Amsterdam UMC Systems Science SIG · 21 May 2026

Causal mapping is not systems mapping

Coding what people say, then deciding what it means

Steve Powell · Causal Map Ltd · hello@causalmap.app

app.causalmap.app · free, unlimited public projects · garden.causalmap.app

Causal mapping and the Causal Map app

Causal mapping is a 50-year tradition which has been used in cognitive psychology, decision science, management, ecology, programme evaluation and more.

I've been providing the Causal Map app for programme evaluation and qualitative research since 2022.

Its niche: coding causal claims in text, closer to CAQDAS/NVivo than to a diagramming tool. It is a tool for qualitative research.

The app: app.causalmap.app. Free for unlimited *public* projects.

Source-aware coding
every link can be traced back to quote from source text

Manual or AI-assisted coding
AI coding is a great fit

Interactive maps
filter, zoom, trace indirect paths, recode, on the fly

Some recent applications

Causal Map has been used for causal mapping in over 100 academic, evaluation- and policy-oriented projects worldwide.

See also garden.causalmap.app/case-studies and working papers.

Exploring the mental models of Central bank governors

GenAI-supported causal mapping on a sample of central bank governor speeches (1996–2023 corpus): what leaders claim drives national and global economic outcomes, and how that shifts over time.

Urban wellbeing · Lonely in London

Working paper built around 48 interviews with young adults (18–24) in deprived London boroughs; compares an AI-assisted thematic analysis with causal mapping.

Impact evaluation: Global · Global Young Academy

Member narratives from an online survey (103 completions); causal claims grouped into themes and a programme theory of change.

REF impact assessment: UK · King's College London

Long-run impact from departmental research through documentation and interviews to downstream clinics and manufacturers worldwide (via Cactus Communications).

Higher-education evaluation: Chile · Duoc UC (university)

Partnership with Duoc UC, a Chilean university: qualitative exploration of gender gaps for women in STEM technical and professional pathways, using conversational interviews (Qualia) and causal mapping.

Food systems evaluation: Southern Africa · World Food Programme

Retailer key informants and focus groups in Lesotho, Malawi, Mozambique and Zimbabwe mapped for the regional market-development thematic evaluation (Forcier Consulting).

Learning synthesis: Global · PLI

Thematic summaries from seven INGO cases on conditions that foster or hinder group learning in international development

Multi-country advocacy evaluation: 13 countries · INTRAC

End-term evaluation for the SCC consortium: causal pathways from harvested outcomes and partner theory-of-change narratives across

Causal mapping and systems mapping: shared ground

Boxes and arrows are a shared visual language.

Many of the same things matter:

- Identifying the salient causal factors
- Understanding feedback and indirect pathways
- Understanding is more important than predicting

The difference is all about what the arrows mean.

What does an arrow mean: the Janus face

Reading 1: a fact about the world

If X goes up, Y goes up (with a certain strength/ effect size/ regression coefficient).

Facts about the system.

Based on measurement, modelling, simulation, or simply expert judgement.

The boxes are variables.

The map is the start of an argument about reality.

Reading 2: a claim with a source

"This person said X influenced Y, in this context, in this quote."

