution does not require any deeper group-theoretic knowledge. Once the group-theoretic layer is stripped away, the core argument appears to be within the reach of Olympiad-level human mathematical reasoning (even though its proof, due to its algebraic nature, is different than the proofs of typical IMO-level problems). One might therefore suspect that the problem's group-theoretic framing steers LLMs toward inappropriate reasoning paths, thereby obscuring the simpler underlying structure. However, we do not believe this explanation is sufficient. To discover Yu Tsumura's 554th problem and the problems it causes for LLMs, 300+ group-theoretic problems were evaluated, and LLMs performed substantially better on many of those. This suggests that the mere presence of group-theoretic language is not enough to explain the failure.

6. Limitations

Goodhart's law, which states that "when a measure becomes a target, it ceases to be a good measure" is pervasive in machine learning: This principle highlights that model creators often optimize the models to score highly on a given benchmark, rather than equipping the models with the skills that are partially captured by that benchmark and that are needed to succeed on that benchmark. In this regard, we expect that, having emphasized the problems commercial LLMs face on Yu Tsumura's 554th problem, models will soon be adapted to solve this issue (we hypothesize that for some state-of-the-art models, techniques as straightforward as improved test-time training will lead to the problem being solved). In fact, since our first release of the preprint in summer 2025, we noticed that some models are able to solve it, albeit inconsistently. Yet, we conjecture that even

in this case, other problems will be found on which LLMs will struggle across the board.

Our evaluation pertained exclusively to models that reasoned and did not use a RAG pipeline – since the solution is publicly available, such an approach would not have assessed the reasoning skills of the evaluated model. In the case of recent models, from o3 Pro onwards, it was necessary to explicitly prompt the model not to look up the solution online and in one case to encourage the LLM to produce a proof, not a Python program.

Our protocol was to give each model a single attempt at a solution. It is reasonable to assume that multiple attempts, especially with the more expensive models, may result in a more complete solution. Of course, commercial models may already do this internally, using techniques such as majority voting, or more sophisticated variants thereof. We did not follow this approach, because our analysis pertains solely to whether the experience of an *end user* interacting with these language models can live up to the expectation generated by the strong performance (such as the medals reported at IMO25).

We have focused on publicly released, widely deployed models, especially flagship models, and their lineages. We cannot exclude that there are boutique models or models that are not yet publicly deployed that can reliably solve the problem.

The difference in model capabilities might also be explained by differences in how much training on the test task was performed (Dominguez-Olmedo et al., 2024).

Lastly, a mathematical problem with a proof that relies mainly on symbolic manipulation will pose few issues for a symbolic solver tailored to this type of reasoning. In this regard we expect that an LLM that has access to a tool, such as Prover9, or Vampire [7] or some other solver, and can translate the problem into the necessary formalism, will be able to solve it.

7. Conclusion

We have demonstrated that there exists at least one problem drawn from a similar distribution in terms of human difficulty and solution strategies as IMO problems, on which LLMs have demonstrated very strong performance to date, on which LLMs nonetheless systematically fail. Thus, subject to the constraints mentioned in Section 6, reasoning in LLMs remains brittle.

The fact that none of the LLMs aced Yu Tsumura's 554th problem, and few solved it, contrary to the performance of a human, provides further evidence that LLMs' "thinking" is different from the thinking of a human, and, in particular, that their reasoning ability is not *transitive*: Solving problems of a similar level of difficulty does not guarantee that another, similar such problem can be solved.

We are cognizant that Yu Tsumura's 554th problem will soon be solved reliably by LLMs, in particular, once attention has been drawn to the fact that it is not currently reliably solved. Our analysis was carried out over a time frame of less than three weeks between July and August 2025 and again a time frame of two weeks in April 2026, and meant to capture a noteworthy snapshot at each time.

Nonetheless, this doesn't change our overall message, as in that case we believe other problems will emerge that should be accessible to LLMs based on their performance on comprehensive benchmarks will turn out not to be solvable by a wide set of LLMs.

We note that the pressure that benchmarking LLMs on increasingly hard problems will put on the correctness of mathematical literature, will lead to an increasingly strong feedback loop where benchmarking efforts improve mathematics. To date, this feedback loop is mostly confined to ensure benchmarks and datasets are error-free. Our benchmarking effort elicited, from this point of view, a very modest, but nonetheless meaningful improvement in terms of a new proof, of motivated type (Pólya, 1949; Morris, 2020), of a tricky math problem, within the reach of an IMO contestant.

We refer to Section C from the Appendix for a detailed description of how we arrived at the current problem. This was not a clear filtering effort, but a long-term evaluation that spun several LLMs and several attempts to find group-theoretic problems that SOTA LLMs were not able to solve. In total over 300 group-theoretic problems were evaluated. A set of historical results on older LLMs is contained in that section from the Appendix. While obsolete, to our knowledge this is the only clear historical record of LLMs performance on group theory problems and serves to record the rapid rise of LLM's mathematical abilities, using group theory as a case study, as well as a qualitative assessment of failure modes throughout time.

We conclude with concerns we have going forward: While LLMs have started to be integrated with mathematical research problems and have emerged as being able to solve open conjectures (Barreto, 2026), our evaluations show that LLMs are far from universal problem solvers, and that their impressive abilities in some mathematical domains are offset by clear failures in others. The lack of systematic studies that promote human assessment may lead to blind spots in terms of training approach that may cover holes in LLMs' reasoning performance.

[7] https://github.com/vprover/vampire

8. Acknowledgements

We thank András Salamon for pointing out the algebraic error in the original Yu Tsumura solution, and providing explanations why it does not invalidate the proof.

References

Ballon, M., Algaba, A., Verbeken, B., and Ginis, V. Benchmarks saturate when the model gets smarter than the judge. *arXiv preprint arXiv:2601.19532*, 2026.

Barreto, K. Problem 728 and the use of AI on Erdős problems. https://www.erdosproblems.com/forum/thread/blog:2, jan 2026. Erdos Problems Forum blog post, 26 January 2026. Accessed 23 April 2026.

Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Dekoninck, J., Petrov, I., Minchev, K., Balunovic, M., Vechev, M., Marinov, M., Drencheva, M., Konova, L., Shumanov, M., Tsvetkov, K., et al. The Open Proof Corpus: A large-scale study of LLM-generated mathematical proofs. *arXiv preprint arXiv:2506.21621*, 2025.

Dominguez-Olmedo, R., Dorner, F. E., and Hardt, M. Training on the test task confounds evaluation and emergence. *arXiv preprint arXiv:2407.07890*, 2024.

Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P., and Berner, J. Mathematical capabilities of ChatGPT. *Advances in neural information processing systems*, 36:27699–27744, 2023.

Frieder, S., Bayer, J., Collins, K. M., Berner, J., Loader, J., Juhász, A., Ruehle, F., Welleck, S., Poesia, G., Griffiths, R.-R., et al. Data for mathematical copilots: Better ways of presenting proofs for machine learning. *arXiv preprint arXiv:2412.15184*, 2024.

Gan, E., Zhao, Y., Cheng, L., Mao, Y., Goyal, A., Kawaguchi, K., Kan, M.-Y., and Shieh, M. Reasoning robustness of LLMs to adversarial typographical errors. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 10449–10459, 2024. doi: 10.18653/v1/2024.emnlp-main.584. URL https://aclanthology.org/2024.emnlp-main.584.

Glazer, E., Erdil, E., Besiroglu, T., Chicharro, D., Chen, E., Gunning, A., Olsson, C. F., Denain, J.-S., Ho, A., Santos, E. d. O., et al. Frontiermath: A benchmark for evaluating advanced mathematical reasoning in AI. *arXiv preprint arXiv:2411.04872*, 2024.

Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., S, Z., et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025. doi: 10.1038/s41586-025-09422-z.

Hao, Y., Wan, X., and Zhai, C. An investigation of robustness of LLMs in mathematical reasoning: Benchmarking with mathematically-equivalent transformation of advanced mathematical problems. *arXiv preprint arXiv:2508.08833*, 2025.

He, C., Luo, R., Bai, Y., Hu, S., Thai, Z. L., Shen, J., Hu, J., Han, X., Huang, Y., Zhang, Y., et al. OlympiadBench: A challenging benchmark for promoting AGI with olympiad-level bilingual multimodal scientific problems. *arXiv preprint arXiv:2402.14008*, 2024.

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., and Steinhardt, J. Measuring mathematical problem solving with the MATH dataset. *NeurIPS*, 2021.

Hou, Y., Zhang, T., Wang, M., Ma, X., and Huang, H. SMART: Self-generating and self-validating multi-dimensional assessment for llms' mathematical problem solving. *arXiv preprint arXiv:2505.16646*, 2025.

Huang, Y. and Yang, L. F. Winning gold at imo 2025 with a model-agnostic verification-and-refinement pipeline. *arXiv preprint arXiv:2507.15855*, 2025.

Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., and Cobbe, K. Let's verify step by step. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=v8L0pN6EOi.

Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S., and Farajtabar, M. Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models. *arXiv preprint arXiv:2410.05229*, 2024.

Morris, R. L. Motivated proofs: What they are, why they matter and how to write them. *The Review of Symbolic Logic*, 13(1):23–46, 2020.

Petrov, I., Dekoninck, J., Baltadzhiev, L., Drencheva, M., Minchev, K., Balunović, M., Jovanović, N., and Vechev, M. Proof or bluff? evaluating LLMs on 2025 USA math olympiad. *arXiv preprint arXiv:2503.21934*, 2025.

Pólya, G. With, or without, motivation? *The American Mathematical Monthly*, 56(10):684–691, 1949.

Schmitt, J., Bérczi, G., Dekoninck, J., Feusi, J., Gehrunger, T., et al. Improofbench: Benchmarking AI on research-level mathematical proof generation. *arXiv preprint arXiv:2509.26076*, 2025.

Sun, Y., Yin, Z., Huang, X., Qiu, X., and Zhao, H. Error classification of large language models on math word problems: A dynamically adaptive framework. *arXiv preprint arXiv:2501.15581*, 2025.

Yang, Y., Yamada, H., and Tokunaga, T. Evaluating robustness of LLMs to numerical variations in mathematical reasoning. In *Proceedings of the Sixth Workshop on Insights from Negative Results in NLP*, pp. 171–180, 2025. doi: 10.18653/v1/2025.insights-1.16. URL https://aclanthology.org/2025.insights-1.16.

Yu, T., Jing, Y., Zhang, X., Jiang, W., Wu, W., Wang, Y., Hu, W., Du, B., and Tao, D. Benchmarking reasoning robustness in large language models. *arXiv preprint arXiv:2503.04550*, 2025.

Zelikman, E., Wu, Y., Mu, J., and Goodman, N. D. STaR: Bootstrapping reasoning with reasoning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 15476–15488, 2022.

A. Additional LLM Information

In order to emulate the end user experience, we used various web GUIs to access the models. In some cases, only these were available.

We accessed DeepSeek R1 via the DeepSeek website. We presume this is DeepSeek R1 0528. This is a 685B model, comprising 671B main model weights.[8] It is accessed by turning on the DeepThink option in the DeepSeek web interface.

The Claude Sonnet 4 model was run with Extended Thinking turned on in the web interface.

The Claude Opus 4 model was accessed via OpenRouter and it spent 146.5s reasoning, indicating that the Extended Thinking mode was switched on.

On release day, GPT-5 suffered technical issues, such as returning responses unrelated to the prompt, or crashes. An output for GPT-5 was obtained, however OpenAI changed the available options quickly over the coming days and we decided not to use the release day response and wait until the models stabilised.

A later attempt on Aug 14 returned an empty answer after 14m 54s.

The included output for GPT-5 was obtained on Aug 16 with GPT-5 Thinking. The mathematical error was essentially the same as the one produced by GPT-5 on release day. However, the Aug 16 generation only thought for 2min 7s vs 4min 43s on release day. We don't know if this was due to server load or a change in the model.

Our summer 2026 evaluations were performed using web interfaces for current models and OpenRouter where possible on XHigh reasoning effort. Where there were problems with incorrect models being served, or server issues, we used LMArena for a small number of models.

B. New Proof of Yu Tsumura's 554th Problem

We present below a new proof that came out of the human study to assess whether an IMO could in principle solve Yu Tsumura's 554th problem.

We reproduce the problem statement below:

Let x, y be generators of a group G with relations

$$xy^2 = y^3x,$$

$$yx^2 = x^3y.$$

Prove that G is the trivial group.

The proof is a streamlined version of the original write-up that we linked to Section 4 and significantly more motivated than the original one.

Proof. As a first step, it makes sense to note how the two relations are symmetric, so for any result we get, we may swap x and y to get a symmetric result. Furthermore, we can rewrite them as $xy^2x^{-1} = y^3$. This is much easier to work with, since both sides behave very nicely under exponentiation (since $(xy^2x^{-1})^2 = xy^2x^{-1}xy^2x^{-1} = xy^4x^{-1}$).
We can now generalise the given identities from this to arrive at

$$xy^{2n}x^{-1} = y^{3n} \tag{1}$$

for any positive integer n. This is true, for example by induction, as $xy^{2n}x^{-1} = xy^{2(n-1)}x^{-1}xy^2x^{-1} = y^{3(n-1)}y^3 = y^{3n}$, and we are given the base case.
For example, we showed $xy^6x^{-1} = y^9$. This equation can be rewritten by replacing y^6 with xy^4x^{-1} to get $x^2y^4x^{-2} = y^9$. We have basically "replaced" a factor of 3 from the exponent of y on the LHS with a factor of 2. We can generalise this in the following way:

[8] https://huggingface.co/deepseek-ai/DeepSeek-V3

Let n be a positive integer divisible by 3, and a a positive integer such that $3^a \mid n$. Then

$$x^a y^{\frac{2^a}{3^a}n} x^{-a} = y^n. \tag{2}$$

Indeed, for $a = 0$ this is clearly true, and for $a = 1$ it is (1). Now assume it is true for all $a_0 \leq a$. We will show it for $a + 1$, assuming that $3^{a+1} \mid n$. We know $x^a y^{\frac{2^a}{3^a}n} x^{-a} = y^n$ from the inductive hypothesis and $y^{\frac{2^a}{3^a}n} = xy^{\frac{2^{a+1}}{3^{a+1}}n}x^{-1}$ from (1). Substituting the second into the first yields the desired identity.
Now with this, we are able to replace factors of 3 by factors of 2 on the LHS, at the cost of getting extra factors of x at the start and end. We now would like to use the symmetric given relation in x, rather than y, on the powers of x at the start/end to get to an equation no longer equivalent to everything so far, so that we do not get a trivial tautology when substituting them into each other. For this we need an exponent of at least 3, so take an n such that $3^3 \mid n$. We know then that the following holds:

$$x^3 y^{\frac{8}{27}n} x^{-3} = y^n. \tag{3}$$

Together with $yx^2y^{-1} = x^3$ and $yx^{-2}y^{-1} = x^{-3}$ this transforms the expression into

$$yx^2 y^{\frac{8}{27}n} x^{-2} y^{-1}.$$

We know $x^2 y^{\frac{8}{27}n} x^{-2} = y^{\frac{2}{3}n}$ once again by (2), substituting that in, we conclude

$$y^n = y^{\frac{2}{3}n}.$$

Hence $y^{\frac{1}{3}n} = e$. The minimal such n for which this holds is 27, so we have $y^9 = e$
We are now finished, since $e = xy^6x^{-1}$, so $y^6 = e = y^9$, hence $y^3 = e$, and $e = xy^2x^{-1}$, so $y^2 = e = y^3$, so $y = e$. Similarly, $x = e$, and as G is generated by x and y, it follows that $G = \{e\}$.

