CHAPTER

Task 3 -- Answering questions -- Individual questions

CONTENTS

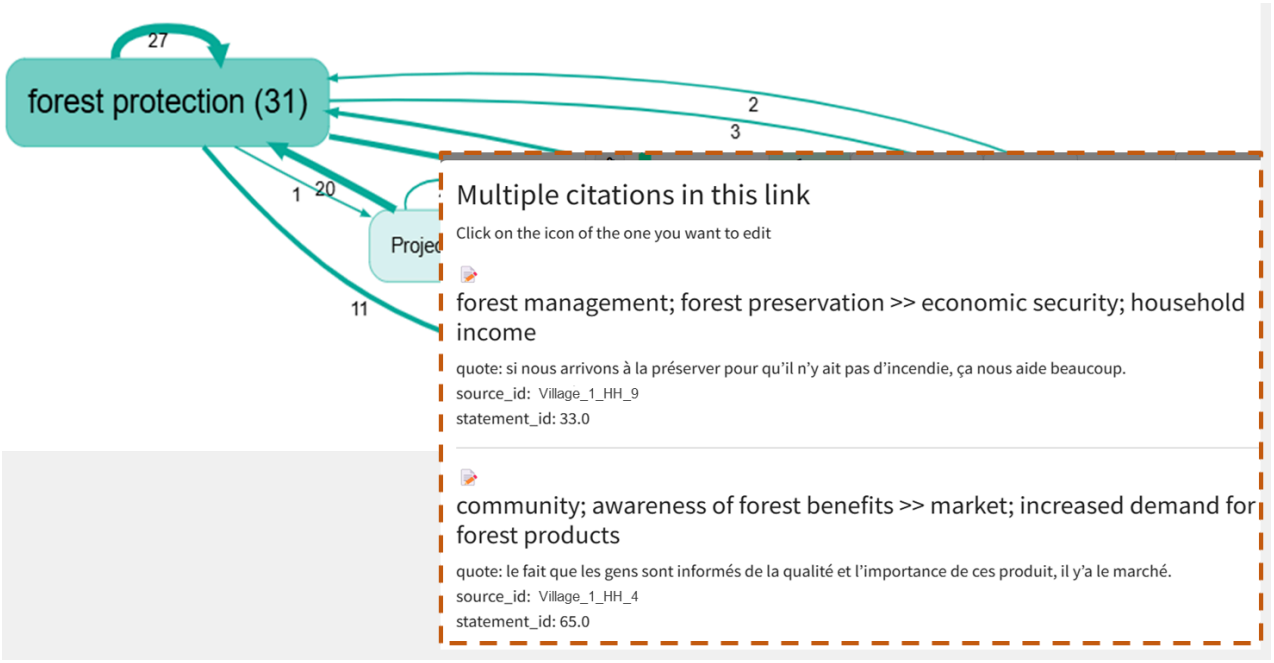
Individual questions -- introduction

Individual views -- How does the system work according to individual sources?

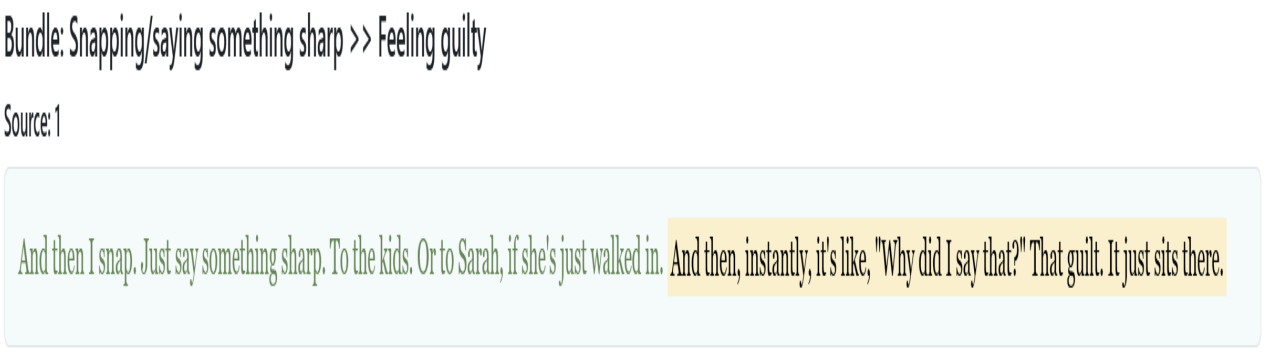
Summarising -- How do the sources claim that the system works, in summary?