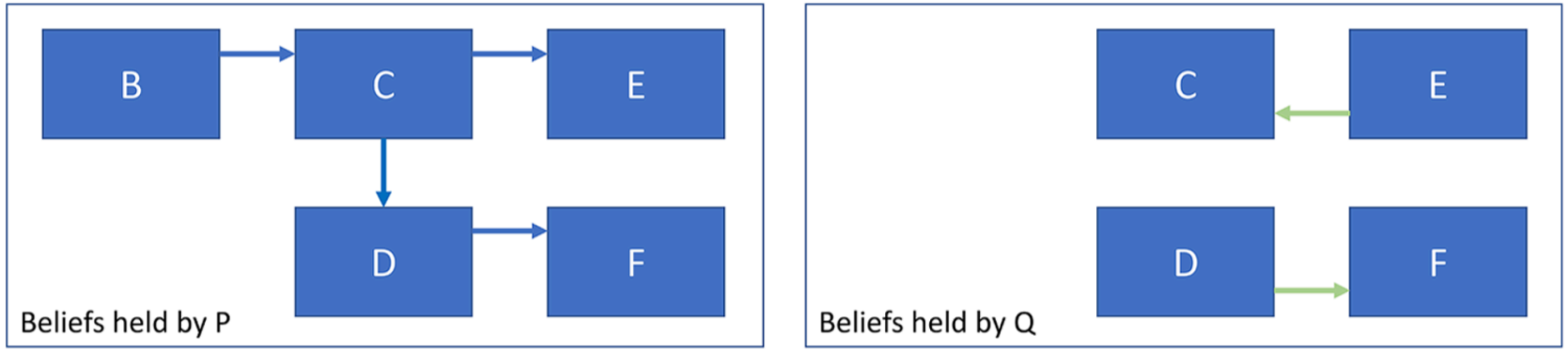
Facts about somebody's account of the world, backed by quotes.

The boxes are events or propositions.

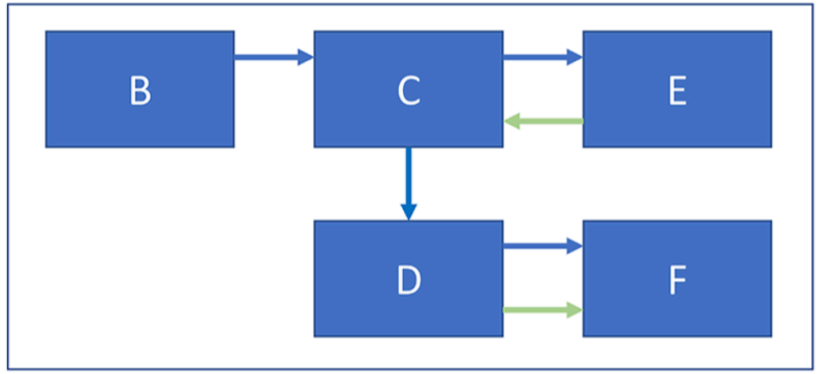
The map is a record of distributed belief.

Systems mapping mostly aligns with the first reading. Causal mapping has historically tended to slide between the two. We very much follow the second reading and are explicit about when we move between the two.

Merging causal maps from multiple sources



Two separate single-source causal maps: Beliefs held by sources P and Q about causal influences between various factors.



A multi-source causal map about beliefs held by sources P and Q.

P and Q's separate beliefs, then combined. Each link still carries source metadata.

Minimalist coding: factors are propositions, not variables

A factor is a short proposition, close to what someone said.

Examples

"Not enough money in the household"

"Did not take a holiday this year"

"Civil society coalitions are present in decision-making bodies"

Not *"money"* as a variable from low to high.

What this minimalist style does not code on the link itself:

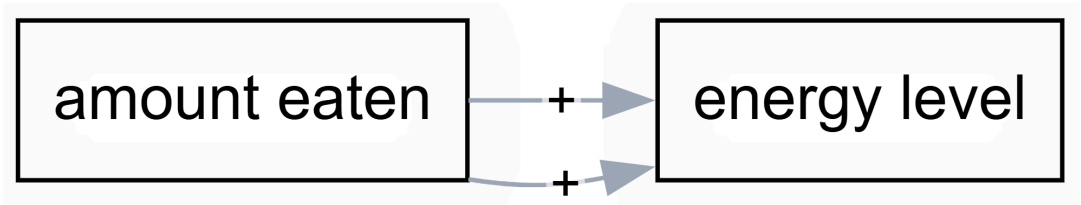
- necessity or sufficiency
- non-linearity or polarity
- moderator status or inhibitor roles
- strength or effect size

People rarely state these explicitly. Coders rarely agree on them from text. Minimalism is honest about what the data can support.