□

C. Finding the problem

The Yu Tsumura problem was identified as the result of a long and evolving process of LLM evaluation, beginning with GPT-4. We describe the history of this effort here.

When early versions of GPT-4 appeared, it became apparent that the model could reproduce standard textbook proofs in undergraduate mathematics. We undertook to comprehensively evaluate this capability.

A list of 321 problems in Group Theory, covering a standard undergraduate course on the subject, was constructed. GPT-4 was then evaluated on all 321 problems by expert human evaluation.

The topics covered were as follows:

- Axioms
- Identities
- Exponent
- Order
- Subgroups
- Cyclic groups
- Conjugacy
- Cosets
- Normal subgroups
- Quotient groups
- Homomorphisms
- Center
- Centralizer
- Normalizer
- Commutators
- Characteristic subgroups
- Direct products
- Direct sums
- Semidirect products
- Free groups
- Torsion
- Symmetric groups
- Alternating groups
- Group actions
- p-groups
- Sylow subgroups
- Groups of prescribed order
- Simple groups

This evaluation focused on proving theorems, and these ranged in difficulty from the trivial

Theorem C.1. *Let G be a group in which $x^2 = 1$ for all $x \in G$. Prove that G is abelian.*

which requires a couple of lines of manipulation of trivial identities, to theorems that would be difficult for a student even after being shown the solution

Theorem C.2. *Let G be a finitely generated group and H a subgroup of finite index. Prove that H is finitely generated.*

which requires construction of a Schreier generating set and a number of clever ideas, including coming up with ways to handle multiple levels of indexing of elements.

Of the 321 theorems, GPT-4 gave proofs with critical errors in 78 cases (24.3%) and 243 basically correct proofs (75.7%) of which 168 (52.3% of the entire collection) were judged to be completely without issue.

When o1-mini became available, a very similar evaluation was performed, this time with 352 theorems, of which o1-mini gave proofs with critical errors in 111 cases (31.8%) and 240 basically correct proofs (68.2%) of which 170 (48.3% of the entire collection) were judged to be completely without issue.

Although these two results could not be directly compared, due to methodological variation, because of differing context lengths, this study gave a baseline for human evaluation of early capable models on undergraduate and early graduate level theorem proving.

We also noted that there were multiple versions of GPT-4 and that later versions had a greater tendency to summarize proofs, which interfered with our evaluation strategy.

We identified numerous issues which made continued evaluation of later models in a similar way, difficult:

- Evaluation required weeks of effort per model
- Some models were frequently updated, exhibiting wildly different performance across versions
- Evaluation of individual problems, batch evaluation within a single context window per batch, or evaluation of the entire set in one context affected results
- Many models proved incapable of outputting their solutions in LaTeX, and requesting this caused some models to perform poorly
- Deciding whether a theorem cited or claim made is a hallucination, was time consuming
- Deciding whether to attribute a gap in a proof to lack of model understanding or to propensity to summarise, was difficult
- Our original dataset of theorems contained some errors, revealed by later evaluations
- Our collection of problems was mainly theorem proving (there were few existentially quantified problems or problems about concrete groups)

After identifying these shortcomings, we decided to construct a smaller collection of problems that could be feasibly evaluated on many LLMs using human evaluation. We also modified our evaluation strategy to deal with some of the known shortcomings.

This led to the creation of a set of 22 problems, which included both problems and theorems, abstract and concrete groups, from trivial to difficult. Some of these problems were from our original dataset, others were new.

The 22 problems were as follows.

Problem C.3. *Let G be a finite group and let exp(G) have the prime factorisation exp$(G) = p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$. Prove that for each $1 \leq i \leq r$ there is a subgroup H_i of G such that exp$(H_i) = p_i^{n_i}$.*

Problem C.4. *Let G be a group and $H \leq Z(G)$. Show that $H \lhd G$.*

Problem C.5. *Suppose P is a finite p-group. Recall the p-rank of G is the maximum rank of any abelian subgroup of P. Show that the p-rank is the rank of the largest elementary abelian subgroup of P.*

Problem C.6. *Let p be an arbitrary prime. Show that there exists a p-group P with N an abelian, normal subgroup and H a subgroup of N such that H is not normal in P.*

Problem C.7. *Show that any group of order 56 is solvable.*

Problem C.8. *Find the order of the smallest nonabelian, simple group.*

Problem C.9. *Let p and q be distinct primes. Find the possible orders of elements in the wreath product $C_p \wr C_q$.*

Problem C.10. *Show that there is a nonabelian group of order 24 with an element of order 4 but no element of order greater than 4.*

Problem C.11. *Show that there exists a nonabelian group of order 81 with elements of order 1, 3, 9 and 27.*

Problem C.12. *Prove that an abelian group G of order p^2 for a prime p is either cyclic or of the form $C_p \times C_p$.*

Problem C.13. *Show that a group G is cyclic iff G has a unique subgroup of each order dividing the order of G.*

Problem C.14. *Show that there exist two non-isomorphic groups of the same order that have the same number of elements of each order.*

Problem C.15. *Show that a group G of order 252 is solvable.*

Problem C.16. *Show that a group of exponent 2 is abelian.*

Problem C.17. *Find a group of exponent 3 which is non-abelian.*

Problem C.18. *Let $G = H \times K$ be a finite group with $gcd(exp(H), exp(K)) = 1$. Suppose that $X \leq G$ and $exp(X) = exp(G)$. Then $X \cong H_1 \times K_1$ for $H_1 \leq H$ and $K_1 \leq K$.*

Problem C.19. *Let K be a field and $K^\times$ be the multiplicative group of nonzero elements. Prove that any finite subgroup of $K^\times$ is a cyclic group.*

Problem C.20. *Let G be a group with generators a and b and relations $a^3 = 1$, $b^2 = 1$ and $a^2ba = bab$. Show the group is trivial.*

Problem C.21. *Let G be a finite group with $H, K \leq G$ with $\gcd(|H|, |K|) = 1$. Suppose $X \leq G \times G$ and the projection maps $\pi_1 : G \times G \to G$ and $\pi_2 : G \times G \to G$ when restricted to X are surjective maps to H and K respectively. Prove that $X = H \times K$.*

Problem C.22. *Let p^i divide the order of a finite group G. Then the number of subgroups of order p^i is congruent to 1 modulo p.*

Problem C.23. *Let P be a finite p-group and $\Phi(P)$ be the Frattini subgroup (intersection of all the maximal subgroups of P). Then*

$$\Phi(P) = P^p[P, P]$$

where $P^p = \langle \{g^p \mid g \in P\} \rangle$.

Problem C.24. *Let G be a finite group and ϕ an automorphism of G that is an involution, i.e. $\phi \circ \phi$ is the identity on G. Suppose that ϕ has no nontrivial fixed points, i.e. $\phi(g) = g$ iff $g = 1$. Show that G is abelian.*

All of these problems and the methods used to solve them, would be very familiar to experts. In fact, basic group theory is so well covered in the literature, that finding some corner of the undergraduate level or early graduate level landscape where truly unexplored directions still exist, would be challenging.

In other words, all of these problems lie very much in-distribution for language models trained on standard corpora.

Unsurprisingly, models improved drastically on our set of 22 problems over time. Indeed, some modern models could be expected to ace this test.

We provide a summary of our historical data on the number of fully correct solutions, by model, for the interest of the reader only. (Note that a later version of GPT-4 than the one mentioned above, was used, which performed poorly in this test.)

Whilst searching for problems that were suitable for such an evaluation, we encountered two notable problems which models seemed to fail on routinely.

One of the problems about finite index subgroups of finitely generated groups, was mentioned above, when discussing our original dataset.

Model	Score
GPT-4	1/22
GPT-4.5	2/22
GPT-4o	5/22
DeepSeek R1	6/22
o1-mini	6/22
o4-mini-high	8/22
Sonnet 3.7 Extended	9/22
Sonnet 3.7	9/22
o3-mini-high	10/22
Gemini 2 Pro Exp	11/22
Gemini 2 Flash Think. Exp	11/22
Grok 3 Beta Reasoning	11/22
Gemini 2.5 Pro Exp	15/22

Table 3. Observed historical model performance on 22 group theory problems

Apart from regular algebraic errors, we noted that many models mixed up left and right cosets when solving this problem. We concluded that this was likely due to the distribution of published solutions available online for this problem, versus that of other standard problems in group theory.

A second problem we identified is the Yu Tsumura problem. This was much more surprising.

Initially, models failed routinely on this problem due to the likelihood of an algebraic error being made before significant progress could be made on the problem.

Later on, the complete lack of *any* correct solutions by *any* LLM we had tried became very surprising (we had informally tried numerous stealth models and many other models not mentioned in this more formal study).

As of summer 2025, we were not aware of *any* models that could solve this problem, no matter how many attempts we allowed.

Not only so, but we were even unable to elicit a correct solution when a qualitative description of the official solution was provided to the model!

The Yu Tsumura problem stood out to us, compared to other problems that are difficult for LLMs because:

- No models had solved the problem by Summer 2025, i.e. failure was universal
- The solution required only unsophisticated algebraic manipulation
- Models rarely made any more than trivial progress before failing

Examination of reasoning traces available to us showed that models often repeatedly covered the same ground, not searching deep enough in the landscape of relations to make real progress. In many cases, this continued until the model made an algebraic error sufficient to convince it that it had a solution.

We came to realise that the Yu Tsumura problem was quite special. This was an elementary group theory exercise, which, at the time, no LLM could solve, while claims were being made in the popular press about "graduate level" advances in problem solving with LLMs.

Since that time, things have improved, but not by much.

Whilst academics are likely to remain skeptical about public marketing claims, the Yu Tsumura problem nevertheless highlights some difficulties with evaluating the extent to which models are truly generalizing versus being trained on specific tasks or problems.

Very few public details of exactly how models are being trained are available on state-of-the-art model. While . Incentives exist for academics to assist with training of models, but few similar incentives exist for arduous, human evaluation of models on problem corpora.

We do not know, for example, which models have been specifically trained to solve Yu Tsumura in response to our wider exposure of the problem in a preprint.

Finding similar problems is a far more difficult task than is widely appreciated. This becomes clear when asking LLMs to come up with novel problems. Anecdotal evidence shows that one can typically trace such problems to the exercise sections of a handful of standard textbooks, many of which are regularly discussed online and therefore enter into LLM pretraining data.

D. Output Traces

This appendix contains the unmodified solutions produced by the model, along with the lines at which critical errors are made. We also consider a substantially incomplete proof to be a critical error.

The prompt for all LLMs was identical: *Let G be a group with generators x and y and relations $xy2 = y3x$ and $yx2 = x3y$. Can you prove that G is the trivial group.* For models that do a web search, we add: *Do not perform a web search to solve this problem* to incentivize the model to reason rather than to regurgitate the web-based proof. For each model below, we first list the errors, and then reproduce the generated proof. At least one error was always critical, and we indicate which.

D.1. o3-pro

`Lines 3, 8 and 12`: The definition of the commutator used is $[x, y] = xyx^{-1}y^{-1}$. The definition of conjugation used is $x^y = y^{-1}xy$. But the identity $[x, yz] = [x, y][x, z]^y$ is not valid with these definitions. (Critical)

```
1  Write the two relations in commutator form.
2
3  (1)  $xy^2 = y^3x  \iff x y^2 x^{-1} = y^3  \iff [x , y^2] = y$
4  (2)  $yx^2 = x^3y  \iff y x^2 y^{-1} = x^3  \iff [y , x^2] = x$
5
6  Step 1.  Express y in terms of the basic commutator
7  Put $c = [x , y]$.
8  Because $[x , y^2] = [x , y][x , y]^y (= c c^y)$ we have
9    $y = c c^y$. (3)
10
11 Multiply (3) on the right by $(c^y)^{-1}$.
12 Since $(c^y)^{-1} = y^{-1} c^{-1} y$,   (3) gives
13
14   $y (c^y)^{-1} = c \iff y y^{-1} c^{-1} y = c \iff c^{-1} y = c$,
15 hence $y = c^2$. (4)
16
17 So $y$ is the square of the commutator $c$ and therefore commutes with $c$.
18
19 Step 2.  Express $x$ in terms of the commutator
20 Put $d = [y , x]$.  From (2) we similarly get
21
22   $x = d d^x$. (5)
23
24 Because $(d^x)^{-1} = x^{-1} d^{-1} x$, the calculation used above gives
25
26   $d^{-1} x = d  \implies x = d^2$. (6)
27
28 Step 3.  Relate d to c
29 But $[y , x] = [x , y]^{-1}, i.e. d = c^{-1}$. (7)
30
31 Combining (4), (6) and (7) we obtain
```

```

  $y = c^2 , x = (c^{-1})^2 = c ^{-2}$. (8)

Step 4.  $xy = 1$
Because powers of the same element commute,
  $xy = c ^{-2} c^2 = 1, so y = x^{-1}$. (9)

Step 5.  Finish with either defining relation
Insert $y = x^{-1}$ in the first original relation:

  $x y^2 = y^3 x \implies x·x ^{-2} = x ^{-3} · x$ ,
                      i.e. $x ^{-1} = x ^{-2}$.

Multiplying by $x^2$ gives $x = 1$, and then (9) forces $y = 1$.

Hence every generator of $G$ is the identity; therefore

  $G$ is the trivial group.
```

D.2. o3

`Lines 36-37`: The value k need not be an integer if n is infinity. (Critical)

