I generally agree with the review provided by the other referee on this manuscript. I think that they summarized the experimental setups and findings well. This is an interesting body of work, and I think that my main concerns lie with how the experimental set-up affects the ability to generate substantial plucking and how to relate this to a field setting. I also have concerns in the definition of plucking as used here, which could be a misconception on my part of the experimental setup. In any case, this work shows interesting relationships on how fracture spacing generates autogenic roughness in a bedrock channel. Additionally, I recommend that the authors read throughout their manuscript for spelling and grammar errors and casual writing. An example is line 326 “yet we have no clue whether it is lower…”

Experimental setup:

I had a hard time understanding the fracture network within your discs. Were the fracture networks 3D, or did you just have 2D fractures? I think that a conceptual figure for both this and how you sum up dip angles and fracture lengths would be really helpful.

With an abrasion mill type setup, you aren’t really able to replicate the processes involved in plucking. With a purely abrasion scenario, grains circulate, eroding the bed to either produce more grains or finer material. In nature, plucking is generally represented as knick progression upstream, a sort of unraveling of the bed. In addition to a fracture network of sufficient density to create blocks that are erodible under realistic hydraulic conditions, this requires a downstream boundary of an exposed block. This can occur due to vertical lift, base level drop, extreme abrasion, etc. In the abrasion mill setup, you don’t have the ability to initiate the plucking with a downstream boundary (at least as you have set it up with a flat bed). Which in and of itself is okay, because you just have to simulate the ‘first’ plucking event by lift to trigger plucking. But then the amount of erosion by plucking that you are able to generate is only one circumference of your circle until you reach the first block again. Then you must re-create the higher erosional threshold first plucking event to trigger a wave around your disk. Further, in the real world these larger blocks would wash downstream and exit the system (or deposit downstream). What happens to them in your mill? This is not necessarily unrealistic, but deserves discussion. I think also this hinges on your blocks being fully separated from the bed and readily mobilized. If you don’t have a 3D fracture network, then are these events really plucking? Or are sediment impacts able to abrade larger chunks of the cement between fractures?

Due to the circular nature of your disk, the fracture orientation relative to flow is never consistent so it is difficult to draw conclusions about fracture orientation and dip angle. These orientations relative to the flow direction are extremely important, and having them change is likely a cause of the variability in your data.

Line 254-255 “and we therefore suggest that the shape of the local erosion rate distribution informs on the occurrence of plucking.” Considering that plucking contributes to the erosion rate distribution, the erosion rate distribution cannot control the occurrence of plucking. Maybe this needs to be reworded for clarity.

What are dimensions of blocks created by your fracture network? Block geometry has been shown to have a significant impact on plucking mode and erodibility (see Lamb et al., 2015; Hurst et al., 2021; Lamb and Dietrich, 2009), and so the dimensions of blocks created by your fracture network should affect how ‘pluckable’ the blocks are.

It would be interesting to run these experiments with a range of sediment inputs. I’m curious if with your setup you would even get any ‘plucking’ if you don’t have sediment impacts since it seems you don’t have fractures at the base. The total percentage of erosion by plucking will depend on how abradable your bed is compared to how pluckable your blocks are.

Why do you think less plucking in the highest fracture density? Are the plucked blocks so small that they are below your threshold of detectability for plucking?

Fig 5/dip angle variability conclusions. How many runs did you have at each dip angle? It looks like you had greater variability in runs where you conducted more repetitions. So that isn’t a strong conclusion if you only had 1-2 runs in the lower variability cases and would need further discussion and analysis.

Discussion:
While your discussion section does a nice job of talking about all of your experimental runs together, I think that more needs to be done to put your work into the context of previous work and discuss how this applies to natural settings.

In discussion and throughout paper you need to be clear that you are exploring a limited scenario where you have a constant supply of sediment. This interaction of sediment can really impact the dynamics of plucking vs. abrasion. You also have a highly erodible bed that is susceptible to abrasion, which can influence the dominance of abrasion over plucking in terms of overall erosion. To me, the fact that there is so little difference in average erosion rates indicates one of two things. Either a) the experimental setup is preventing a greater magnitude of plucking from occurring and since the bed can't unravel, the stochastic plucking events are contributing a great magnitude of erosion that is averaged out over time (a real thing that happens where rapid events with high magnitudes of erosion are interspersed with long periods of no erosion!!! i.e. jokulhaups in Iceland. So a cool interpretation if you can back it up with observations or data) or b) erosion by abrasion dominates the long-term erosion rates. It would be useful to discuss how these dynamics would change with a different experimental setup or different erodibilities of the bed. I think that looking at your video footage of your experiments in depth could start to untangle some of these causes. I think it is at least important to discuss why you think these rates are so similar.

In discussion, it would be useful to compare erosional thresholds for each of your fracture orientations and spacings. Each fracture network geometry results in different block geometries, which have been shown to have a significant impact on plucking mode (see Lamb et al. 2015, Hurst et al. 2021, Lamb et al. 2009)

References
Hurst, A. A., Anderson, R. S., & Crimaldi, J. P. (2021). Toward entrainment thresholds in fluvial plucking. *Journal of Geophysical Research: Earth Surface*, *126*(5), e2020JF005944.

Lamb, M. P., Finnegan, N. J., Scheingross, J. S., & Sklar, L. S. (2015). New insights into the mechanics of fluvial bedrock erosion through flume experiments and theory. *Geomorphology*, *244*, 33-55.

Lamb, M. P., & Dietrich, W. E. (2009). The persistence of waterfalls in fractured rock. *Geological Society of America Bulletin*, *121*(7-8), 1123-1134.

Esurf review aspects:
1. Does the paper address relevant scientific questions within the scope of ESurf?
Yes
2. Does the paper present novel concepts, ideas, tools, or data?
Yes
3. Are substantial conclusions reached?
Yes
4. Are the scientific methods and assumptions valid and clearly outlined?
Yes
5. Are the results sufficient to support the interpretations and conclusions?
Not necessarily – see review.
6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?
Yes
7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution?
Yes
8. Does the title clearly reflect the contents of the paper?
Yes

9. Does the abstract provide a concise and complete summary?
Yes
10. Is the overall presentation well structured and clear?
Yes
11. Is the language fluent and precise?
Yes
12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?
Yes
13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
Yes
14. Are the number and quality of references appropriate?
Not necessarily – see other referee review.
15. Is the amount and quality of supplementary material appropriate?
Yes