Vignettes -- What is a typical source and what is their story?

What are the narratives behind a specific link?



It is important to frequently return to the original quotes associated with each factor or link to **understand how different stakeholders interpret and talk about key concepts**. This can be done in Causal Map by clicking on a link in the interactive map, or by printing out quotes for a particular filter (e.g. just for a single bundle of links) with additional context and metadata.



Relevant page from Causal Map help:
Print View

Which factors and links were most frequently mentioned?

Which factors and links are mentioned by the most sources?

Main outcomes. Which factors are mentioned most often as outcomes?

Main drivers. Which factors are mentioned most often as drivers?

Splitting by groups. Are different groups involved in different ways?

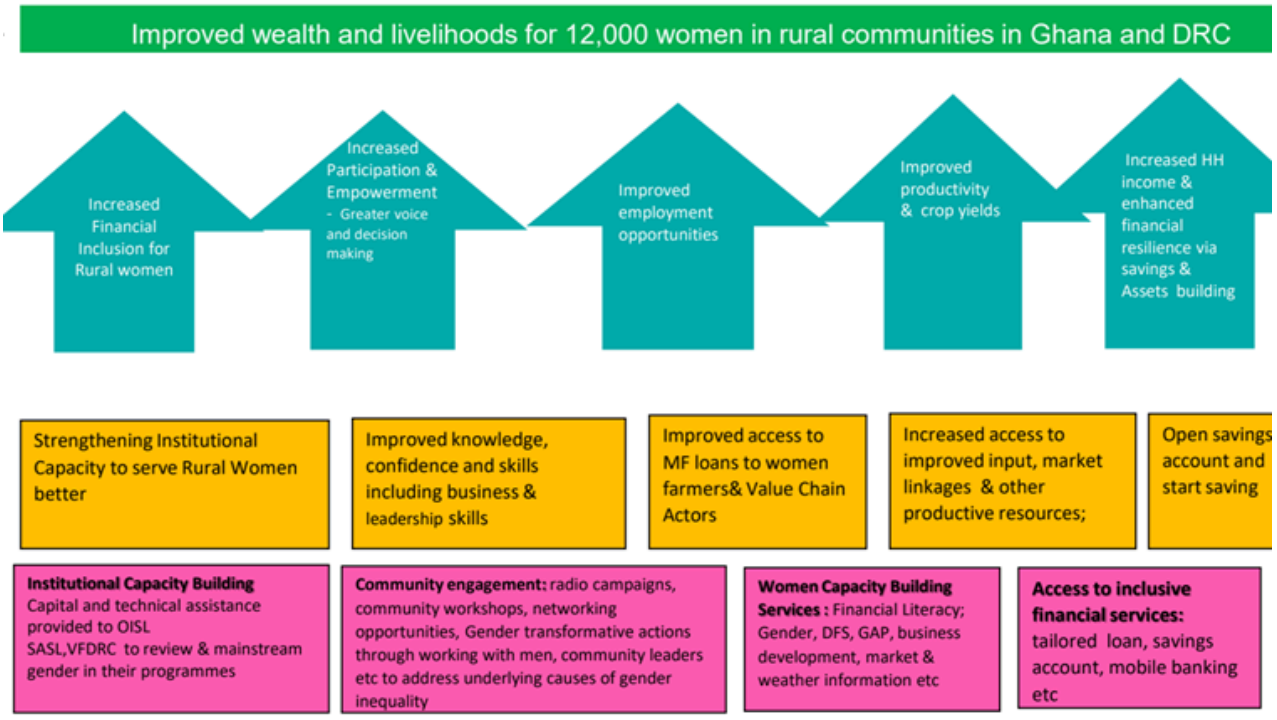
Comparing groups -- What factors or links were mentioned more by some groups than others, in the same map?

