

# AIRR Coding Codebook (Training Manual v1.0)

---

## Introduction

Purpose:

This codebook trains human coders to classify posts from the Polymath project dataset into thematic categories. The goal is calibration between human and AI coding and to measure reliability.

Coder expectations:

- Work independently (no discussion until after coding round).
- Follow the decision key and rules for consistency.
- If uncertain, choose the best-fit code and add a note.
- Avoid overthinking: rely on definitions and precedence rules.

## Dichotomous Key (Decision Aid)

1. Does the post propose or suggest a new approach to the problem? Yes → Idea Generation. No → Continue.
2. Does the post test, critique, or evaluate an approach (assumptions, counterexamples, limits)? Yes → Evaluation/Testing. No → Continue.
3. Is the post primarily exploring or working through problem-solving steps, calculations, or provisional reasoning (without a framed proposal or critique)? Yes → Working/TAP. No → Continue.
4. Is the post about organizing collaboration or process (summaries, workflow, resource sharing)? Yes → Meta/Coordination. No → Continue.
5. Is the post clarifying communication (definitions, explanations, acknowledgments, or requests)? Yes → Clarification/Communication. No → Flag as unclear/other.

## Themes, Rules, and Subtypes

### Working / TAP (Think-Aloud Problem-Solving)

Definition: Posts that show exploratory reasoning, calculations, or provisional thoughts without framing them as proposals or critiques.

Precedence: TAP dominates when the main function is exploration, even if it contains fragments of proposals.

Examples:

- "If we take  $k = 3$ , then the sum looks like this... maybe it simplifies?"
- "I don't know if this is useful, but computing the first few terms gives..."

## Idea Generation

Definition: Posts that introduce a new approach, refinement, or framework.

Precedence: Takes priority if a post frames a new idea clearly, even if exploratory work is present.

Subtypes:

1. Problem Reformulation — Recasting the problem in a new framework.
2. Special Case Proposal — Treating a restricted case to simplify.
3. Refinement — Improving or adjusting an existing step.
4. Alternate Technique — Suggesting a different method entirely.
5. Case Unification — Merging multiple subcases into a single framework.

Examples:

- "Can we think of this as a graph coloring problem?"
- "Let's just analyze the  $n=2$  case first."

## Evaluation / Testing

Definition: Posts that test, critique, or probe validity (assumptions, counterexamples, edge cases).

Precedence: Takes priority over Clarification when assumptions are explicitly examined.

Examples:

- "That argument breaks if we try  $n=17$ ."
- "This relies on an assumption that doesn't hold for primes."

## Meta / Coordination

Definition: Posts about managing collaboration or process, rather than advancing the mathematics itself.

Subtypes:

- Process Proposal — Suggestions for workflow or task management.
- Shared Resource Proposed/Updated — Sharing or updating collaborative tools/resources.

Examples:

- "Maybe we should split the proof cases into separate threads."
- "I've updated the wiki page with the new lemma."

## Clarification / Communication

Definition: Posts that request or provide clarification, ensuring shared understanding.

Precedence: Lower priority than Evaluation when the post explicitly challenges validity.

Examples:

- "By 'regular sequence,' I meant..."
- "Could you explain what you meant by the bound?"

## Precedence Rules Summary

- TAP > others when exploration dominates.

- Idea > TAP if a new proposal is explicitly framed.
- Evaluation > Clarification when assumptions are tested.
- Meta > Substantive if organizational intent is the main focus.

### **Edge Cases & Pitfalls**

- Posts with mixed TAP + Idea → code Idea only if the proposal is explicit.
- Posts with Eval + Clarification → code Evaluation if it tests or challenges.
- Posts with Meta + content → code Meta if the focus is organization.

## Examples Appendix (Populated)

Below are sample Polymath post excerpts for each theme. These are anonymized but representative.

### Working/TAP

Post 1631 (Participant A008): The two examples we have for a lower bound for  $c_5$  being 150 have cubes with 14 and 12 points is there an example with all cubes 16 or higher and total 150 or higher? Or failing that all cubes 15 or higher and total 150 or higher?