`Lines 36-41`: The given map involves x and thus y need not be sent to an element of $\langle y\rangle$. The model has made the assumption that the automorphism of G (conjugation by x) induces an automorphism of the subgroup $\langle y\rangle$. (Critical)

```
1  Below is one way to see that the presentation
2
3  $$
4  G=\langle x,y\mid xy^{2}=y^{3}x,\;yx^{2}=x^{3}y\rangle
5  $$
6
7  defines nothing more than the trivial group.
8
9  ---
10
11 ### 1.  The abelianisation is trivial
12
13 Send $x\mapsto \bar x,\;y\mapsto\bar y$ in the free abelian group on two
   ↪  generators.
14 The two relators become
15
16 $$
17 \bar x+2\bar y=3\bar y+\bar x,\qquad
18 \bar y+2\bar x=3\bar x+\bar y,
19 $$
20
21 and both reduce to $ \bar x=\bar y=0$.
22 Hence $G^{\text{ab}}=G/[G,G]=1$; in particular $G=[G,G]$ is *perfect*.
23 If we can show that $G$ is abelian, it will therefore be trivial.
24
25 ---
26
27 ### 2.  Conjugating the cyclic subgroups
28
29 Write $n=\lvert y\rvert\;(1\le n\le\infty)$.
30 From $xy^{2}=y^{3}x$ we get
31
32 $$
33 x y^{2}x^{-1}=y^{3}. \tag{2.1}
34 $$
35
36 Because the subgroup $\langle y\rangle$ is cyclic, there is an integer
37 $k\;(1\le k\le n)$ such that
38
39 $$
40 x y x^{-1}=y^{k}. \tag{2.2}
41 $$
42
43 Squaring (2.2) and comparing with (2.1) yields the congruence
44
```

```
$$
2k\equiv 3\pmod n, \qquad\gcd(k,n)=1. \tag{2.3}
$$

Exactly the same argument with the second relator $yx^{2}=x^{3}y$ gives numbers
$m=\lvert x\rvert$ and $\ell$ satisfying

$$
y x y^{-1}=x^{\ell},\qquad 2\ell\equiv 3\pmod m,\qquad\gcd(\ell,m)=1. \tag{2.4}
$$

*If $n=\infty$* then (2.3) would read $2k=3$ in the integers, impossible.
Thus $y=1$ and, inserting this in the first relator, $x=1$.
So from now on we may assume $n,m<\infty$; necessarily $n,m$ are **odd**
↪ (otherwise $2$ would not be invertible mod $n$ or $m$).

---

### 3. A useful commutator

By (2.2)-(2.4),

$$
[x,y]=x y x^{-1}y^{-1}=y^{k-1}=x^{1-\ell}. \tag{3.1}
$$

Because of the congruences $2k\equiv3$ and $2\ell\equiv3$ we have

$$
\gcd(k-1,n)=\gcd(\ell-1,m)=1. \tag{3.2}
$$

(Indeed, if a divisor $d$ satisfied $d\mid n$ and $d\mid k-1$ then
↪ $k\equiv1\pmod d$; reducing $2k\equiv3\pmod n$ modulo $d$ would give
↪ $2\equiv3\pmod d$, hence $d=1$.)

From (3.1)-(3.2) the element

$$
c:=y^{k-1}=x^{1-\ell}
$$

has order

$$
\lvert c\rvert=n=\lvert y\rvert = m=\lvert x\rvert,
$$

so $\langle c\rangle=\langle y\rangle=\langle x\rangle$.
Thus **both** generators are powers of the *same* element $c$; consequently $G$
↪ is cyclic and therefore abelian.
```

```markdown

---

### 4.  Finish the argument

We have shown

* $G$ is abelian (Step 3),
* $G$ is perfect (Step 1).

The only perfect abelian group is the trivial one, so $G=1$.
Equivalently $x=1$ and $y=1$.

$$
\boxed{\;G\text{ is the trivial group.}\;}
$$

---

### Why the key step works

The heart of the proof is the observation that the first relator lets us
*measure* how conjugation by $x$ acts on the cyclic subgroup generated by $y$;
 ↪  it must multiply exponents by $k$ with $2k\equiv3$.
Because $k-1$ is coprime to $|y|$, the commutator $[x,y]=y^{k-1}$ still has full
 ↪  order and ends up generating the **same** cyclic subgroup as $y$.
The second relator yields the mirror-image statement for $x$.
That forces the whole group to be cyclic-hence, after the abelianisation
 ↪  calculation, trivial.


```

D.3. o4 mini high

`Lines 21-37`: It is shown that $\langle y^2 \rangle$ and $\langle y^3 \rangle$ are conjugate in G. But the following comments require that they be conjugate in $\langle y \rangle$, which is not demonstrated. (Critical)

`Lines 53-71`: Again, conjugation is in G, not in the cyclic subgroup generated by y. Thus the stated conclusions are not supported. (Critical)

`Lines 93-100`: The given map involves x and may not even be a map from $\langle y \rangle$ to itself, let alone an automorphism. (Critical)

```
1  Here is a self-contained proof that
2
3  $$
4  G = \langle x,y \mid xy^2 = y^3x,\; yx^2 = x^3y\rangle
5  $$
6
7  is the trivial group.
8
9  ---
10
11 ## 1. No element can have infinite order
12
13 From
14
15 $$
16 xy^2 = y^3x
17 \quad\Longrightarrow\quad
18 x\,y^2\,x^{-1} = y^3
19 $$
20
21 we see that the subgroups
22
23 $$
24 \langle y^2\rangle
25 \quad\text{and}\quad
26 \langle y^3\rangle
27 $$
28
29 are conjugate in $G$.  But in an infinite cyclic group $\langle y\rangle$, the
   ↪  index of $\langle y^k\rangle$ is $|k|$.  Conjugate subgroups have the same
   ↪  index, so you would get
30
31 $$
32 [\<y\>:\<y^2\>] \;=\;2,
33 \qquad
34 [\<y\>:\<y^3\>] \;=\;3,
35 $$
36
37 a contradiction.  Hence **$y$ must have finite order**.  By an entirely
   ↪  symmetric argument from
38
39 $$
```

```
yx^2 = x^3y
\quad\Longrightarrow\quad
y\,x^2\,y^{-1} = x^3,
$$

also **$x$ has finite order**.

---

## 2. Reduction to a cyclic, perfect group

Let

$$
m = \mathrm{ord}(y),\quad n = \mathrm{ord}(x).
$$

Because $xy^2x^{-1}=y^3$, conjugation by $x$ carries
$\langle y^2\rangle$ onto $\langle y^3\rangle$.  In a cyclic group of order $m$, the index of $\langle y^k\rangle$ is

$$
\bigl[\<y\>:\<y^k\>\bigr]
=\frac{m}{\gcd(m,k)}.
$$

Equating the two indices gives

$$
\frac{m}{\gcd(m,2)} \;=\;\frac{m}{\gcd(m,3)}
\;\Longrightarrow\;
\gcd(m,2)=\gcd(m,3).
$$

Since $\gcd(m,2)\in\{1,2\}$ and $\gcd(m,3)\in\{1,3\}$, the only way they are equal is

$$
\gcd(m,2)=\gcd(m,3)=1,
$$

i.e.\ $m$ is coprime to $6$.  In particular, **2 and 3 are invertible in $\mathbb{Z}/m\mathbb{Z}$.**

By exactly the same reasoning from the second relation one sees that

$$
\gcd(n,2)=\gcd(n,3)=1
$$

and so **2 and 3 are invertible mod $n$.**
```

```markdown

---

## 3. Linearising the conjugations

Because $\gcd(2,m)=1$, the map

$$
\varphi\colon\<y\>\to\<y\>,\quad
\varphi(y^k)=x\,y^k\,x^{-1}
$$

is an automorphism of the finite cyclic group of order $m$.  From

$$
\varphi(y^2)=y^3
\quad\Longrightarrow\quad
\varphi(y)^2 = y^3,
$$

we deduce that

$$
\varphi(y)=y^c,
$$

where $c$ is the unique integer mod $m$ satisfying

$$
2c\equiv3\pmod m.
$$

Since $\gcd(2,m)=1$, we have

$$
c-1 \;\equiv\;2^{-1}\;\;(\bmod\,m),
$$

so $\gcd(c-1,m)=1$.  Thus

$$
x\,y\,x^{-1} \;=\; y^c.
$$

Similarly, from $\,y\,x^2\,y^{-1}=x^3$ and $\gcd(2,n)=1$ one shows

$$
y\,x\,y^{-1} \;=\; x^d,
$$

where $2d\equiv3\pmod n$ and hence $\gcd(d-1,n)=1$.
```

```

---

## 4. The commutator forces cyclicity {&} perfection

Compute the two "commutators":

$$
[x,y] \;=\; x\,y\,x^{-1}\,y^{-1}
\;=\; y^c\,y^{-1} = y^{\,c-1},
$$

$$
[y,x] = y\,x\,y^{-1}\,x^{-1}
= x^d\,x^{-1} = x^{\,d-1}.
$$

But always $[y,x]=[x,y]^{-1}$, so

$$
x^{\,d-1} \;=\; \bigl[y,x\bigr]
\;=\;\bigl[x,y\bigr]^{-1}
\;=\;y^{\,1-c}.
\tag{*}
$$

Now:

1. Since $\gcd(c-1,m)=1$, there is an integer $u$ with

$$
u\,(c-1)\equiv1\pmod m.
$$

Raising $[x,y]=y^{c-1}$ to the $u$th power gives

$$
[x,y]^u \;=\; y^{\,u(c-1)} = y,
$$

i.e.\\

$$
y \;=\; [x,y]^u
\;\in\; [G,G].
\tag{1}
$$

2. Since $\gcd(d-1,n)=1$, there is an integer $v$ with

$$
```

```
v\,(d-1)\equiv1\pmod n.
$$

Raising equation $(*)$ to the $v$th power gives

$$
x^{\,v(d-1)} = y^{\,v(1-c)},
$$

but $v(d-1)\equiv1\pmod n$ so $x=y^{\,v(1-c)}$.  Hence

$$
x \;\in\;\langle y\rangle.
\tag{2}
$$

From $(1)$ and $(2)$, **both** generators $x,y$ lie in the *derived subgroup*
↪  $[G,G]$, so

$$
[G,G] = G,
$$

i.e.\ $G$ is **perfect**.  On the other hand $(2)$ shows $G$ is cyclic
↪  (generated by $y$), hence **abelian**.  The only group which is *both*
↪  abelian and perfect is the trivial group.

---

### Conclusion

No non-trivial group can satisfy simultaneously

$$
xy^2=y^3x
\quad\text{and}\quad
yx^2=x^3y,
$$

so

$$
\boxed{G\text{ is the trivial group.}}
$$

```

D.4. GPT-4o

`Line 118`: The preceding lines don't demonstrate that conjugation of y^n by x increases the exponent by 1. (Incorrect but not used)

`Line 348`: It is not clear which equation $x = yxy$ comes from. (Incorrect but not used)

`Line 354`: Having made no progress whatsoever, the model concludes the "only consistent solution" is the trivial group. (Critical)

```
1  We are given a group $G$ with generators $x$ and $y$, and two relations:
2
3  $$
4  (1) \quad xy^2 = y^3x, \quad \text{and} \quad (2) \quad yx^2 = x^3y.
5  $$
6
7  Our goal is to **prove that $G$ is trivial**, i.e., $x = y = 1$.
8
9  ---
10
11 ### Step 1: Rewrite the Relations
12
13 Let's manipulate each relation to bring like terms together.
14
15 #### From (1): $xy^2 = y^3x$
16
17 Multiply both sides on the right by $x^{-1}$:
18
19 $$
20 xy^2x^{-1} = y^3.
21 $$
22
23 So,
24
25 $$
26 x y^2 x^{-1} = y^3. \quad \text{(1')}
27 $$
28
29 This says that conjugating $y^2$ by $x$ gives $y^3$, or
30
31 $$
32 x y^2 x^{-1} = y^3 \quad \Rightarrow \quad x y^2 x^{-1} y^{-3} = 1.
33 $$
34
35 This is equivalent to:
36
37 $$
38 x y^2 x^{-1} y^{-3} = 1.
39 $$
40
41 #### From (2): $yx^2 = x^3y$
42
43 Multiply both sides on the right by $y^{-1}$:
```

```

$$
yx^2 y^{-1} = x^3.
$$

So,

$$
y x^2 y^{-1} = x^3. \quad \text{(2')}
$$

Similarly,

$$
y x^2 y^{-1} x^{-3} = 1.
$$

---

### Step 2: Define Conjugation Maps

Let's define the following:

* From (1'): $x y^2 x^{-1} = y^3 \Rightarrow x y^2 x^{-1} y^{-3} = 1$
* From (2'): $y x^2 y^{-1} = x^3 \Rightarrow y x^2 y^{-1} x^{-3} = 1$

These say that:

$$
x y^2 x^{-1} y^{-3} = 1 \quad \text{and} \quad y x^2 y^{-1} x^{-3} = 1.
$$

These are conjugation identities showing how $x$ and $y$ act on each other.

Let's now take (1') and apply the same conjugation again.

---

### Step 3: Conjugate (1') Again

From (1'):

$$
x y^2 x^{-1} = y^3.
$$

Now conjugate both sides by $x$:

Left-hand side:

$$
```

```
x (x y^2 x^{-1}) x^{-1} = x^2 y^2 x^{-2}.
$$

Right-hand side:

$$
x y^3 x^{-1}.
$$

So,

$$
x^2 y^2 x^{-2} = x y^3 x^{-1}.
$$

But from (1') again:

$$
x y^2 x^{-1} = y^3 \Rightarrow x y^3 x^{-1} = x (x y^2 x^{-1}) x^{-1} = x^2 y^2
↪  x^{-2}.
$$

So this checks out.

But this also tells us that each time we conjugate $y^n$ by $x$, the exponent
↪  increases by one.

So maybe try to **write $x$ and $y$ in terms of each other**.

---

### Step 4: Solve for One Variable

Go back to:

$$
x y^2 = y^3 x.
$$

Let's try to **move all $x$'s to one side**.

Start with:

$$
x y^2 = y^3 x
$$

Multiply both sides on the left by $x^{-1}$, then on the right by $x^{-1}$:

$$
y^2 = x^{-1} y^3 x \Rightarrow x y^2 x^{-1} = y^3.
```

```
$$

As before.

Similarly, from (2):

$$
yx^2 = x^3 y \Rightarrow y x^2 y^{-1} = x^3.
$$

Now, write these as:

* $x y^2 x^{-1} = y^3$
* $y x^2 y^{-1} = x^3$

Let's now define $a = x$, $b = y$, to simplify notation.

So:

* $a b^2 a^{-1} = b^3$  (1)
* $b a^2 b^{-1} = a^3$  (2)

Let's try to eliminate one of the variables using these relations.

---

### Step 5: Try to Find Orders of $x$, $y$

Suppose $x = 1$, the identity.

Then from (1):

$$
1 \cdot y^2 = y^3 \Rightarrow y^2 = y^3 \Rightarrow y^{-1} = 1 \Rightarrow y = 1.
$$

So $x = 1 \Rightarrow y = 1$.

Similarly, if $y = 1$:

From (2):

$$
1 \cdot x^2 = x^3 \Rightarrow x^2 = x^3 \Rightarrow x^{-1} = 1 \Rightarrow x = 1.
$$

So $y = 1 \Rightarrow x = 1$.

So **if either generator is identity, the other must be too**.
```

```

Now suppose neither is the identity, but we try to find an order.

Suppose $x$ and $y$ have finite order.

Let's say $x^n = 1$ and $y^m = 1$.

Try to find a contradiction.

But a better approach is to **define a new variable** and work through
 ↪  simplification.

---

### Step 6: Try to Simplify with Substitutions

From the conjugation relation:

$$
x y^2 x^{-1} = y^3 \Rightarrow x y^2 = y^3 x.
$$

Let's isolate $x$. Multiply both sides on the left by $y^{-3}$:

$$
y^{-3} x y^2 = x.
$$

So:

$$
x = y^{-3} x y^2.
$$

Now cancel $x$ from both sides:

Multiply both sides on the left by $x^{-1}$:

$$
1 = x^{-1} y^{-3} x y^2.
$$

Rewriting:

$$
x^{-1} y^{-3} x y^2 = 1.
$$

Similarly, from the second relation:

$$
```

```
y x^2 = x^3 y \Rightarrow x^2 = y^{-1} x^3 y \Rightarrow 1 = x^{-2} y^{-1} x^3
 ↪  y.
$$

So:

$$
x^{-2} y^{-1} x^3 y = 1.
$$

We now have:

* $x^{-1} y^{-3} x y^2 = 1$  (A)
* $x^{-2} y^{-1} x^3 y = 1$  (B)

Let's work with (A):

$$
x^{-1} y^{-3} x y^2 = 1 \Rightarrow y^{-3} x y^2 = x.
$$

Let's multiply both sides by $x^{-1}$ on the right:

$$
y^{-3} x y^2 x^{-1} = 1.
$$

So:

$$
y^{-3} x y^2 x^{-1} = 1.
$$

That is:

$$
x y^2 x^{-1} = y^3.
$$

We've gone in circles.

Let's take a step back.

---

### Step 7: Try a Free Group Homomorphism

Suppose $G$ is a group with these relations:

$$
G = \langle x, y \mid x y^2 = y^3 x,\ y x^2 = x^3 y \rangle.
```

```
$$

Let's define a homomorphism $\phi: G \to \mathbb{Z}$ to abelianize the group.

In the abelianization, $xy = yx$, so we treat all elements as commuting.

From the relations:

1. $x y^2 = y^3 x \Rightarrow x + 2y = 3y + x \Rightarrow 2y = 3y \Rightarrow y = 0$
2. $y x^2 = x^3 y \Rightarrow y + 2x = 3x + y \Rightarrow 2x = 3x \Rightarrow x = 0$

So in the abelianization, $x = y = 0$, so both are trivial in $G^{ab}$.

Therefore, $G$ is a **perfect group** (equal to its commutator subgroup).

Since the abelianization is trivial, we suspect $G$ is trivial or very small.

---

### Step 8: Try the Identity Directly

From:

$$
x y^2 = y^3 x
\Rightarrow x y^2 x^{-1} = y^3
\Rightarrow x y^2 x^{-1} y^{-3} = 1.
$$

Let's define:

$$
r = x y^2 x^{-1} y^{-3}, \quad s = y x^2 y^{-1} x^{-3}.
$$

Then $G = \langle x, y \mid r = 1,\ s = 1 \rangle$.

Now define a **map from $G$ to the trivial group**, mapping $x \mapsto 1, y \mapsto 1$. Clearly both relations are satisfied. So the trivial group is a quotient of $G$.

But we now show that **$x = y = 1$** in $G$.

From:

$$
x y^2 = y^3 x \Rightarrow x y^2 x^{-1} = y^3.
$$