Identifying groups -- Are there different subgroups within the data?

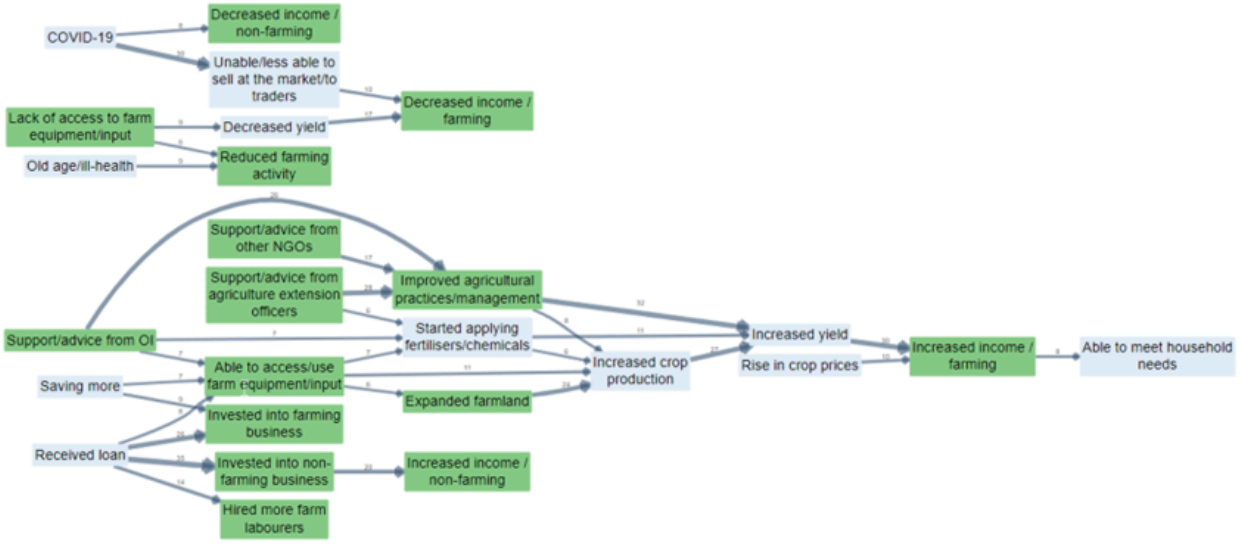
What are the emerging or unexpected factors?

One way to do identify emerging or unexpected factors is to use the elements from your theory of change as your codebook while coding and only adding other elements when necessary, making a note of these additional elements.

Does the evidence support your theory of change?



One way to do this is to use the elements from your theory of change as your codebook and only add other elements when necessary. Validated pathways of change (by showing which mechanisms are observed on the ground) and find gaps where expected pathways might be missing or where stakeholders list elements were not anticipated in the theory of change?



We discuss this approach at more length in (Powell et al., 2023).

References

Powell, Larquemin, Copestake, Remnant, & Avard (2023). *Does Our Theory Match Your Theory? Theories of Change and Causal Maps in Ghana*. In *Strategic Thinking, Design and the Theory of Change. A Framework for Designing Impactful and Transformational Social Interventions.*

Showing group data as custom link labels

One of the most exciting applications of causal mapping is to assess change over time within a system. If we apply a systematic approach to coding (using blindfolded manual coding or AI-supported coding) we can compare the frequencies with which links or factors are mentioned over time. This becomes particularly interesting when applying inductive coding, so that new and emerging phenomena can be included into the codebook. Re-applying new codes to previously coded data would be very tedious with manual coding but is easy to do with AI-supported coding: Transforms Filters -- Soft Recode with Magnetic Labels

More details are given in this paper: (Powell et al., 2025)

Unfortunately, there is not much consensus about what assessing systems change means. Sometimes we read about *measuring* systems change, which would imply assigning numbers to change, but often just means "assessing".