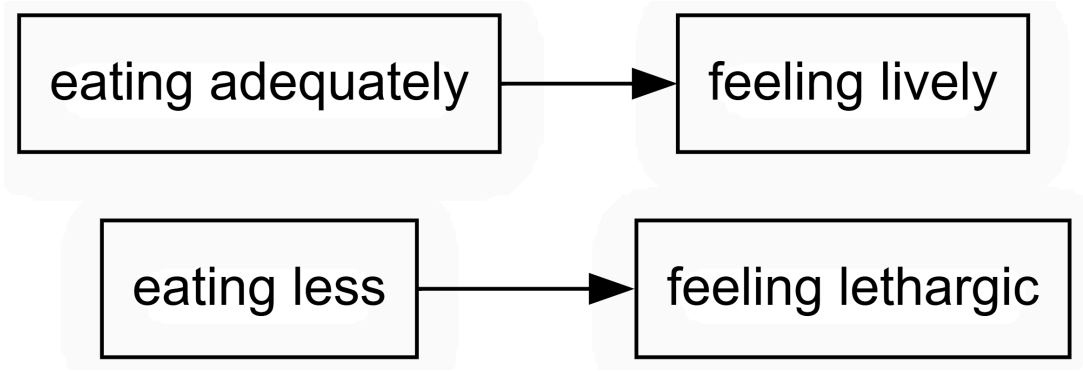
Illustration: variable-style factors vs minimalist propositions

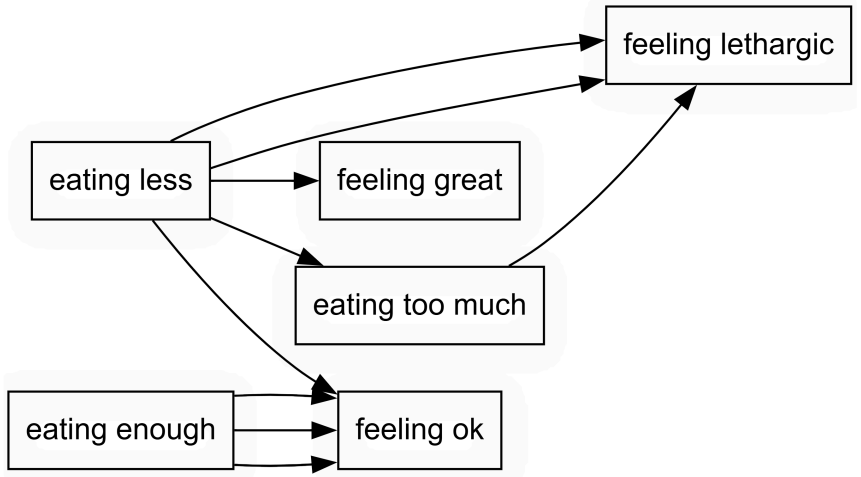
> *Today I ate enough and felt lively; Yesterday I ate less and felt really weak*

Variable-style coding maps different statements onto abstract factor names.



Minimalist coding keeps distinct propositional endpoints so parallel claims stay parallel, not averaged into one link type.





Coding of some more typical statements around the topic; cannot be neatly mapped onto variables.