```

```
Let's suppose $y \ne 1$, and see that the order of $y$ must be finite.

Let's try small orders.

Suppose $y^n = 1$.

Try $n = 1$: $y = 1 \Rightarrow x = 1$.

Try $n = 2$: then $y^2 = 1 \Rightarrow x = y x y \Rightarrow x = x$, okay.

Try $n = 3$: $y^3 = 1 \Rightarrow x y^2 = x \Rightarrow y^2 = 1 \Rightarrow y =
↪  1$. Contradiction.

So we always get back to $y = 1 \Rightarrow x = 1$.

Therefore, the only consistent solution is:

$$
x = y = 1.
$$

---

###  Final Conclusion:

$$
\boxed{G \text{ is the trivial group.}}
$$

All elements reduce to the identity under the given relations.
```

D.5. Gemini 2.5 Pro

`Lines 39-40`: a cannot be cancelled from the right, and even if this is a typo, the claimed identity $a = b^3ab^{-1}$ does not follow from $bab = b^4ab^{-1}$. This is an algebra error. (Critical)

```
1  Here is a proof that the group $G$ with generators $x$ and $y$ and relations
   ↪  $xy^2 = y^3x$ and $yx^2 = x^3y$ must be the trivial group.
2
3  ### **1. Rewriting the Relations**
4
5  The given relations are:
6  1.  $xy^2 = y^3x$
7  2.  $yx^2 = x^3y$
8
9  To simplify the manipulation of these relations, we can introduce new
   ↪  generators. Let $a = x$ and $b = y^{-1}$. Then $y = b^{-1}$. We can rewrite
   ↪  the original relations in terms of $a$ and $b$.
10
11 * **Relation 1:** $xy^2 = y^3x$ becomes $ab^{-2} = b^{-3}a$.
12 * **Relation 2:** $yx^2 = x^3y$ becomes $b^{-1}a^2 = a^3b^{-1}$.
13
14 Now, we can rearrange these new relations to get expressions for $a$ and $a^2$:
15 * From $ab^{-2} = b^{-3}a$, if we multiply by $b^3$ on the left, we get
   ↪  $b^3ab^{-2} = a$. If we then multiply by $b^2$ on the right, we obtain:
16     $b^3a = ab^2$ (R1')
17 * From $b^{-1}a^2 = a^3b^{-1}$, if we multiply by $b$ on the left, we get $a^2 =
   ↪  ba^3b^{-1}$. If we then multiply by $b$ on the right, we obtain:
18     $a^2b = ba^3$ (R2')
19
20 So, the original relations are equivalent to the pair of relations:
21 * $ab^2 = b^3a$
22 * $a^2b = ba^3$
23
24 ---
25 ### **2. Deriving a Key Identity**
26
27 Now, we will use these relations to show that $b$ must be the identity element.
28
29 From $ab^2 = b^3a$, we can derive expressions by progressively multiplying by
   ↪  $b^{-1}$ on the right:
30 * $a = b^3ab^{-2}$
31 * $ab = b^3ab^{-1}$
32 * $ab^2 = b^3a$ (This is R1')
33
34 Let's focus on the identity $ab = b^3ab^{-1}$. We can use this to analyze the
   ↪  expression $bab$:
35 $bab = b(ab)$
36 Substitute the expression for $ab$:
37 $bab = b(b^3ab^{-1}) = b^4ab^{-1}$
38
39 So, we have found that $bab = b^4ab^{-1}$. By canceling $b$ from the left and
   ↪  $a$ from the right (if $a$ is not the identity), we get:
```

```markdown
$a = b^3ab^{-1}$

Now we have two expressions for $a$:
1.  $a = b^3ab^{-2}$ (from R1')
2.  $a = b^3ab^{-1}$ (derived above)

Equating these two expressions gives:
$b^3ab^{-2} = b^3ab^{-1}$

By canceling $b^3a$ from the left, we get:
$b^{-2} = b^{-1}$

Multiplying by $b^2$ on the right gives:
$e = b$

---
### **3. Proving Triviality**

We have shown that $b = e$. Since we defined $b = y^{-1}$, this means $y^{-1}=e$, which implies **$y=e$**.

Now we can substitute $y=e$ back into one of the original relations. Let's use the second relation:
$yx^2 = x^3y$
Substituting $y=e$:
$ex^2 = x^3e$
$x^2 = x^3$

By canceling $x^2$ from the left, we find that **$x=e$**.

Since both generators $x$ and $y$ are the identity element ($e$), the group $G$ can only contain the identity element. Therefore, **$G$ is the trivial group**.

```

D.6. DeepSeek R1

`Lines 11-13`: It is stated that $xyx^{-1} = y^3$. This is claimed to follow from $gyg^{-1} = y^3$. But this is not proved. The claimed proof only shows that $(xyx^{-1})^2 = y^3$. (Critical)

```
1  The group \(G\) is defined by generators \(x\) and \(y\) and relations \(xy^2 =
   ↪  y^3x\) and \(yx^2 = x^3y\). The goal is to show that \(G\) is trivial,
   ↪  meaning \(x = 1\) and \(y = 1\), where 1 is the identity element.
2
3  From the given relations, it is possible to derive additional relations:
4  - From \(xy^2 = y^3x\), conjugation by \(x\) gives \(x y^2 x^{-1} = y^3\). Since
   ↪  conjugation is an automorphism, \(x y^2 x^{-1} = (x y x^{-1})^2\), so \((x y
   ↪  x^{-1})^2 = y^3\).
5  - Similarly, from \(yx^2 = x^3y\), conjugation by \(y\) gives \(y x^2 y^{-1} =
   ↪  x^3\), so \((y x y^{-1})^2 = x^3\).
6
7  Consider the element \(g = x y\). Conjugating \(y\) by \(g\):
8  \[
9  g y g^{-1} = (x y) y (x y)^{-1} = (x y) y (y^{-1} x^{-1}) = x y x^{-1}.
10 \]
11 From the earlier derivation, \(x y x^{-1} = k\) where \(k^2 = y^3\), but also
   ↪  from the relation \(g y g^{-1} = y^3\) (since \(g y g^{-1} = x y x^{-1}\)
   ↪  and \((x y x^{-1})^2 = y^3\) implies that conjugating again or directly, it
   ↪  follows that \(x y x^{-1} = y^3\)). Thus:
12 \[
13 g y g^{-1} = x y x^{-1} = y^3.
14 \]
15 Therefore:
16 \[
17 x y x^{-1} = y^3. \quad (1)
18 \]
19 Similarly, for the other relation, consider \(h = y x\). Conjugating \(x\) by
   ↪  \(h\):
20 \[
21 h x h^{-1} = (y x) x (y x)^{-1} = (y x) x (x^{-1} y^{-1}) = y x y^{-1}.
22 \]
23 From the derivation, \(y x y^{-1} = m\) where \(m^2 = x^3\), and similarly, it
   ↪  follows that:
24 \[
25 h x h^{-1} = y x y^{-1} = x^3.
26 \]
27 Thus:
28 \[
29 y x y^{-1} = x^3. \quad (2)
30 \]
31
32 From equations (1) and (2):
33 - Equation (1): \(x y x^{-1} = y^3\) implies \(x y = y^3 x\). \quad (3)
34 - Equation (2): \(y x y^{-1} = x^3\) implies \(y x = x^3 y\). \quad (4)
35
36 Now, use the original relations. From the original relation \(xy^2 = y^3x\)
   ↪  \quad (5), and from equation (3), multiply both sides by \(y\):
```

```
\[
x y^2 = y^3 x y. \quad (6)
\]
Equating (5) and (6):
\[
y^3 x = y^3 x y.
\]
Since the group operation is cancellative, left-multiply both sides by \(y^{-3}\):
\[
x = x y,
\]
which implies:
\[
x y x^{-1} = x x^{-1} \implies y = 1. \quad (7)
\]

Similarly, from the original relation \(yx^2 = x^3y\) \quad (8), and from equation (4), multiply both sides by \(x\):
\[
y x^2 = x^3 y x. \quad (9)
\]
Equating (8) and (9):
\[
x^3 y = x^3 y x.
\]
Left-multiply both sides by \(x^{-3}\):
\[
y = y x,
\]
which implies:
\[
y x y^{-1} = y y^{-1} \implies x = 1. \quad (10)
\]

From (7) and (10), \(x = 1\) and \(y = 1\). Therefore, the group \(G\) is trivial.

\boxed{\text{The group } G \text{ is trivial.}}
```

D.7. Claude Sonnet 4 (Extended Thinking)

`Line 39`: The model promises to verify the assumption $y^2 = e$, but this is never independently verified. (Critical)

`Lines 51-54`: The model assumes $x^2 = e$ without justification. (Critical)

```
1  Looking at this group presentation, I'll work systematically with the given
   ↪  relations to show that both generators must equal the identity.
2
3  Given relations:
4  1) $xy^2 = y^3x$
5  2) $yx^2 = x^3y$
6
7  **Step 1: Finding the orders of x and y**
8
9  From relation (1): $xy^2 = y^3x$
10
11 Let me multiply both sides on the left by $y^{-3}$ and on the right by $y^{-2}$:
12 $$y^{-3}xy^2y^{-2} = y^{-3}y^3xy^{-2}$$
13 $$y^{-3}x = xy^{-2}$$
14
15 Multiplying both sides on the right by $y^2$:
16 $$y^{-3}xy^2 = x$$
17
18 But from our original relation, $xy^2 = y^3x$, so:
19 $$y^{-3}(y^3x) = x$$
20 $$x = x$$
21
22 Let me try a different approach. From $xy^2 = y^3x$, I'll multiply on the left
   ↪  by $y^{-3}$:
23 $$y^{-3}xy^2 = y^{-3}y^3x = x$$
24
25 So $y^{-3}xy^2 = x$, which means $y^{-3}x = xy^{-2}$.
26
27 Multiplying both sides on the right by $y^2$:
28 $$y^{-3}xy^2 = xy^{-2}y^2 = x$$
29
30 From the original relation $xy^2 = y^3x$, we have $y^{-3}xy^2 = y^{-3}y^3x = x$.
31
32 This confirms our manipulation. Now, from $y^{-3}x = xy^{-2}$, multiplying on
   ↪  the right by $y^2$:
33 $$y^{-3}xy^2 = x$$
34
35 And since $xy^2 = y^3x$:
36 $$y^{-3}y^3x = x$$
37 $$x = x$$
38
39 But more importantly, from $y^{-3}x = xy^{-2}$, if we set $y^2 = e$ (which we'll
   ↪  verify), then:
40 $$y^{-3}x = x$$
41 $$y^{-3} = e$$
42 $$y^3 = e$$
```

```

**Step 2: Using symmetry for x**

Similarly, from relation (2): $yx^2 = x^3y$

Following the same process, multiply on the left by $x^{-3}$:
$$x^{-3}yx^2 = x^{-3}x^3y = y$$

This gives us $x^{-3}yx^2 = y$, so $x^{-3}y = yx^{-2}$.

Following similar logic, we get:
$$x^3 = e$$

**Step 3: Deriving stronger conditions**

Now I have $x^3 = e$ and $y^3 = e$.

Substituting back into the original relations:

From $xy^2 = y^3x$ with $y^3 = e$:
$$xy^2 = ex = x$$
$$y^2 = e$$

From $yx^2 = x^3y$ with $x^3 = e$:
$$yx^2 = ey = y$$
$$x^2 = e$$

**Step 4: Conclusion**

Now I have:
- $x^2 = e$ and $x^3 = e$
- $y^2 = e$ and $y^3 = e$

Since $\gcd(2,3) = 1$, there exist integers $a,b$ such that $2a + 3b = 1$.

For $x$: $x = x^1 = x^{2a + 3b} = (x^2)^a(x^3)^b = e^a \cdot e^b = e$

For $y$: $y = y^1 = y^{2a + 3b} = (y^2)^a(y^3)^b = e^a \cdot e^b = e$

Therefore, both generators equal the identity, which means $G$ is the trivial group.
```