(Rizzardi, 2025)

References

Powell, Cabral, & Mishan (2025). *A Workflow for Collecting and Understanding Stories at Scale, Supported by Artificial Intelligence*. SAGE PublicationsSage UK: London, England. https://doi.org/10.1177/13563890251328640.

Rizzardi (2025). *Systems Change | Modern Slavery*. https://www.freedomfund.org/news/systems-change-pathways-measurement/.

Sentiment -- Which changes are perceived as most positive or negative?

Focusing on specific factors. What influences and outcomes are connected to a specific factor?

Looking downstream. What are the direct and indirect consequences of one or more factors?

Looking upstream. What are the direct and indirect influences on one or more factors?

Names of tables and fields

Columns and tables

We can think of a causal map as a database consisting of two tables, the links table and the sources table. We don't need to have a separate table for the factors because the factors can be derived from the links table.

Columns in both the links and factors tables

Field name	Explanation
Citation count aka Link count	Number of citations of a given factor or link.
Source count	Number of sources mentioning a given factor or link. Source count cannot be higher than citation count and may be a lot lower if some sources mentioned the same factor or link many times.

Columns in the factors table

Field name	Explanation
incoming_links, indegree	Number of citations of all the incoming links to a particular factor.
outgoing_links, outdegree	Number of citations of all the outgoing links from a particular factor.
outcome-ness (%)	A factor with a high outcomeness percentage is mostly an outcome; it has mostly incoming links. If it has low outcomeness it has mostly outgoing links so it is mostly a driver. Outcomeness is the proportion of citations of incoming links out of all the citations of a particular factor: a normalised version of the Copeland Score (Copeland, 1951). So factors with high *outcomeness* can be thought of as "outcomes". And factors with low outcomeness can be thought of as inputs or drivers.

Columns in the links table

Field name	Explanation

Columns in the sources table

Field name	Explanation
source_id	
title	
filename	

Glossary ▸

References

Copeland (1951). *A Reasonable Social Welfare Function*.

Path tracing -- How do one or more causes affect one or more effects, including indirect pathways?

Source tracing -- What are the consequences of one or more factors, looking only at stories told in their entirety by individual sources?

Robustness -- How robust is the evidence for that X influences Y?

Counting and comparing influences

How much evidence is there for the influence of our intervention on a valued outcome? Is that a lot? Can we compare these numbers across pathways?

In this map, six sources told stories which start with our intervention and end with Wellbeing (tracing threads). Is this a lot?

Well, we can compare that with the total number of sources to make a source-mention-proportion, to say that 6 out of 24 sources mentioned this pathway.

Or we can compare it with the number of sources mentioning Wellbeing at all, to say that 6 out of the 16 sources mentioning Wellbeing told stories which began with our intervention.

Or we can compare it with the number of sources telling stories ending with Wellbeing and beginning with a different intervention.

Or we can compare any of these figures with the same figures from a previous time point.

Using numbers and proportions like this in a fundamentally qualitative approach like causal mapping can be very useful but we have to be careful. These are quite fragile indicators which can be easily influenced by other factors (for example, how visible was our intervention?) and can be hard to generalise.

We should always also consider the evidence itself behind each link by looking at the quotes.

Properties of the causal map – Which factors are reported as being causally central or causally peripheral?

Properties of the causal map -- What is the overall structure of the network?

Properties of the causal map -- Are there leverage points?

Properties of the causal map -- Are there feedback loops?

Combining questions

Causal mapping gets really useful when you start to combine the different questions you might want to ask in order to answer more sophisticated questions. We can think of many of the techniques as filters which filter the view in a particular way. Using multiple filters allows you to build up an answer to a question. Usually, order matters.

tribes-intro

Tribes: the most relevantly different subgroups in your data (by causal story)

The **Tribes** filter answers a very specific analysis question:

> **What are the most relevantly different subgroups in the data in terms of the causal stories they tell?**

In other words: if your sources contain *different narratives about how the system works*, Tribes tries to find those subgroups for you.