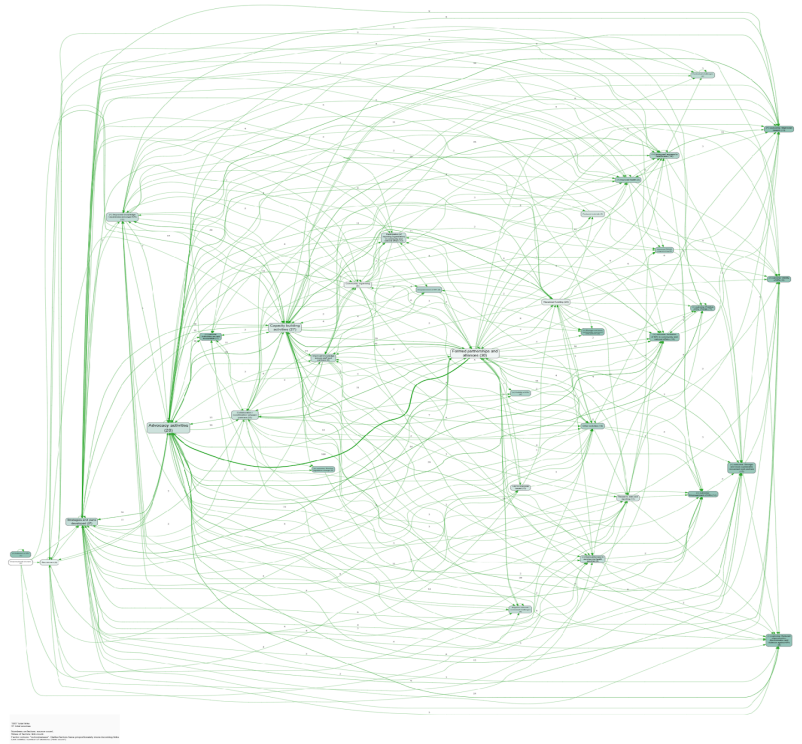
Treating opposed expressions of causes and effects as poles of one latent variable often hides something of what each source actually claimed and at the same time over-generalises.

Why minimalism scales

Bare links plus good labels let analysts:

- aggregate across sources without arguing about mathematical meaning (e.g. + or -, necessary or sufficient, etc.)
- easily compare across groups, sites, projects
- easily summarise with simple counts (citations, sources)
- code thousands of claims at workable speed

It works fully by hand. AI makes it faster at scale.



Hierarchical coding: showing the big picture without deleting the detail

Hierarchical labels let a dense map zoom out without deleting detail. In Causal Map there is no separate factors table: factors are implied by the links table. The way to record attributes of a factor is by using labelling conventions.

Convention

`General; specific`

`New intervention; midwife training`

`Healthy behaviour; hand washing`

Zoom to level 1:

`New intervention -> Healthy behaviour`

Keep detail
Quotes stay attached to specific factors

Report clearly
Roll up to level 1 or 2 for slides

Search better
Seaching for parent labels also finds the children

Count at the right level
Broad family counts without losing sub-items

Use a hierarchy only when the parent is a valid causal factor. `A; B` means B can be reported as evidence for A.

Opposites, polarity and sentiment

Opposites coding handles pairs like:

- `Employed` and `~Employed`
- `Good health` and `~Good health`
- `Eating vegetables` and `~Eating vegetables`

The `~` marks the opposite pole. Optionally combining opposites rewrites both poles to one canonical label, while keeping polarity metadata which can be expressed e.g. by colouring each end of the link.

Sentiment coding adds link metadata, usually `-1`, `0`, `1`. It says whether the claim is positive, neutral or negative in context.

Coding for sentiment does not change the structure of the map, but it marks the tone or valence of a particular claim.

Six moments for quality assurance, especially when inferring facts from claims

Moving from claims to conclusions is a sequence of checks.

1

Coding individual claims
ensure causal claims are coded correctly (high recall and precision).

2

Checking individual claims
reject hallucinated, hypothetical, reversed or over-interpreted links.

3

Moving from claims to bundles
group all X→Y claims; decide whether to create one assessed synthetic link.

4

From bundles to pathways
check whether links form coherent paths and are not assembled from possibly incompatible sources/concepts.

5

Judging value and relative contribution
ask whether the pathway matters and how contribution compares with alternatives.

6

Holistic judgements, the whole thing
does each finding fit the underlying evidence well enough?

Moment 1: coding individual claims

Code the smallest defensible claim: *this source says X influenced Y*.

Cause
Short phrase, close to the text

Effect
Short phrase, close to the text

Quote
Original passage supporting the link

Source
Already known from the source metadata

> Manual or AI, the check is the same: is there a causal claim here, and are the endpoints right?

Moment 2: checking individual claims

Check each raw claim before it feeds the map.