D.8. Claude Opus 4 (Extended Thinking)

`Line 20`: This is an unwarranted assumption. The conjugation map is a map on G and so the result of conjugating y need not be a power of y. (Critical)

```
1  Looking at the given relations:
2  - $xy^2 = y^3x$
3  - $yx^2 = x^3y$
4
5  I'll analyze what these relations tell us about the group structure.
6
7  From the first relation $xy^2 = y^3x$, I can rewrite this as:
8  $$xy^2x^{-1} = y^3$$
9
10 This means that conjugating $y^2$ by $x$ gives us $y^3$.
11
12 Similarly, from the second relation $yx^2 = x^3y$:
13 $$yx^2y^{-1} = x^3$$
14
15 This means that conjugating $x^2$ by $y$ gives us $x^3$.
16
17 Let me define the conjugation map $\phi_x(g) = xgx^{-1}$. Since conjugation is a
   ↪  homomorphism, and $\phi_x(y^2) = y^3$, I have:
18 $$\phi_x(y)^2 = y^3$$
19
20 If I let $\phi_x(y) = y^a$ for some integer $a$, then:
21 $$(y^a)^2 = y^3$$
22 $$y^{2a} = y^3$$
23
24 This means $y^{2a-3} = 1$.
25
26 For this to hold, either:
27 1. $y$ has infinite order, which would require $2a = 3$ in $\mathbb{Z}$
   ↪  (impossible for integer $a$)
28 2. $y$ has finite order $m$, and $2a \equiv 3 \pmod{m}$
29
30 So $y$ must have finite order $m$. For the congruence $2a \equiv 3 \pmod{m}$ to
   ↪  have a solution, we need $\gcd(2,m) | 3$. Since 3 is odd, we need $\gcd(2,m)
   ↪  = 1$, which means $m$ must be odd.
31
32 By the same reasoning applied to the second relation, $x$ must have finite odd
   ↪  order $n$, and there exists $b$ such that $yxy^{-1} = x^b$ where $2b \equiv
   ↪  3 \pmod{n}$.
33
34 Now I'll use both relations together. From our conjugation equations:
35 - $xyx^{-1} = y^a$ where $2a \equiv 3 \pmod{m}$
36 - $yxy^{-1} = x^b$ where $2b \equiv 3 \pmod{n}$
37
38 Let me compute $xyxy^{-1}x^{-1}$ in two ways:
39
40 First way: $xyxy^{-1}x^{-1} = x(yxy^{-1})x^{-1} = xx^bx^{-1} = x^b$
41
```

```
Second way: $xyxy^{-1}x^{-1} = (xyx^{-1})(yy^{-1}) = y^a \cdot 1 = y^a$

Therefore: $x^b = y^a$

This is a key relationship. Now, since $x^b = y^a$:
- Applying $\phi_y$: $y(x^b)y^{-1} = y(y^a)y^{-1}$, which gives $(x^b)^b = y^a$, so $x^{b^2} = y^a$
- But also $x^{b^2} = (x^b)^b = (y^a)^b = y^{ab}$

So $y^a = y^{ab}$, which means $y^{a(b-1)} = 1$.

Since $y$ has order $m$, we need $m | a(b-1)$.

But we also know:
- From $x^b = y^a$, raising to the $a$-th power: $x^{ab} = y^{a^2}$
- Applying $\phi_y$ to this: $y(x^{ab})y^{-1} = y(y^{a^2})y^{-1}$
- This gives $(x^b)^{ab} = y^{a^2}$, so $x^{ab^2} = y^{a^2}$
- But we already had $x^{ab} = y^{a^2}$

So $x^{ab} = x^{ab^2}$, which means $x^{ab(b-1)} = 1$.

Since $x$ has order $n$, we need $n | ab(b-1)$.

Now, since $m$ and $n$ are odd, and $a$ and $b$ are odd (as they satisfy $2a \equiv 3 \pmod{m}$ and $2b \equiv 3 \pmod{n}$ with odd $m,n$), we have:
- $\gcd(m, ab) = 1$ (since $m$ is odd and $a,b$ are odd)
- Therefore $m | (b-1)$ (since $b-1$ is even and $m | a(b-1)$)

But wait - this gives us $m | (b-1)$ where $b$ satisfies $2b \equiv 3 \pmod{n}$.

Similarly, we get $n | (a-1)$ where $a$ satisfies $2a \equiv 3 \pmod{m}$.

Let me check if any odd integer $m > 1$ can satisfy these conditions:

For $m = 3$: $a \equiv 3 \cdot 2^{-1} \equiv 3 \cdot 2 \equiv 0 \pmod{3}$
So $a-1 \equiv -1 \equiv 2 \pmod{3}$
But we need $3 | (a-1)$, which means $a-1 \equiv 0 \pmod{3}$. Contradiction!

For $m = 5$: $a \equiv 3 \cdot 3 \equiv 4 \pmod{5}$ (since $2^{-1} \equiv 3 \pmod{5}$)
So $a-1 \equiv 3 \pmod{5}$
But we need $5 | (a-1)$, which means $a-1 \equiv 0 \pmod{5}$. Contradiction!

By similar analysis, no odd $m > 1$ can satisfy the required conditions.

Therefore $m = 1$, which means $y = 1$.

By symmetry, $n = 1$, which means $x = 1$.

Therefore $G = \{1\}$ is the trivial group.
```

D.9. Grok 4 0709

`Note`: Grok 4 repeatedly returned empty output in OpenRouter. Thus Grok 4 0709 was accessed via LMArena. It is unclear if the two versions are identical.

`Line 5`: Nothing of the sort is established. (Critical)

`Line 7`: No reason is given for this. (Critical)

```
1  The group \(G\) is presented as \(\langle x, y \mid x y^2 = y^3 x, \, y x^2 =
   ↪  x^3 y \rangle\).
2
3  Assume for contradiction that \(G\) is non-trivial. Then at least one of \(x\)
   ↪  or \(y\) is non-trivial.
4
5  As established, the orders \(n\) of \(y\) and \(m\) of \(x\) must be finite and
   ↪  odd.
6
7  The relations imply that conjugation by \(x\) induces an automorphism on
   ↪  \(\langle y \rangle\) multiplying exponents by \(k \equiv 3 \cdot 2^{-1}
   ↪  \pmod{n}\), with \(\gcd(k, n) = 1\).
8
9  Similarly, conjugation by \(y\) induces an automorphism on \(\langle x \rangle\)
   ↪  multiplying exponents by \(l \equiv 3 \cdot 2^{-1} \pmod{m}\), with
   ↪  \(\gcd(l, m) = 1\).
10
11 The order \(e\) of \(k\) modulo \(n\) divides \(m\) and \(l - 1\).
12
13 The order \(d\) of \(l\) modulo \(m\) divides \(n\) and \(k - 1\).
14
15 Now, \(k - 1 \equiv 3 \cdot 2^{-1} - 1 \pmod{n}\). Suppose an odd prime \(p\)
   ↪  divides both \(n\) and \(k - 1\). Then \(3 \cdot 2^{-1} \equiv 1 \pmod{p}\),
   ↪  so \(3 \equiv 2 \pmod{p}\), hence \(p \mid 1\), impossible.
16
17 Thus, \(\gcd(n, k - 1) = 1\), so \(d = 1\).
18
19 Then \(l \equiv 1 \pmod{m}\), so \(2 \cdot 1 \equiv 3 \pmod{m}\), hence \(m \mid
   ↪  1\), so \(m = 1\) and \(x = 1\).
20
21 With \(x = 1\), the first relation becomes \(y^2 = y^3\), so \(y = 1\), hence
   ↪  \(n = 1\).
22
23 This contradicts \(G\) non-trivial.
24
25 Therefore, \(G\) is trivial.
26
27 \boxed{G is trivial}
28
```

D.10. Kimi K2

`Note`: The fact that at the end of the output generated by Kimi K2 a `$$` is missing is not a typo. The model terminates its output without claiming to have found a solution. The error below occurs well before that time.

`Line 184`: The identity

$$y^3 = x^2y^4x^{-2}$$

does not follow by conjugating again by x, and in fact, does not follow from the previous line at all. (Critical)

```
### Understanding the Problem

We have a group $$ G $$ defined by two generators $$ x $$ and $$ y $$ with two relations:
1. $$ xy^2 = y^3x $$
2. $$ yx^2 = x^3y $$

Our goal is to prove that $$ G $$ is the trivial group, meaning $$ x = 1 $$ and $$ y = 1 $$ in $$ G $$, so $$ G $$ has only one element.

### Initial Approach

To show that $$ G $$ is trivial, we need to derive from the given relations that $$ x = 1 $$ and $$ y = 1 $$. This means we should manipulate the given equations to find simpler relationships or equalities that force $$ x $$ and $$ y $$ to be the identity.

### Manipulating the First Relation: $$ xy^2 = y^3x $$

Let's start with the first relation:
$$ xy^2 = y^3x $$

We can multiply both sides on the right by $$ x^{-1} $$ (the inverse of $$ x $$) to isolate terms:
$$ xy^2x^{-1} = y^3 $$

This shows that $$ y^3 $$ is equal to $$ x y^2 x^{-1} $$, which is the conjugation of $$ y^2 $$ by $$ x $$.

### Manipulating the Second Relation: $$ yx^2 = x^3y $$

Similarly, take the second relation:
$$ yx^2 = x^3y $$

Multiply both sides on the right by $$ y^{-1} $$:
$$ yx^2y^{-1} = x^3 $$

Here, $$ x^3 $$ is the conjugation of $$ x^2 $$ by $$ y $$.

### Exploring Conjugation and Powers

From the first manipulation, we have:
$$ x y^2 x^{-1} = y^3 $$
```

```

This implies that $$ x $$ conjugates $$ y^2 $$ to $$ y^3 $$. Similarly, from the second manipulation:
$$ y x^2 y^{-1} = x^3 $$

So $$ y $$ conjugates $$ x^2 $$ to $$ x^3 $$.

### Trying to Find Simpler Relations

Let's see if we can find a relationship between $$ x $$ and $$ y $$ that leads to $$ x = 1 $$ or $$ y = 1 $$.

From $$ x y^2 x^{-1} = y^3 $$, we can raise both sides to the power of $$ n $$ (for any integer $$ n $$):
$$ (x y^2 x^{-1})^n = y^{3n} $$
$$ x y^{2n} x^{-1} = y^{3n} $$

Similarly, from $$ y x^2 y^{-1} = x^3 $$:
$$ (y x^2 y^{-1})^n = x^{3n} $$
$$ y x^{2n} y^{-1} = x^{3n} $$

These show how conjugation by $$ x $$ and $$ y $$ affects powers of $$ y $$ and $$ x $$, respectively.

### Attempting to Find a Fixed Point

Let's consider the exponent of $$ y $$ in the first relation. The exponents on both sides of $$ x y^2 x^{-1} = y^3 $$ don't match directly, which is unusual. This suggests that the only way this can hold is if $$ y = 1 $$, making both sides equal to 1.

Similarly, if $$ y = 1 $$, then plugging into the second relation:
$$ 1 \cdot x^2 \cdot 1 = x^3 \cdot 1 $$
$$ x^2 = x^3 $$
$$ x = 1 $$

By symmetry, if we assume $$ x = 1 $$, then:
First relation:
$$ 1 \cdot y^2 = y^3 \cdot 1 $$
$$ y^2 = y^3 $$
$$ y = 1 $$

This suggests that $$ x = y = 1 $$ is a solution that satisfies both relations, making $$ G $$ trivial. But we need to ensure that this is the only possible solution within the group $$ G $$.

### Proving Uniqueness of Trivial Solution

Let's formalize this without assuming $$ x $$ or $$ y $$ is the identity upfront.

```

```
From $$ x y^2 x^{-1} = y^3 $$, we can consider the exponent as a kind of homomorphism. Specifically, the mapping that sends $$ y $$ to $$ y^3 $$ when conjugated by $$ x $$. Similarly, the other relation gives a mapping sending $$ x $$ to $$ x^3 $$ under conjugation by $$ y $$.

This is reminiscent of the concept where in a group, if conjugation by an element raises powers to higher exponents, the group might be forced to be trivial.

Let's try to express everything in terms of a single element. Suppose we express one element in terms of the other.

From $$ x y^2 x^{-1} = y^3 $$, we can write:
$$ x y^2 = y^3 x $$

Similarly, from $$ y x^2 y^{-1} = x^3 $$:
$$ y x^2 = x^3 y $$

Now, let's try to find a word that must equal the identity.

Compute $$ x y x^{-1} y^{-1} $$, but this might not directly help. Instead, let's try to find a power that collapses to identity.

### Using Commutators

Notice that the relations resemble "expansive" behaviour-conjugation increases exponents. In some contexts, this implies the group must be trivial unless the elements commute in a certain way.

Let's attempt to compute some commutators or express higher powers.

From $$ x y^2 x^{-1} = y^3 $$, we can apply this multiple times:

First, raise both sides to the power of $$ k $$:
$$ (x y^2 x^{-1})^k = y^{3k} $$
$$ x y^{2k} x^{-1} = y^{3k} $$

Now, set $$ k = 0 $$, but that gives $$ 1 = 1 $$, which is trivial.

Alternatively, express $$ y $$ in terms of itself:
$$ x y^2 x^{-1} = y^3 $$
$$ x y x^{-1} x y x^{-1} = y^3 $$
$$ (x y x^{-1})^2 = y^3 $$

Similarly, from $$ y x^2 y^{-1} = x^3 $$:
$$ (y x y^{-1})^2 = x^3 $$

This seems to imply that the conjugates have certain power relationships.

### Trying Specific Elements and Equalities