This aligns with the broader idea that causal mapping helps you make sense of **many causal claims from many sources** (not to build a single "true" causal model). See the broader framing in the CausalMap "Ideas Garden" PDF. Causal Mapping: 97 ideas

When is Tribes useful?

Use it when you want to:

- surface **distinct narrative patterns** without predefining groups
- compare "how the system works" across sources (not just what's frequent overall)
- identify disagreement or tension (e.g. different explanations of the same outcomes)

It's less useful when:

- you have very few sources (any clustering will be unstable)
- the sources mostly tell the same story (there may be no strong subgroups to find)

What Tribes clusters (conceptually)

Tribes clusters **sources**, not links.

Each source is summarised by the pattern of causal links it contains (cause→effect, plus sentiment buckets), then **k-means** groups sources that have similar patterns.

The output is a new field on links: `custom_tribeId` (plus similarity diagnostics), so you can analyze the same map "through the lens of tribes".

Controls (how to think about them)

Number of clusters (k)

This is "how many subgroups do you want to see?".

Practical workflow:

1. start with k = 2–4

2. inspect the result (do the tribes look meaningfully different?)
3. adjust k up/down (too low merges stories; too high splits hairs)

Similarity cutoff + Drop unmatched

Each source has a similarity to its assigned tribe.

- **Similarity cutoff**: minimum similarity for a source to count as a "good fit"
- **Drop unmatched**:
- ON: remove links from weakly-assigned sources (cleaner tribes, less coverage)
- OFF: keep them (more coverage, more mixing)

Min cluster %

Prevents the "1 big cluster + many tiny clusters" pattern.

Clusters below the threshold are discarded and their sources are reassigned to the nearest surviving cluster (subject to similarity + drop rules).

View Tribe Report: Sources vs Citations

The **View Tribe Report** button generates chi-square tables for categorical fields by tribe. You can choose what "counts" mean:

- **Sources** (default): each source contributes at most 1 to a cell (more robust if some participants produce many links)
- **Citations**: each link contributes 1 (more sensitive to "talkative" sources)

Using the Statistics (Pivot) tab to explain *why* tribes differ

Once you have `custom_tribeId`, the next question is usually:

> "Do these tribes line up with anything we already know about our sources (gender, region, program arm, interview round…)?"

A practical workflow:

1. run Tribes (so links have `custom_tribeId`)
2. open **Statistics / Pivot**
3. use the dataset that reflects your **current filtered links**
4. compare tribes against known characteristics (e.g. `custom_gender`, `custom_region`)

This helps you distinguish:

- tribes that are "purely story-based" (not explained by known demographics)
- tribes that correlate with known subgroups (e.g. region-specific causal mechanisms)

Toy example

Imagine 30 sources. Tribes with k=3 might reveal:

- Tribe A: “training → productivity → income”
- Tribe B: “prices → debt → stress”
- Tribe C: “weather → crops → migration”

Next steps:

- use Statistics/Pivot to see whether (say) Tribe C is concentrated in drought-prone regions
- use Custom Links Label (next post) to label edges by tribe composition and spot where the narratives diverge in the map

Showing group data as custom link labels

Custom Link Labels (and chi-square “significance”) on the map

Custom Links Label can use **any field that exists in the current links table**.

That includes:

- **existing fields** already in links (e.g. `sentiment`, tags, AI metadata you’ve stored)
- **source custom columns** that are joined onto links (e.g. `custom_gender`, `custom_village`, `custom_region`)
- **fields created by filters** (e.g. **Tribes** adds `custom_tribeId`)

That’s what **Custom Links Label** does: it configures the map’s **Link Labels** to show a compact summary per connection.

What is being summarised?

The map groups links into **bundles** by cause→effect.

For each bundle (say `Education → Employment`) Custom Links Label looks at the links inside that bundle and summarises the selected field.