Real causal claim
Not just association, chronology or topic similarity

Correct direction
X influences Y, not Y influences X

Grounded endpoints
Factor labels match what the source said

If hypothetical, mark it
Separate speculation from reported evidence

Moment 3: moving from claims to bundles

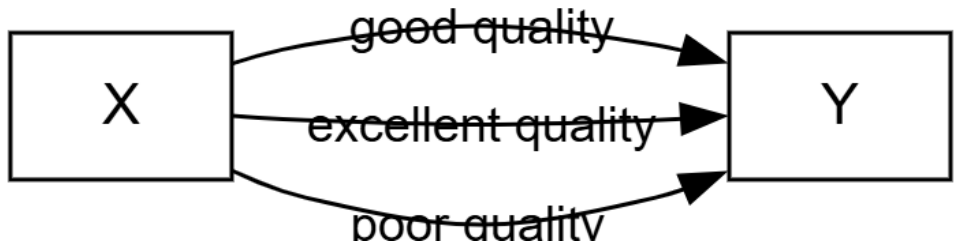
A *bundle* is all claims X→Y across all sources. Either collapse it into one *assessed link*, or skip it if evidence is too thin.

The workflow:

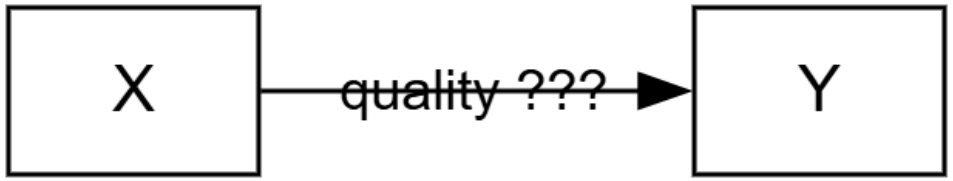
1. Finish coding; focus on the bundles that survive your filters
2. For each bundle: read the quotes, apply your rubric, judge quality
3. Create one assessed link, or none, with optional metrics (e.g. confidence 1-5)
4. Switch the view: unassessed claims stay in the database. the assessed map is what you argue from

The app requires a written rubric before allowing assessed links to be created.

Typical scale: 1,000 raw claims → 30 bundles → 25 assessed links



Unassessed view: many raw claims per bundle



Assessed view: one vouched link per bundle

Moment 4: from bundles to pathways

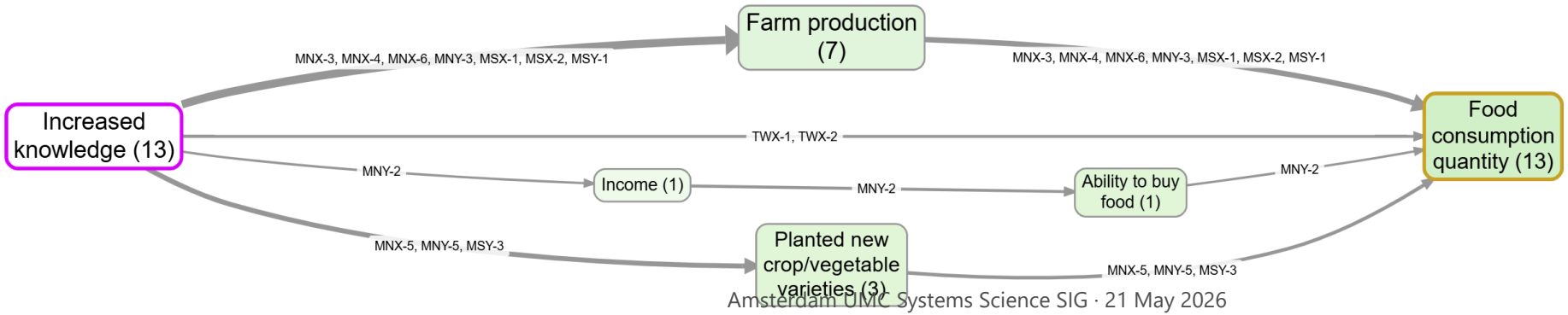
A→B plus B→C does *not* automatically mean A→C. Contexts may not line up.

Path tracing: keeps links lying on *some* path from A to C.