```

```
Let's consider the order of $$ x $$ and $$ y $$. Suppose $$ x $$ has finite order $$ n $$, then $$ x^n = 1 $$.

From $$ y x^2 y^{-1} = x^3 $$, we have:
$$ y x^2 y^{-1} = x^3 $$
$$ y x^4 y^{-1} = x^6 $$
$$ y x^{2k} y^{-1} = x^{3k} $$

For $$ k = n $$, since $$ x^n = 1 $$, we have:
$$ y 1 y^{-1} = 1 = x^{3n} $$
But $$ x^{3n} = (x^n)^3 = 1^3 = 1 $$, which doesn't give new information.

Similarly, from $$ x y^2 x^{-1} = y^3 $$, we have:
$$ x y^{2k} x^{-1} = y^{3k} $$

If $$ y $$ has finite order $$ m $$, then for $$ k = m $$:
$$ x 1 x^{-1} = 1 = y^{3m} $$
$$ y^{3m} = 1 $$

But since the order of $$ y $$ is $$ m $$, $$ m $$ divides $$ 3m $$, which is always true, so no contradiction arises directly from this.

### Introducing a Normal Subgroup Approach

Perhaps we can consider that these relations imply all elements are trivial. Let's try to prove that $$ x $$ and $$ y $$ commute, and then see if this forces them to be trivial.

Assume $$ xy = yx $$. Then from $$ xy^2 = y^3x $$:
$$ xy^2 = y^3x $$
$$ y^2x = y^3x $$ (since $$ x $$ and $$ y $$ commute)
Multiply both sides on the right by $$ x^{-1} y^{-2} $$:
$$ 1 = y $$

Similarly, from $$ yx^2 = x^3y $$:
$$ yx^2 = x^3y $$
$$ x^2y = x^3y $$
$$ x^2 = x^3 $$
$$ x = 1 $$

Thus, if $$ x $$ and $$ y $$ commute, both must be identity. But we don't know if they commute.

### Using the Composition of Conjugations

We have:
$$ \phi_x: y \mapsto x y x^{-1} $$
with $$ \phi_x(y^2) = y^3 $$, so $$ (\phi_x(y))^2 = y^3 $$, implying $$ \phi_x(y) $$ is a square root of $$ y^3 $$, i.e., $$ x y x^{-1} = y^{3/2} $$, but this suggests fractional exponents, which are typically not in groups.
```

```

This seems problematic unless $$ y = 1 $$.

Let me rephrase: from $$ x y^2 x^{-1} = y^3 $$, we can define $$ z = x y x^{-1} $$, then:
$$ z^2 = y^3 $$

Similarly, from $$ y x^2 y^{-1} = x^3 $$, let $$ w = y x y^{-1} $$, then:
$$ w^2 = x^3 $$

Now we have:
$$ x y x^{-1} = z $$ where $$ z^2 = y^3 $$
$$ y x y^{-1} = w $$ where $$ w^2 = x^3 $$

This seem to imply that $$ z $$ is some root of $$ y^3 $$ and $$ w $$ of $$ x^3 $$, but in groups, this often forces elements to be identity.

### Deriving Contradictions or Identities

Let's try to express everything in terms of $$ x $$ and its conjugates.

From $$ x y^2 x^{-1} = y^3 $$, we have:
$$ y^3 = x y^2 x^{-1} $$
$$ y^3 = x^2 y^4 x^{-2} $$ (conjugating by $$ x $$ again)
$$ y^3 = x^2 y^4 x^{-2} $$

But also:
$$ y^3 = x y^2 x^{-1} $$
So:
$$ x y^2 x^{-1} = x^2 y^4 x^{-2} $$
$$ y^2 = x y^4 x^{-1} $$
But $$ x y^4 x^{-1} = (x y^2 x^{-1})^2 = (y^3)^2 = y^6 $$, so:
$$ y^2 = y^6 $$
$$ y^4 = 1 $$

Similarly, from $$ y x^2 y^{-1} = x^3 $$, applying $$ y $$ twice:
$$ y^2 x^4 y^{-2} = x^9 $$
But also:
$$ x^3 = y x^2 y^{-1} $$
So:
$$ y x^2 y^{-1} = y^2 x^4 y^{-2} $$
$$ x^2 = y x^4 y^{-1} $$
$$ x^2 = (x^3)^2 = x^6 $$
$$ x^4 = 1 $$

Now we have $$ x^4 = 1
```

D.11. Qwen3 235B A22B Thinking 2507

`Line 25`: The model has assumed commutativity. (Critical)

```
1  To determine whether the group $ G = \langle x, y \mid xy^2 = y^3x,\ yx^2 = x^3y
   ↪  \rangle $ is trivial, we analyze the given relations and manipulate them
   ↪  using group-theoretic techniques.
2
3  ---
4
5  ### Step 1: Expressing Generators in Terms of a Single Element
6
7  Define the **commutator** $ a = [x, y] = x y x^{-1} y^{-1} $. Then by
   ↪  definition:
8
9  $$
10 x y = a y x.
11 $$
12
13 Now apply this to **Relation 1**:
14 $$
15 xy^2 = y^3x.
16 $$
17
18 Left-hand side:
19 $$
20 x y^2 = x y \cdot y = (a y x) \cdot y = a y x y.
21 $$
22
23 Using $ x y = a y x $ again inside:
24 $$
25 a y x y = a y (a y x) = a^2 y^2 x.
26 $$
27
28 So:
29 $$
30 x y^2 = a^2 y^2 x.
31 $$
32
33 Equating with the right-hand side $ y^3 x $, we get:
34 $$
35 a^2 y^2 x = y^3 x \Rightarrow a^2 y^2 = y^3 \Rightarrow a^2 = y.
36 $$
37
38 Thus, we have:
39 $$
40 y = a^2. \tag{1}
41 $$
42
43 ---
44
45 ### Step 2: Using the Second Relation
```

```markdown

Now consider **Relation 2**:
$$
yx^2 = x^3 y.
$$

We define the inverse commutator $ b = [y, x] = y x y^{-1} x^{-1} = a^{-1} $, so:
$$
y x = b x y = a^{-1} x y.
$$

Apply this to Relation 2:
$$
y x^2 = y x \cdot x = (a^{-1} x y) \cdot x = a^{-1} x y x.
$$

Then:
$$
y x^2 = a^{-1} x (a^{-1} x y) = a^{-2} x^2 y.
$$

Equating with RHS $ x^3 y $, we get:
$$
a^{-2} x^2 y = x^3 y \Rightarrow a^{-2} x^2 = x^3 \Rightarrow x = a^{-2}. \tag{2}
$$

---

### Step 3: Substituting Back into Relations

From (1) and (2), we now express both generators in terms of $ a $:
$$
x = a^{-2}, \quad y = a^2.
$$

Substitute these into **Relation 1**:
$$
xy^2 = a^{-2}(a^2)^2 = a^{-2} a^4 = a^2,
$$
$$
y^3 x = (a^2)^3 a^{-2} = a^6 a^{-2} = a^4.
$$

So:
$$
a^2 = a^4 \Rightarrow a^2 = e.
$$

Therefore, $ a = e $, which implies:
```

```
$$
x = a^{-2} = e, \quad y = a^2 = e.
$$

---

### Step 4: Verifying Triviality

With $ x = e $ and $ y = e $, all relations are trivially satisfied:
- $ xy^2 = e \cdot e = e $, and $ y^3x = e \cdot e = e $,
- $ yx^2 = e \cdot e = e $, and $ x^3y = e \cdot e = e $.

Hence, the only solution is the trivial group.

---

### Conclusion

By expressing the generators in terms of a single element derived from the commutator and analyzing the resulting constraints, we find that the relations force both $ x $ and $ y $ to be the identity element. Therefore, the group defined by the presentation:

$$
G = \langle x, y \mid xy^2 = y^3x,\ yx^2 = x^3y \rangle
$$

is **trivial**.

$$
\boxed{G \text{ is the trivial group}}
$$

```

D.12. GLM-4.5

`Note`: The generation began in the middle of the sentence, our reproduction below is not a typo.

`Line 19`: This identity is incorrect. The model has made an algebra error from the previous line. (Critical)

```
1 that the group $G = \langle x, y \mid xy^2 = y^3x, yx^2 = x^3y \rangle$ is the trivial group, I'll derive that both generators $x$ and $y$ must be the identity element.
2
3 Starting with the given relations:
4 1. $xy^2 = y^3x$
5 2. $yx^2 = x^3y$
6
7 From relation (1), multiply both sides by $y^{-2}$ on the left:
8 $y^{-2}xy^2 = yx$  (3)
9
```

```
From relation (2), multiply both sides by $x^{-2}$ on the left:
$x^{-2}yx^2 = xy$  (4)

Let $A = y^{-1}xy$ and $B = x^{-1}yx$. Then:
- From (3): $y^{-1}Ay = yx$
- From (4): $x^{-1}Bx = xy$

Since $A = y^{-1}xy$, we have $y^{-1}Ay = y^{-2}xy^2$. From equation (3), $y^{-2}xy^2 = yx$, so:
$y^{-1}Ay = yx$
which implies $A = y^2x$.

Similarly, since $B = x^{-1}yx$, we have $x^{-1}Bx = x^{-2}yx^2$. From equation (4), $x^{-2}yx^2 = xy$, so:
$x^{-1}Bx = xy$
which implies $B = x^2y$.

Now we have:
$A = y^{-1}xy = y^2x$
$B = x^{-1}yx = x^2y$

From $yA = xy$ and $A = y^2x$, we get:
$y(y^2x) = xy$
$y^3x = xy$

From the original relation (1), $xy^2 = y^3x$, so:
$xy^2 = xy$

Now, multiply both sides of $xy^2 = xy$ by $x^{-1}$ on the left:
$y^2 = y$

Multiply both sides of $y^2 = y$ by $y^{-1}$ on the right:
$y = e$

Since $y$ is the identity element, we can substitute $y = e$ into relation (1):
$xe^2 = e^3x$
$x = x$

This is a tautology and doesn't give us new information. Let's substitute $y = e$ into relation (2):
$ex^2 = x^3e$
$x^2 = x^3$

Multiply both sides of $x^2 = x^3$ by $x^{-2}$ on the right:
$e = x$

Therefore, both $x$ and $y$ are the identity element, which means $G$ is the trivial group.
```

D.13. Gemini 2.5 Deep Think

`Line 18`: The model claims $yx^3 = (yx)x^3$, which is an algebra error. (Critical)

```
1  Yes, we can prove that G is the trivial group.
2
3  Let $G$ be a group with generators $x$ and $y$ and the relations:
4  1) $xy^2 = y^3x$
5  2) $yx^2 = x^3y$
6
7  We want to show that $x=e$ and $y=e$, where $e$ is the identity element.
8
9  Let's start by manipulating the second relation (2). We can rewrite it as:
10 $x^3y = yx^2$ (2')
11
12 Now, we examine how the element $x^3$ interacts with the element $yx$. We look
   ↪  at the product $x^3(yx)$:
13 $x^3(yx) = (x^3y)x$ (by associativity)
14
15 We use the rearranged relation (2') to substitute $x^3y$ with $yx^2$:
16 $(x^3y)x = (yx^2)x = y(x^2x) = yx^3$
17
18 So we have found that $x^3(yx) = yx^3$. Since $yx^3 = (yx)x^3$, this means that
   ↪  $x^3$ commutes with $yx$. In terms of commutators, we have $[x^3, yx] = e$.
19
20 Next, let's analyze the commutator $[x^3, yx]$. Recall that the commutator $[A,
   ↪  B]$ is defined as $ABA^{-1}B^{-1}$.
21 $[x^3, yx] = x^3 (yx) (x^3)^{-1} (yx)^{-1}$
22 $= x^3 y x x^{-3} (x^{-1}y^{-1})$
23 $= x^3 y (x^{1-3-1}) y^{-1}$
24 $= x^3 y x^{-3} y^{-1}$
25
26 This expression is exactly the definition of the commutator of $x^3$ and $y$,
   ↪  i.e., $[x^3, y]$.
27
28 Since we established that $[x^3, yx] = e$ and $[x^3, yx] = [x^3, y]$, we must
   ↪  have:
29 $[x^3, y] = e$.
30
31 This means that $x^3$ commutes with $y$:
32 $x^3y = yx^3$.
33
34 Now we have two equations involving $x^3y$:
35 From relation (2'): $x^3y = yx^2$
36 From our derivation: $x^3y = yx^3$
37
38 We can equate the right-hand sides:
39 $yx^2 = yx^3$
40
41 We multiply by $y^{-1}$ on the left:
42 $x^2 = x^3$
43
```

```
We multiply by $x^{-2}$ (on either side):
$e = x$.

Now that we have shown $x=e$, we substitute this back into the first relation
 ↪  (1):
$xy^2 = y^3x$
$e y^2 = y^3 e$
$y^2 = y^3$

We multiply by $y^{-2}$:
$e = y$.

Since both generators $x$ and $y$ are equal to the identity element, the group
 ↪  $G$ is the trivial group.


```

D.14. Llama 4 Maverick

`Lines 4-8`: A lot of effort is expended proving $y^3x = y^3x$.

`Line 11`: The model claims $x(y^2)x^{-1}(y^2)^{-1} = e$ but never proves this. (Critical)

```
1  ## Step 1: Understand the given problem and the relations provided for the group $G$.
2  The group $G$ is generated by $x$ and $y$ with the relations $xy^2 = y^3x$ and $yx^2 = x^3y$. We need to determine if $G$ is the trivial group, meaning it contains only the identity element.
3
4  ## Step 2: Manipulate the given relations to simplify or derive useful equations.
5  Starting with $xy^2 = y^3x$, we can rearrange this to $xy^2x^{-1} = y^3$, implying that $x$ conjugates $y^2$ to $y^3$. Similarly, from $yx^2 = x^3y$, we get $yxy^{-1} = x^3x^{-2} = x$, showing that $y$ conjugates $x$ to $x$.
6
7  ## Step 3: Further simplify the relations to check for any patterns or simplifications that could lead to proving $G$ is trivial.
8  From $xy^2 = y^3x$, we can also derive $x^{-1}y^3x = y^2$. This implies $x^{-1}y^3x = y^2$ and thus $y^3 = xy^2x^{-1}$. Substituting $y^2 = x^{-1}y^3x$ into the original equation yields $x(x^{-1}y^3x) = y^3x$, simplifying to $y^3x = y^3x$, which is always true and doesn't provide new information directly.
9
10 ## Step 4: Analyze the relation $xy^2 = y^3x$ further to derive a condition on $y$.
11 Rearranging $xy^2 = y^3x$ gives $xy^2x^{-1}y^{-3} = e$, where $e$ is the identity. This can be written as $xy^2x^{-1}y^{-2}y^{-1} = e$, or $x(y^2)x^{-1}(y^2)^{-1}y^{-1} = e$, simplifying to $y^{-1} = e$ because $x(y^2)x^{-1}(y^2)^{-1} = e$ implies $y^2 = e$ after simplification, but directly we see $y = e$ from $y^{-1} = e$.
12
13 ## Step 5: Examine the implication of $y = e$ on the group $G$ and its generators.
14 If $y = e$, then substituting into one of the original relations, $xy^2 = y^3x$ becomes $x = x$, which is true for any $x$. However, we need to check the other relation as well. $yx^2 = x^3y$ becomes $ex^2 = x^3e$, simplifying to $x^2 = x^3$. This implies $x^3x^{-2} = e$, or $x = e$.
15
16 ## Step 6: Verify the result by checking if both generators being the identity implies $G$ is trivial.
17 Since both $x = e$ and $y = e$, and $G$ is generated by $x$ and $y$, $G$ must be the trivial group because it is generated by the identity.
18
19 The final answer is: $\boxed{Yes}$
20
```

D.15. DeepSeek v3 0324

`Line 46`: The order is only a positive integer if it is finite. The model has ignored the infinite case. However, it recovers from this on line 62.