Post 1852 (Participant A006): 955., The term  $c_7$  in my last comment should be  $c'_7$ .

Post 1115 (Participant A012): 1004.1. Wikified here. Hopefully this will help with the generalization of line-free sets correlating with 12-sets.

### Idea Generation

Post 1276 (Participant A005): We definitely could use more visualizations (I think Thomas Sauvaget is the only one so far who has depicted  $c_n$  visually), although things would first have to be packaged in such a way that a casual participant could understand and make such a thing.

Post 1882 (Participant A002): 980 Linear Programming., In 770 Terry described that improvements on bounds of type 4D: etc help improving bounds in dimension 7., I expect, that for In dimension 4, with  $e=1$  a much better result holds. In 778 KS Chua mentioned that the best example of this type he found had 36 points, (and his programme found a 124 point example...

Post 1314 (Participant A015): For me the exciting parts seem to be the simpler proof and the need to rewrite in accessible language to enable people to join in. Maybe this can be made one of the explicit goals of the project. Not just to create new maths but to work on making mathematics that can be accessible., I say the same thing, but...

### Evaluation/Testing

Post 2039 (Participant A009): 1125. c-statistics, I ran a search for the sixteen c-statistics ( $2 \times 22$ ,  $22 \times 2$ , etc) I could only find 65 rules, from the 65 combinatorial lines., My results limit the total possible points for 4D Moser below 73, which is higher than 43. Either I have a bug, or there was too little connection between the various c statistics., For the record,...

Post 331 (Participant A016): 20. A related observation is that if you've got three rich triples in the right configuration, you aren't guaranteed a combinatorial line. For example, in two of the sets the first coordinate could always be 1 and in the third it could always be 2.

Post 1757 (Participant A009): 772. Moser(3), Terry, in your writeup of Kristal's proof, you said that any solution with  $d > 2$  had at most 40 points. From reading this blog and the wiki, I can only tell that such a solution has at most 41 points. What have I missed?

## Meta/Coordination

Post 2060 (Participant A008): 1142. lower bounds on line free sets, I think the results of 1141 can be used to get lower bounds on lines free sets for large  $n$  for all values of  $k$ . For any  $k$  and any  $n$  we can find a prime prelatively close to  $k^n$  then we remove the first  $k+1$  values mod  $p$  then we pick a...

Post 872 (Participant A012): 803.1 Positivity: , In my experience, obvious positivity on the physical side doesn't always imply obvious positivity on the Fourier side. Here is an example (although it goes in the reverse direction). , Let  $\mu$  be arbitrary and consider  $\nu$ , where  $\nu$  is uniform random and is formed by flipping each bit of  $\mu$  with probability  $\frac{1}{2}$ . (Note that  $\nu$  has the same distribution...

Post 2043 (Participant A008): 1129.  $c^{\{\mu\}}_4=9$ , If  $c^{\{\mu\}}_4$  is greater than 9 then the counterexample must occur when there is exactly one point of the form  $aaaa$  removed. The other three cases are below., If we have all three points of the form  $aaaa$  removed then the remaining points have value 12 and we have covered all lines any set of moving coordinates with...

## Clarification/Communication

Post 1418 (Participant A012): 1068. Hi Tim, re the outline in 1065: , Looks good. One question: It looks like you've divided the proof into three main lemmas: multidim-Sperner (more generally, multidim-DHJ( $k-1$ )), line-free set correlating with intersections of  $ij$ -insensitive sets, and  $ij$ -insensitive sets being partitionable. , It seems to me that the Varnavides-version of multidim-Sperner (more generally, multidim-DHJ( $k-1$ )) may as well be considered the...

Post 66 (Participant A001): Thanks for Gowers's comments. I really understand that the author may forget his paper after a long time. However, I talk with some authors just for the very recently papers..... Maybe, the case I meeting is a little special. Anyway, as a begginer, I do not understand why the authors order is alphabetic in mathematics.

Post 1483 (Participant A016): The problem was that, if I understand correctly, Terry had some way of discretizing Tim Austin's argument. But perhaps that argument isn't sufficiently written to make it a problem to call ours the first finitary proof. We should probably see what Terry thinks about it.