Source tracing: keeps only sources whose *own account* runs all the way from A to C.

The transitivity trap: fragments from different sources can create a pathway nobody actually told.

Source tracing asks: *did anyone tell the whole story from intervention to outcome?*



Moment 5: judging value and relative contribution

Once a pathway is credible, ask about value, practical significance and relative contribution:

- Did this pathway matter?
- Was the contribution large enough to care about?
- How does it compare with alternative explanations/pathways?
- Did it produce valuable, harmful or ambiguous consequences?

> Going from “is there evidence?” to “what should be concluded?”: evaluative judgement.

Counts and metrics help, but it’s the analyst who weighs significance, context and rival explanations.

Moment 6: holistic judgement, the whole thing

The final judgement is about the whole finding.

Using vignettes to draft or check a holistic judgement

Example vignette output

> *Quotes 23, 44 and 45 provided by sources C and D largely represent coherent causal stories explaining the journey from increased knowledge to food consumption quantity, as follows:… One source links lessons from the organisation to increased output, then links that output to better household food consumption.*
>
> *"I now produce a lot and I have a good income from my products. I now produce and sell and grow well." - MNX-6*

The other half of the workflow: questions

Coding is half the job. The database under the map is for questions.

You can stack filters and you can answer many research questions, and most evaluation questions.

> Any map visualisation is one view.
>
> Behind it sits a queryable database of links, sources, quotes, tags and metadata.