`Line 82`: The model worries that $x^3 = x^2$ can only be simplified to $x = e$ if x is cancellable. But this is an unfounded concern, as we are working in a group.

`Line 194`: The model makes no progress with the problem, other than showing $|x|$ and $|y|$ are coprime to 6, but then concludes the group is trivial. No part of the proof comes close to showing this. (Critical)

```
1  Alright, let's tackle this problem step by step. We have a group $$ G $$ generated by two elements $$ x $$ and $$ y $$ with the following relations:
2
3  1. $$ xy^2 = y^3x $$
4  2. $$ yx^2 = x^3y $$
5
6  Our goal is to show that $$ G $$ is the trivial group, meaning that $$ x $$ and $$ y $$ are both the identity element, and thus $$ G = \{e\} $$.
7
8  ### Understanding the Problem
9
10 First, let's recall what these relations mean. In a group presentation, relations are equations that hold true among the generators. Here, the two relations give us ways to manipulate expressions involving $$ x $$ and $$ y $$.
11
12 Our strategy will be to derive consequences from these relations, hoping to find that $$ x $$ and $$ y $$ must be the identity.
13
14 ### Starting with the First Relation: $$ xy^2 = y^3x $$
15
16 Let's see what this tells us. If we multiply both sides on the left by $$ x^{-1} $$, we get:
17
18 $$ x^{-1}xy^2 = x^{-1}y^3x $$
19
20 Which simplifies to:
21
22 $$ y^2 = x^{-1}y^3x $$
23
24 This can be rewritten as:
25
26 $$ x y^2 x^{-1} = y^3 \tag{Equation 1}$$
27
28 This suggests that conjugating $$ y^2 $$ by $$ x $$ gives $$ y^3 $$.
29
30 Similarly, from the second relation $$ yx^2 = x^3y $$, let's multiply both sides on the left by $$ y^{-1} $$:
31
32 $$ y^{-1}yx^2 = y^{-1}x^3y $$
33
34 Simplifying:
```

```

$$ x^2 = y^{-1}x^3y $$

Which can be rewritten as:

$$ y x^2 y^{-1} = x^3 \tag{Equation 2}$$

This suggests that conjugating $$ x^2 $$ by $$ y $$ gives $$ x^3 $$.

### Exploring Equation 1: $$ x y^2 x^{-1} = y^3 $$

This equation tells us that $$ y^2 $$ and $$ y^3 $$ are conjugates, implying that they have the same order in the group. The order of an element $$ g $$, denoted $$ |g| $$, is the smallest positive integer $$ n $$ such that $$ g^n = e $$.

From $$ y^2 $$ and $$ y^3 $$ being conjugates:

$$ |y^2| = |y^3| $$

We know that in general, $$ |g^k| = \frac{|g|}{\gcd(|g|, k)} $$. So:

$$ \frac{|y|}{\gcd(|y|, 2)} = \frac{|y|}{\gcd(|y|, 3)} $$

This implies:

$$ \gcd(|y|, 2) = \gcd(|y|, 3) $$

The only way this can happen is if $$ \gcd(|y|, 6) = 1 $$, meaning that $$ |y| $$ is coprime with 6. The simplest possibility is $$ |y| = 1 $$, i.e., $$ y = e $$.

But let's not jump to conclusions; maybe $$ |y| $$ is infinite or another number coprime to 6. Let's explore further.

### Exploring Equation 2: $$ y x^2 y^{-1} = x^3 $$

Similarly, this tells us that $$ x^2 $$ and $$ x^3 $$ are conjugates, so:

$$ |x^2| = |x^3| $$

Again, using the order formula:

$$ \frac{|x|}{\gcd(|x|, 2)} = \frac{|x|}{\gcd(|x|, 3)} $$

Thus:

$$ \gcd(|x|, 2) = \gcd(|x|, 3) $$

Similarly, this implies $$ \gcd(|x|, 6) = 1 $$, so $$ |x| $$ is coprime with 6. Again, the simplest solution is $$ |x| = 1 $$, i.e., $$ x = e $$.
```

```markdown

### Considering $$ x $$ and $$ y $$ are the Identity

If $$ y = e $$, then the first relation $$ xy^2 = y^3x $$ becomes $$ x = x $$, which holds trivially. The second relation $$ yx^2 = x^3y $$ becomes $$ x^2 = x^3 $$, which implies $$ x = e $$ (assuming $$ x $$ has finite order or is cancellable).

Similarly, if $$ x = e $$, the second relation becomes $$ y = y $$, and the first relation becomes $$ y^2 = y^3 $$, implying $$ y = e $$.

This suggests that the only solution is $$ x = y = e $$.

### Trying to Find Non-Trivial Solutions

But let's not assume the group is finite yet. Maybe the group is infinite, and $$ x $$ and $$ y $$ have infinite order. Let's see if we can find any non-trivial relations.

From $$ x y^2 x^{-1} = y^3 $$, we can think of this as an action of $$ x $$ on $$ y $$. In general, this could suggest a semidirect product, but the relations seem too restrictive.

Similarly, $$ y x^2 y^{-1} = x^3 $$ is an action of $$ y $$ on $$ x $$.

Perhaps we can find a consistency condition by combining these.

Let's see if we can express both $$ x $$ and $$ y $$ in terms of one another.

From $$ x y^2 = y^3 x $$, we can try to solve for $$ y $$ in terms of $$ x $$, but it's not straightforward.

Alternatively, let's see if we can find that $$ x $$ and $$ y $$ commute.

Assume for a moment that $$ x $$ and $$ y $$ commute, i.e., $$ xy = yx $$. Then:

From $$ xy^2 = y^3x $$, with commutativity:

$$ y^2x = y^3x $$

Cancel $$ x $$ on the right (assuming $$ x $$ is cancellable, which it is in a group):

$$ y^2 = y^3 $$

Thus $$ y = e $$.

Similarly, from $$ yx^2 = x^3y $$, with commutativity:

$$ x^2y = x^3y $$

```

```
Cancel $$ y $$:

$$ x^2 = x^3 $$

Thus $$ x = e $$.

So if $$ x $$ and $$ y $$ commute, they must both be the identity.

But we haven't shown that $$ x $$ and $$ y $$ must commute. Maybe the relations force them to commute.

Let's attempt to see if $$ xy = yx $$ follows from the given relations.

From $$ xy^2 = y^3x $$, let's try to express $$ xy $$:

Multiply both sides by $$ y^{-1} $$ on the right:

$$ xy = y^3x y^{-1} $$

Similarly, from $$ yx^2 = x^3y $$, multiply by $$ x^{-1} $$ on the right:

$$ yx = x^3y x^{-1} $$

This seems circular. Maybe another approach is better.

### Attempting to Find Orders

Let me think about the orders of $$ x $$ and $$ y $$. From earlier, $$ |y^2| = |y^3| $$, which implies $$ \gcd(|y|, 2) = \gcd(|y|, 3) $$.

Possible orders for $$ y $$:

If $$ |y| $$ is finite, then $$ \gcd(|y|, 2) = \gcd(|y|, 3) $$. The numbers where this holds are those not divisible by 2 or 3 (or both), i.e., numbers coprime to 6: 1, 5, 7, 11, etc.

Similarly, $$ |x| $$ must be coprime to 6.

The simplest case is $$ |x| = |y| = 1 $$, i.e., trivial group.

If $$ |y| > 1 $$, say $$ |y| = 5 $$, then $$ y^2 $$ has order $$ 5 / \gcd(5,2) = 5 $$, and $$ y^3 $$ has order $$ 5 / \gcd(5,3) = 5 $$, so the equation holds. But does this lead to a contradiction elsewhere?

From $$ x y^2 x^{-1} = y^3 $$, if $$ y $$ has order 5, then $$ y^3 = y^{-2} $$, so:

$$ x y^2 x^{-1} = y^{-2} $$

But $$ y^2 $$ has order 5, and its inverse also has order 5, so this is possible, but doesn't immediately force $$ y = e $$.
```

```

Similarly, for $$ x $$, if $$ |x| = 5 $$, from $$ y x^2 y^{-1} = x^3 $$, and $$ x^3 = x^{-2} $$, so:

$$ y x^2 y^{-1} = x^{-2} $$

Again, no immediate contradiction.

However, the group seems too constrained to allow non-trivial elements, but I'm not seeing an immediate contradiction for higher orders coprimes to 6.

### Trying Specific Cases

Perhaps assume $$ y = e $$, then see what $$ x $$ must be.

If $$ y = e $$, then relations become:

1. $$ x = x $$ (holds)
2. $$ x^2 = x^3 \implies x = e $$

So $$ x = e $$, $$ y = e $$. Trivial group.

Similarly, if $$ x = e $$, then relations become:

1. $$ y^2 = y^3 \implies y = e $$
2. $$ y = y $$ (holds)

Again, trivial group.

This suggests that the only solution is the trivial group.

### Conclusion

After exploring these relations, it appears that the only group $$ G $$ satisfying both $$ xy^2 = y^3x $$ and $$ yx^2 = x^3y $$ is the trivial group where both generators $$ x $$ and $$ y $$ are the identity element. Any non-trivial assignment leads to contradictions or forces the generators to be the identity.

### Final Answer

After carefully analyzing the given relations, we conclude that the group $$ G $$ must be the trivial group. That is, $$ G = \{e\} $$, where $$ e $$ is the identity element, and $$ x = y = e $$.

### Verification

Let's verify:

1. $$ xy^2 = e \cdot e^2 = e $$ and $$ y^3x = e^3 \cdot e = e $$, so $$ xy^2 = y^3x $$.
```

```
2. $$ yx^2 = e \cdot e^2 = e $$ and $$ x^3y = e^3 \cdot e = e $$, so $$ yx^2 =
↪  x^3y $$.

Both relations hold, and no smaller group satisfies them unless both generators
↪  are the identity. Hence, $$ G $$ is indeed the trivial group.

```

D.16. QwQ 32B

`Lines 7-9`: There is an algebra error in deriving $x^{-1}y^3xy^{-3} = e$. (Critical)

`Line 13`: Commutativity is incorrectly assumed. (Critical)

`Line 25`: Commutativity is incorrectly assumed. (Critical)

```
1  To prove that the group \( G \) generated by \( x \) and \( y \) with relations
   ↪  \( xy^2 = y^3x \) and \( yx^2 = x^3y \) is trivial, we start by analyzing
   ↪  the given relations.
2
3  First, consider the relation \( xy^2 = y^3x \). Rearranging this, we have:
4  \[
5  x^{-1} y^3 x = y^2
6  \]
7  Multiplying both sides on the right by \( y^{-3} \), we get:
8  \[
9  x^{-1} y^3 x y^{-3} = e
10 \]
11 This simplifies to:
12 \[
13 x^{-1} (y^3 y^{-3}) x = e \implies x^{-1} y^0 x = e \implies x^{-1} e x = e
14 \]
15 However, we need to re-express this correctly. Starting again:
16 \[
17 x y^2 x^{-1} = y^3
18 \]
19 Multiply both sides by \( y^{-3} \):
20 \[
21 x y^2 x^{-1} y^{-3} = e
22 \]
23 This can be rewritten as:
24 \[
25 x (y^{-1}) x^{-1} = e
26 \]
27 Multiplying both sides on the left by \( x^{-1} \) and on the right by \( x \),
   ↪  we get:
28 \[
29 y^{-1} = e \implies y = e
30 \]
31 Thus, \( y \) must be the identity element.
32
33 Substituting \( y = e \) into the second relation \( yx^2 = x^3y \), we get:
34 \[
35 e \cdot x^2 = x^3 \cdot e \implies x^2 = x^3
36 \]
37 This simplifies to:
38 \[
39 x^2 = x^3 \implies x^2 x^{-2} = x^3 x^{-2} \implies e = x
40 \]
41 Therefore, \( x \) must also be the identity element.
42
```

```
Since both \( x \) and \( y \) are the identity element, the group \( G \) is
↪  trivial.

\[
\boxed{1}
\]