Common choices:

- after Tribes: `custom_tribeId`
- for subgroup comparisons: `custom_gender`, `custom_village`, `custom_region`
- for link-level fields: `sentiment` (or any other link metadata you’ve added)

Counts: Sources vs Citations

Before choosing a display mode, decide what a “count” means:

- **Sources** (default): each source counts once per value in the bundle (unique `source_id`)
- **Citations**: each link counts (raw link count)

If some sources produce many more links than others, **Sources** is usually the more interpretable option.

Display modes

Assume you pick field = `custom_tribeId` (or `custom_gender`) and Counts = Sources.

Tally

Shows counts per value in the bundle.

Example label on an edge:

- `T1:4 T2:1`

Interpretation: 4 sources from tribe T1 and 1 source from tribe T2 contain at least one link in this cause→effect bundle.

International development-style example:

- Edge bundle: `Cash transfers → Food security`
- Label (Counts = Sources): `T1:9 T2:2`
- Interpretation: this connection is mainly supported by sources in tribe T1 (e.g. “economic buffering” story), with a smaller contribution from T2 (e.g. “market/price volatility” story).

Example using an existing source custom column (Counts = Sources):

- Field: `custom_gender`
- Edge bundle: `School fees → Dropout`
- Label: `Female:7 Male:2`
- Interpretation: more sources coded this link in the “Female” group than in the “Male” group (but you still need a baseline-aware view before calling it “a real difference” — see chi-square).

Percentage

For each value, shows:

[

\text{percent} = \frac{\text{count of this value in this bundle}}{\text{total count of this value across all currently filtered links}} \times 100

]

Example label:

- `T1:18% T2:6%`

Interpretation: 18% of T1’s total (source-based) participation in the *current filtered map* appears in this one bundle.

This is a “share of that group’s activity” view.

Chi-square (significance)

Why would you want this?

When you look at subgroup tallies on a map (tribes, gender, village, etc.), it's very easy to over-interpret small differences that are just **random noise** (especially if you have many bundles and many groups).

Chi-square is a quick way to focus your attention on differences that are **probably not noise** given the overall mix of your data.

What the test does:

For each value (e.g. a tribe ID), it checks whether that value appears in this bundle **more (↑)** or **less (↓)** than you'd expect given its overall prevalence in the filtered links.

For each value:

- **Observed** = its count in this bundle
- **Expected** = (bundle size) × (value total) / (grand total)
- Mark as significant if chi-square contribution (> 3.84) (≈ p < .05, df=1)

You have three chi-square display styles:

1) **Chi-square** (no counts)

Example: `45 (T1↑, T3↓)`

2) **Chi-square (with counts)**

Example: `45 (T1 4↑, T3 3↓)`

3) **Chi-square (with counts/totals)**

Example: `45 (T1 4/5↑, T3 3/6↓)`

Where:

- `45` is the bundle size in the chosen Counts unit
- `4/5` means observed 4 out of value-total 5 (again in the chosen Counts unit)

International development-style example (Counts = Sources):

- Bundle: `Drought → Migration`
- Label: `18 (T3↑, T1↓)`
- Interpretation: the "climate shock" tribe (say T3) shows up in this link bundle more than expected given how often T3 appears in the filtered map overall, while T1 shows up less than expected.

A worked mini-example (Sources)

Suppose in your current filtered map:

- grand total sources represented across all links = 50
- value total for T1 = 5 sources
- bundle size for `Education → Employment` = 45 sources
- observed T1 sources in that bundle = 4

Expected T1 in bundle = (45 × 5 / 50 = 4.5).

Observed (4) is slightly below expected, so it would *not* show as over-represented.

If instead observed were 5 with the same totals, it would skew upward and might appear as `T1↑` depending on the chi-square contribution.

How to use (quick)

1. Add **Custom Links Label** to the filter pipeline
2. Pick **Field** (e.g. `custom_tribeId`)
3. Pick **Counts** (Sources/Citations)
4. Pick **Display mode**
5. In **Map Formatting → Link Labels**, select **Custom Links label**

That's it: the map labels update immediately as you tweak the settings.