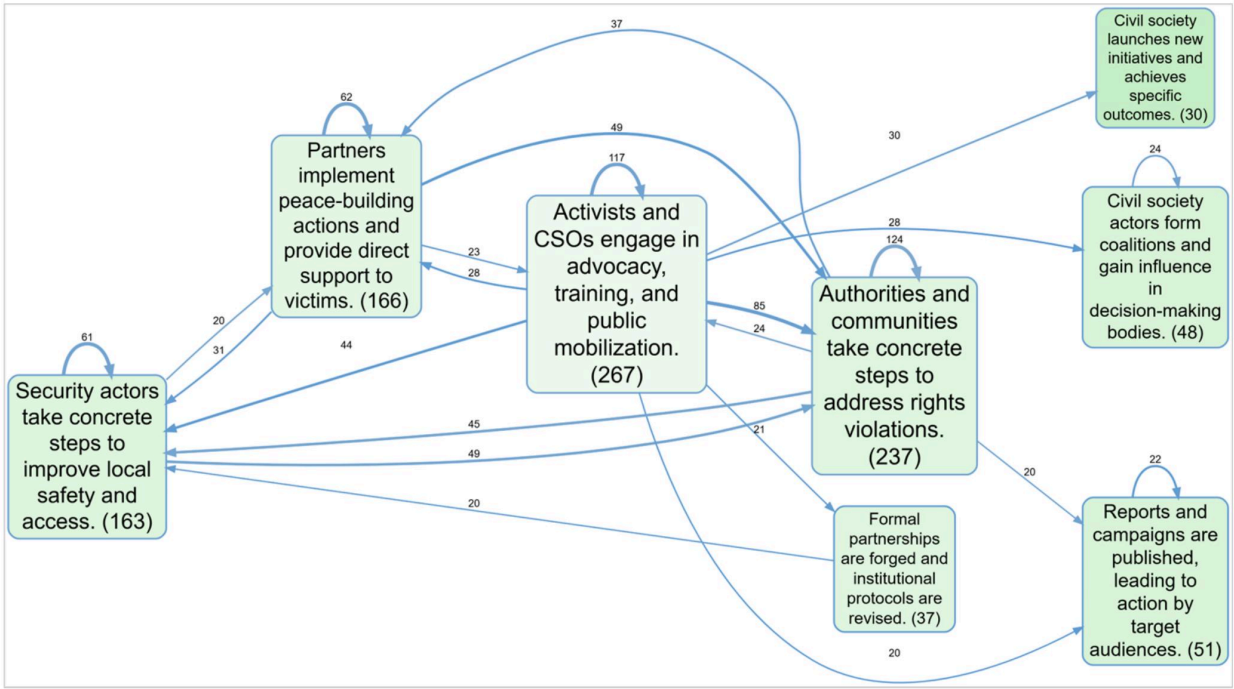
Case study: INTRAC, SCC programme

13 countries **692 outcomes** **5,430 claims**

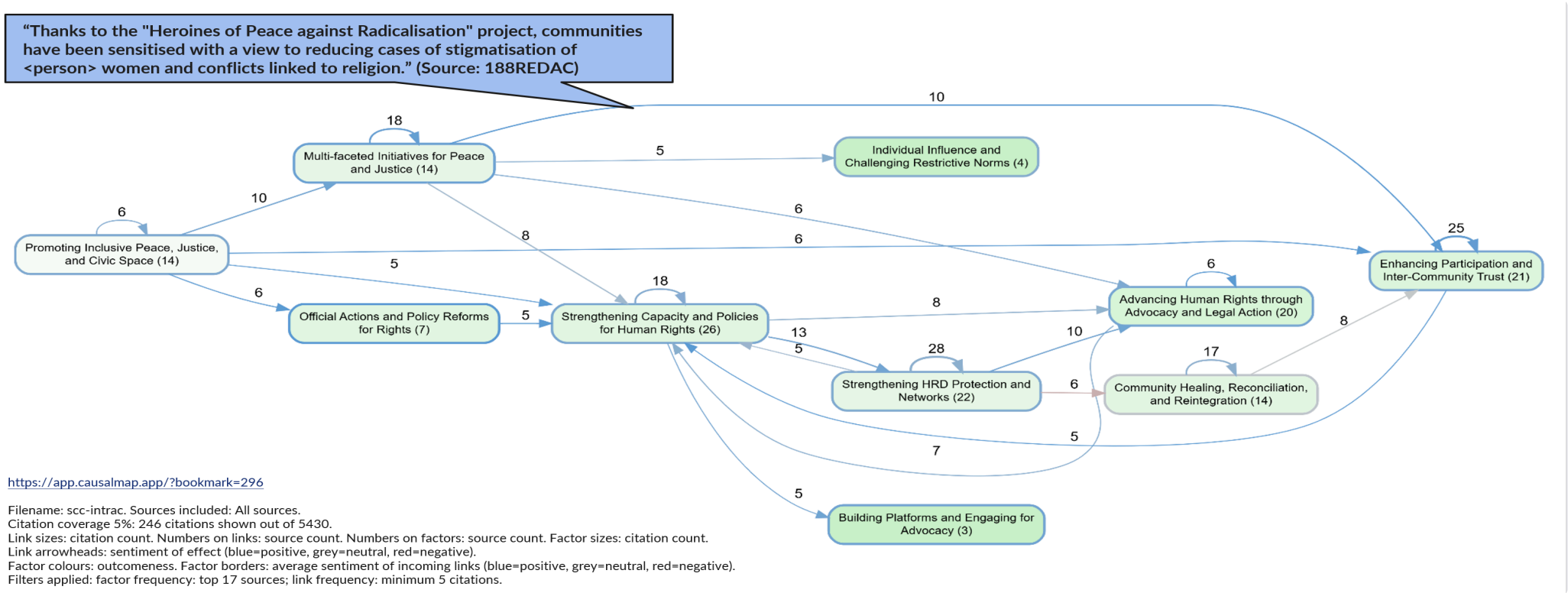
INTRAC was monitoring/evaluation partner for a 13-country civil-society programme. Outcome Harvesting gave rich data, but not the big picture.

> *Couldn't understand the big picture, for the whole programme but also for specific countries too.*
>
> **- Alastair Spray, INTRAC**

Zero-shot AI coding: no pre-set codebook. Outputs: one programme map, 13 country maps, filtered views by objective and time period.



INTRAC: same data, one country



Burkina Faso, country view

Advocacy, training and public mobilisation feed action on rights violations.