```

D.17. GPT-OSS-120B

`Line 24`: The given conjugation relations only show that the subgroups generated by y^2 and x^2 are normal, not those generated by x and y. (Critical)

```
1  $$
2  G=\langle x ,y\mid xy^{2}=y^{3}x ,\; yx^{2}=x^{3}y\rangle .
3  $$
4
5  ---------------------------------------------------------------------
6  ### 1.  Conjugation relations
7
8  From the defining equalities we can rewrite
9
10 $$
11 \begin{aligned}
12 xy^{2}&=y^{3}x \quad\Longrightarrow\quad x\,y^{2}x^{-1}=y^{3},\$$2mm]
13 yx^{2}&=x^{3}y \quad\Longrightarrow\quad y\,x^{2}y^{-1}=x^{3}.
14 \end{aligned}
15 \tag{1}
16 $$
17
18 Thus
19
20 *conjugation by $$x$$ sends $$y^{2}$$ to $$y^{3}$$;*
21
22 *conjugation by $$y$$ sends $$x^{2}$$ to $$x^{3}$$.
23
24 Hence the subgroups $$\langle y\rangle$$ and $$\langle x\rangle$$ are normal,
   ↪  and the
25 conjugation maps restrict to automorphisms of these cyclic subgroups.
26
27 ---------------------------------------------------------------------
28 ### 2.  The automorphisms are given by a power map
29
30 Let
31
32 $$
33 x y x^{-1}=y^{\ell},\qquad y x y^{-1}=x^{k}
34 \tag{2}
35 $$
36
37 for some integers $$\ell ,k$$.
38 Because the restriction of conjugation to a cyclic group is an automorphism,
39 $$\gcd(\ell ,|y|)=\gcd(k ,|x|)=1$$ (if the orders are infinite the condition is
40 $$\ell ,k\neq0$$).
41
42 Apply the first equality of (2) to the element $$y^{2}$$:
43 $$
44 x y^{2}x^{-1} = (x y x^{-1})^{2}=y^{2\ell}=y^{3}.
45 $$
46 Hence
```

```

$$
2\ell\equiv 3\pmod{|y|}\qquad\text{(or }2\ell=3\text{ if }|y|=\infty).
\tag{3}
$$

Similarly, applying the second equality of (2) to $$x^{2}$$ gives

$$
y x^{2}y^{-1}= (y x y^{-1})^{2}=x^{2k}=x^{3},
$$
so

$$
2k\equiv 3\pmod{|x|}\qquad\text{(or }2k=3\text{ if }|x|=\infty).
\tag{4}
$$

Because the congruences (3) and (4) have a solution, the orders of $$x$$ and
$$y$$ cannot be even and cannot be multiples of $$3$$; in particular they are
odd (or infinite).  In any case the integers $$\ell ,k$$ are well defined and
satisfy $$\ell\not\equiv 0$$ and $$k\not\equiv 0$$.

---------------------------------------------------------------------
### 3.  $$x$$ and $$y$$ lie in the same cyclic subgroup

From (2) we compute the commutator:

$$
\begin{aligned}
x y x^{-1}y^{-1}
   &= (x y x^{-1})y^{-1}
    = y^{\ell}y^{-1}
    = y^{\ell-1},\$$2mm]
   &= x (y x y^{-1}) x^{-1}
    = x x^{k}x^{-1}
    = x^{k}.
\end{aligned}
$$

Thus

$$
x^{k}=y^{\ell-1}. \tag{5}
$$

Since $$k\not\equiv0$$ (mod $$|x|$$) the element $$x^{k}$$ is a non-trivial
↪ power
of $$x$$; (5) shows that this power is a power of $$y$$.  Consequently

$$
```

```
x\in\langle y\rangle .
$$

A symmetric computation (or simply interchanging the roles of $$x$$ and $$y$$)
gives

$$
y\in\langle x\rangle .
$$

Hence both generators belong to the same cyclic subgroup; the whole group
$$G$$ is cyclic.

---------------------------------------------------------------------
### 4.  A cyclic group satisfying the relations is trivial

Write the cyclic group as $$\langle g\rangle$$ and set

$$
x=g^{r},\qquad y=g^{s}.
$$

Insert these expressions in the defining relations:

$$
\begin{aligned}
xy^{2}=y^{3}x
   &\Longrightarrow g^{r+2s}=g^{3s+r},\\
yx^{2}=x^{3}y
   &\Longrightarrow g^{s+2r}=g^{3r+s}.
\end{aligned}
$$

In a cyclic group equality of elements means equality of the exponents
(modulo the order of $$g$$, or as integers if the order is infinite).  From the
first equality we obtain

$$
2s\equiv 3s\pmod{|g|}\;\;\Longrightarrow\;\;s\equiv0,
$$

and from the second equality

$$
2r\equiv 3r\pmod{|g|}\;\;\Longrightarrow\;\;r\equiv0.
$$

Thus $$r=s=0$$; consequently $$x=y=e$$.

---------------------------------------------------------------------
### 5.  Conclusion
```

```

The group defined by

$$
\langle x ,y\mid xy^{2}=y^{3}x ,\; yx^{2}=x^{3}y\rangle
$$

is cyclic, and the only cyclic group that satisfies the two relations is the
trivial group.  Hence

$$
\boxed{G\text{ is the trivial group}.}
$$
```

D.18. GPT-5 Thinking

`Lines 12, 20, 35`: The definition of $[s,t]$ used is $sts^{-1}t^{-1}$. The commutator identity given then holds only if $s^t = tst^{-1}$ but the commutator definition is later taken to be $s^t = t^{-1}st$. (Critical)

```
1  Yes. Write the relations in commutator form and use a short lemma.
2
3  From
4
5  $$
6  xy^2=y^3x\quad\text{and}\quad yx^2=x^3y
7  $$
8
9  we get
10
11 $$
12 x y^2 x^{-1}=y^3\ \Longrightarrow\ [x,y^2]=x y^2 x^{-1}y^{-2}=y,
13 $$
14
15 $$
16 y x^2 y^{-1}=x^3\ \Longrightarrow\ [y,x^2]=y x^2 y^{-1}x^{-2}=x.
17 $$
18
19 Let $u=[x,y]$. Using the standard commutator identity
20 $[a,bc]=[a,b]\,[a,c]^b$, we have
21
22 $$
23 [x,y^2]=[x,y]\,[x,y]^y=u\,u^y,\qquad
24 [y,x^2]=[y,x]\,[y,x]^x=u^{-1}(u^{-1})^x.
25 $$
26
27 Hence
28
29 $$
30 y=u\,u^y,\qquad x=u^{-1}(u^{-1})^x. \tag{$\ast$}
31 $$
32
33 **Lemma.** If $a=w\,w^a$ in a group, then $a=w^2$.
34
35 *Proof.* $a=w\,w^a\Rightarrow 1=w\,w^a a^{-1}=w(a^{-1}wa)a^{-1}=w a^{-1} w$, so
   ↪  $a^{-1}=w^{-2}$ and thus $a=w^{2}$. $\square$
36
37 Apply the lemma to $(a,w)=(y,u)$ and $(a,w)=(x,u^{-1})$ in $(\ast)$:
38
39 $$
40 y=u^2,\qquad x=(u^{-1})^2=u^{-2}.
41 $$
42
43 Thus $x=y^{-1}$.
44
45 Substitute $x=y^{-1}$ into $xy^2=y^3x$:
46
```

```
$$
y^{-1}y^2=y^3y^{-1}\ \Longrightarrow\ y=y^2 \ \Longrightarrow\ y=1,
$$

whence $x=1$ as well. Therefore $G$ is the trivial group. $\square$
```

D.19. GPT-5.5 Thinking

`Line 42`: The expression yy^4y^{-1} in the middle of the equation has been incorrectly expanded to y^5.

`Line 48`: Left multiplication by y and right multiplication by y^{-1} should not change y^9 to y^{10}.

```
1  Yes. Write the defining relations as
2
3  [
4  xy^2x^{-1}=y^3,\qquad yx^2y^{-1}=x^3.
5  ]
6
7  We will prove (y=1), and then (x=1) follows immediately.
8
9  First note that
10
11 [
12 x^n y^{2^n}x^{-n}=y^{3^n}
13 ]
14
15 for every (n\ge 1). In particular,
16
17 [
18 x^2y^4x^{-2}=y^9,\qquad x^3y^8x^{-3}=y^{27}.
19 ]
20
21 From the second defining relation,
22
23 [
24 yx^2y^{-1}=x^3,
25 ]
26
27 we get
28
29 [
30 x^2=y^{-1}x^3y.
31 ]
32
33 Substitute this into (x^2y^4x^{-2}=y^9):
34
35 [
36 (y^{-1}x^3y)y^4(y^{-1}x^{-3}y)=y^9.
37 ]
38
39 So
40
41 [
42 y^{-1}x^3y^5x^{-3}y=y^9,
43 ]
44
45 and therefore
46
```

```
[
x^3y^5x^{-3}=y^{10}.
]

Now compare two ways of conjugating (y^{40}) by (x^3). Since (40=5\cdot 8=8\cdot
↪  5),

[
x^3y^{40}x^{-3}
===============

# (x^3y^8x^{-3})^5

# (y^{27})^5

y^{135}.
]

But also

[
x^3y^{40}x^{-3}
===============

# (x^3y^5x^{-3})^8

# (y^{10})^8

y^{80}.
]

Thus

[
y^{135}=y^{80},
]

so

[
y^{55}=1.
]

By symmetry, the same argument gives

[
x^{55}=1.
]

Now (2^{-1}\equiv 28\pmod{55}). Since (y^{55}=1), the relation (xy^2x^{-1}=y^3)
↪  implies
```

```

[
xyx^{-1}
========

# x(y^2)^{28}x^{-1}

# (y^3)^{28}

# y^{84}

y^{29}.
]

Similarly,

[
yxy^{-1}=x^{29}.
]

Now conjugate the relation (xyx^{-1}=y^{29}) by (y). We get

[
(yxy^{-1})y(yx^{-1}y^{-1})=y^{29}.
]

Using (yxy^{-1}=x^{29}), this becomes

[
x^{29}yx^{-29}=y^{29}.
]

But conjugation by (x) sends (y) to (y^{29}), so conjugation by (x^{29}) sends
↪ (y) to

[
y^{29^{29}}.
]

Modulo (55),

[
29^{29}\equiv 19 \pmod{55}.
]

Hence

[
y^{19}=y^{29},
]

```

```
so

[
y^{10}=1.
]

Together with (y^{55}=1), this gives

[
y^5=1.
]

By symmetry,

[
x^5=1.
]

Now reduce the conjugation relations modulo (5). Since (2^{-1}\equiv 3\pmod 5),
↪  the original relation (xy^2x^{-1}=y^3) gives

[
xyx^{-1}=y^{3\cdot 3}=y^9=y^4.
]

Similarly,

[
yxy^{-1}=x^4.
]

Conjugate (xyx^{-1}=y^4) by (y):

[
(yxy^{-1})y(yx^{-1}y^{-1})=y^4.
]

Using (yxy^{-1}=x^4), this becomes

[
x^4yx^{-4}=y^4.
]

But conjugation by (x) sends (y) to (y^4), so conjugation by (x^4) sends (y) to

[
y^{4^4}=y^{256}=y
]

because (256\equiv 1\pmod 5). Therefore

```

```
[
y=y^4,
]

so

[
y^3=1.
]

Since also (y^5=1), we must have

[
y=1.
]

Finally, substituting (y=1) into

[
yx^2=x^3y
]

gives

[
x^2=x^3,
]

so (x=1). Therefore both generators are trivial, and hence

[
G={1}.
]

So (G) is the trivial group.
```

D.20. Gemini 3.1 Pro

`Line 27`: The expression $y^{-1}y^8y$ in the middle of the equation has been incorrectly expanded to y^7.

```
1  Here is the step-by-step mathematical proof that $G$ is the trivial group.
2
3  ### **1. Set Up the Relations**
4  We are given the two relations:
5  1. $xy^2 = y^3x$
6  2. $yx^2 = x^3y$
7
8  From these, we can isolate powers of $y$ and $x$ through conjugation:
9  * From the first relation: $xy^2x^{-1} = y^3$ and $x^{-1}y^3x = y^2$
10 * From the second relation: $yx^2y^{-1} = x^3$ and $yx^{-2}y^{-1} = x^{-3}$
11
12 ### **2. Expand $y^{27}$ Using the First Relation**
13 Using the identity $y^3 = xy^2x^{-1}$, we can systematically find expressions
   ↪  for higher powers of $y$.
14
15 First, cube both sides to find $y^9$:
16 $$y^9 = (xy^2x^{-1})^3 = x(y^2)^3x^{-1} = xy^6x^{-1}$$
17 Since $y^6 = (y^3)^2 = (xy^2x^{-1})^2 = xy^4x^{-1}$, we can substitute this back
   ↪  in:
18 $$y^9 = x(xy^4x^{-1})x^{-1} = x^2y^4x^{-2}$$
19
20 Next, cube both sides again to find $y^{27}$:
21 $$y^{27} = (x^2y^4x^{-2})^3 = x^2(y^4)^3x^{-2} = x^2y^{12}x^{-2}$$
22 Since $y^{12} = (y^3)^4 = (xy^2x^{-1})^4 = xy^8x^{-1}$, we substitute this in:
23 $$y^{27} = x^2(xy^8x^{-1})x^{-2} = x^3y^8x^{-3}$$
24
25 ### **3. Substitute Using the Second Relation**
26 We now have $y^{27} = x^3y^8x^{-3}$. We can substitute $x^3 = yx^2y^{-1}$ and
   ↪  $x^{-3} = yx^{-2}y^{-1}$ into this equation:
27 $$y^{27} = (yx^2y^{-1})y^8(yx^{-2}y^{-1}) = yx^2y^7x^{-2}y^{-1}$$
28
29 Because $y^{27}$ commutes with $y^{-1}$, we can multiply by $y^{-1}$ on the left
   ↪  and $y$ on the right to conjugate the equation:
30 $$y^{-1}y^{27}y = x^2y^7x^{-2}$$
31 $$y^{27} = x^2y^7x^{-2}$$
32
33 Multiplying by $x^{-2}$ on the left and $x^2$ on the right, we isolate $y^7$:
34 $$x^{-2}y^{27}x^2 = y^7$$
35
36 ### **4. Compute $x^{-2}y^{27}x^2$ Directly**
37 Let's find a different expression for $x^{-2}y^{27}x^2$ using the known relation
   ↪  $x^{-1}y^3x = y^2$.
38
39 First, conjugate $y^{27}$ by $x$ once:
40 $$x^{-1}y^{27}x = (x^{-1}y^3x)^9 = (y^2)^9 = y^{18}$$
41
42 Now, conjugate it by $x$ a second time:
43 $$x^{-2}y^{27}x^2 = x^{-1}(y^{18})x = (x^{-1}y^3x)^6 = (y^2)^6 = y^{12}$$
```

```

### **5. Deduce the Order of $y$**
From Step 3, we know $x^{-2}y^{27}x^2 = y^7$.
From Step 4, we know $x^{-2}y^{27}x^2 = y^{12}$.

Equating the two results:
$$y^{12} = y^7$$
$$y^5 = 1$$

### **6. Show That $x = 1$ and $y = 1$**
Since $y^5 = 1$, we know that $y^6 = y$ and $y^4 = y^{-1}$.

Let's return to $x^{-1}y^3x = y^2$ and square both sides:
$$x^{-1}y^6x = y^4$$
Substituting $y^6 = y$ and $y^4 = y^{-1}$, we get:
$$x^{-1}yx = y^{-1}$$

This means conjugating $y$ by $x$ inverts $y$. If we conjugate by $x$ twice, we get $y$ back:
$$x^{-2}yx^2 = x^{-1}(x^{-1}yx)x = x^{-1}(y^{-1})x = (x^{-1}yx)^{-1} = (y^{-1})^{-1} = y$$
Because $x^{-2}yx^2 = y$, we can deduce that $y$ commutes with $x^2$.

Now, let's revisit the second main relation: $yx^2y^{-1} = x^3$.
Because $y$ commutes with $x^2$, the left side simplifies perfectly:
$$yx^2y^{-1} = x^2$$

Therefore:
$$x^2 = x^3 \implies x = 1$$

Finally, substitute $x = 1$ back into our inversion equation $x^{-1}yx = y^{-1}$:
$$(1)y(1) = y^{-1} \implies y = y^{-1} \implies y^2 = 1$$

Since $y^5 = 1$ and $y^2 = 1$, the only possible solution is $y = 1$ (as the greatest common divisor of 5 and 2 is 1).

Because $x = 1$ and $y = 1$, the group $G$ is generated entirely by the identity element. Thus, $G$ is the trivial group.
```