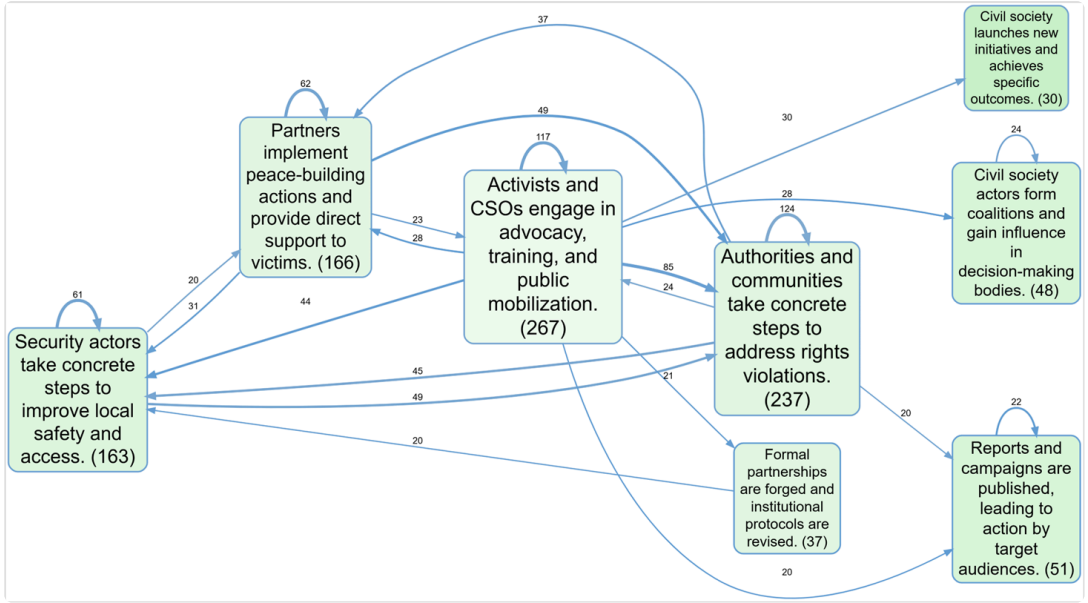
Two complementary uses

Raw cognition maps

For understanding the lived worlds of stakeholders for policy design, behaviour change and (social) psychology: what do people think drives what?

Assessed maps

For evaluation, high-stakes decisions, judgements about facts on the basis of claims.



INTRAC: pathways from interventions to outcomes in a 13-country lobby and advocacy programme.

Causal mapping and AI: a good fit

The method works fully without AI. Minimalist coding is a simple task which suits LLMs.

Validation study (Powell and Cabral, 2025), QuIP corpus, GPT-4 temperature 0:

84%
OPEN CODING
180 LINKS

87%
CODEBOOK-ASSISTED
172 LINKS

Composite score: correct endpoints, real causal claim, not hypothetical, correct direction.

Top-level overview maps from AI and human coding come out broadly the same.

It works because the prompt is narrow: “where is a causal claim, and what influences what?”

Three things to take away

1

The arrow question matters

The key difference between system mapping and causal mapping:

SD	CM
• arrows are facts	• arrows are claims
• variables drive one another via (linear) functions	• events/situations cause other events/situations
• single composite model	• many sources, traceable

2

Minimalism simplifies the task

Minimalist coding lets analysts aggregate, compare and query thousands of claims without semantic disputes.

Enables coding at scale.

3

Janus matters

Claims are not facts. Going from claims to facts needs checkable steps.

Resources

App (free for core functionality and unlimited projects, as long as they are public.)

app.causalmap.app

Knowledge garden

garden.causalmap.app

Bibliography

garden.causalmap.app/bibliography

Powell and Cabral (2025) AI-assisted causal mapping: a validation study. *IJSRM*.

Powell, Cabral and Mishan (2025) A workflow for collecting and understanding stories at scale. *Evaluation*.

Powell, Copestake and Remnant (2024) Causal mapping for evaluators. *Evaluation*.

Thank you. Questions, please.

hello@causalmap.app