

# Causal Mapping

*Steve Powell*

2019-11-25

## Contents

<b>Contents</b>	<b>1</b>
<b>1 What's in this guide?</b>	<b>2</b>
<b>I Introduction</b>	<b>3</b>
<b>2 Manifesto: piecing together fragments of causal information</b>	<b>4</b>
<b>3 Causal maps: a unifying idea</b>	<b>6</b>
3.1 Causal map <i>paradigms</i> . . . . .	8
3.2 Mini-maps with more than one influencing item . . . . .	9
3.3 Semi-formal definition of a causal map paradigm . . . . .	10
3.4 Hybrid sets of rules . . . . .	11
3.5 Causal <i>multi-maps</i> ? . . . . .	11
3.6 A brief overview of different causal map paradigms . . . . .	12
3.7 Summary . . . . .	14
<b>4 Causal mapping</b>	<b>16</b>
<b>5 The causal map app in four pictures</b>	<b>18</b>
<b>6 We need a “soft arithmetic” for causal maps</b>	<b>20</b>
6.1 An ambitious project . . . . .	21
6.2 Users of a causal map expect to be able to deduce some kind of comparative information from it . . . . .	21
6.3 Asking and answering those kinds of “typical questions” of a causal map boils down to assigning some kinds of numbers to its elements . . . . .	22
6.4 Aren't there strategies to encode causal information without using any kind of number? . . . . .	23
<b>7 We need rules about how to encode causal information in a causal map (and decode it again)</b>	<b>24</b>

<b>CONTENTS</b>	<b>2</b>
7.1 How do causal claims work within ordinary language? . . . . .	25
7.2 Understanding the elements of a causal map by agreeing how to make deductions with them . . . . .	25
<b>8 Soft arithmetic is also the answer to understanding causal maps</b>	<b>27</b>
<b>9 Metalanguage: some words for talking about causal maps</b>	<b>28</b>
9.1 “The cause”, “a cause”, “(causal) influence” . . . . .	28
9.2 Nodes, variables, vertices . . . . .	28
9.3 Influence variable, consequence variable, package of variables .	29
9.4 Mechanism, theory . . . . .	29
<b>II The rules of Soft Arithmetic for causal maps</b>	<b>30</b>
<b>10 The mini-map coding rule</b>	<b>31</b>
10.1 Meaning depends also on elicitation context . . . . .	32
<b>11 The mini-map coding rule – functional form</b>	<b>34</b>
<b>12 The extras rule: adding extra information, in particular about the levels or values of the variables</b>	<b>36</b>
12.1 Interpretation . . . . .	37
12.2 Corollary: Ordinary reasoning . . . . .	37
<b>13 The juxtaposition rule</b>	<b>39</b>
<b>14 Rules for joining maps</b>	<b>41</b>
14.1 The chaining rule . . . . .	41
14.2 The shared consequence rule . . . . .	42
14.3 The shared influence rule . . . . .	43
14.4 The shared arrow rule . . . . .	43
<b>15 Rules for joining maps: what counts as “the same” variable?</b>	<b>46</b>
<b>16 The chaining rule, functional form (zooming out / black-box rule)</b>	<b>47</b>
16.1 Two special cases . . . . .	48
16.2 Technical note . . . . .	49
<b>17 The shared consequence rule, functional version</b>	<b>51</b>
17.1 Contradictory combinations . . . . .	53
17.2 But which case is which? . . . . .	54
17.3 Extension to causal <i>packages</i> . . . . .	55
<b>18 The shared influence rule, functional version</b>	<b>57</b>
<b>19 The shared arrow rule, functional form.</b>	<b>58</b>
<b>20 The rule for merging arrows. Weight of evidence.</b>	<b>60</b>
20.1 Problems with evidence . . . . .	61

<b>CONTENTS</b>	<b>3</b>
<b>21 The chaining rule with loops</b>	<b>63</b>
<b>22 The rule for conceptual links</b>	<b>64</b>
22.1 Variations . . . . .	65
22.2 Problems . . . . .	65
<b>23 Combining “extra” information like the values or levels of variables</b>	<b>66</b>
<b>24 Context</b>	<b>67</b>
24.1 Intersections . . . . .	67
24.2 Replacing contexts with variables . . . . .	67
<b>25 Maybe the nodes are not variables, maybe they are ... propositions, events, schemas?</b>	<b>69</b>
25.1 Events . . . . .	69
25.2 Propositions . . . . .	69
25.3 Schemas / Schemata . . . . .	70
<b>26 Variable-based and propositional-based approaches to causal diagrams: chalk and cheese??</b>	<b>71</b>
26.1 The variable-based approach . . . . .	72
26.2 More on the proposition-based approach . . . . .	73
26.3 Statistics? . . . . .	79
26.4 Conclusion . . . . .	79
<b>27 Limitations to the proposition-based approach</b>	<b>80</b>
27.1 Problems <i>within</i> the proposition-based approach . . . . .	80
27.2 Limitations: Problems with <i>extending</i> the proposition-based approach . . . . .	82
27.3 Weaknesses of the variable-based approach . . . . .	84
27.4 What is to be done? Suggestions for QuIP . . . . .	84
<b>28 Clarifying how to ask QuIP questions in contexts when the final outcomes are not necessarily “changes”</b>	<b>85</b>
28.1 Classic QuIP: the focus on changes . . . . .	85
28.2 Even final outcomes don’t have to be changes . . . . .	85
28.3 Preceding items don’t have to be changes either . . . . .	87
28.4 How to focus on the relevant contrast . . . . .	87
28.5 Drivers . . . . .	88
28.6 Focus on initial items (“drivers”) . . . . .	88
<b>29 Coding individual causal fragments using propositions (revised)</b>	<b>90</b>
29.1 Basic coding rule . . . . .	90
29.2 Mini Extension: causal chains . . . . .	93
29.3 Extension: propositional claims about <i>additive contributions</i> . . . . .	93
29.4 Examples we can code . . . . .	96
29.5 Examples we cannot code . . . . .	97
29.6 That’s it . . . . .	98

<b>30 Combining causal fragments from different sources using propositions</b>	<b>99</b>
30.1 Same item . . . . .	99
30.2 What these extensions still don't do . . . . .	102
<b>31 Coding causal maps with propositions</b>	<b>106</b>
31.1 Basic coding rule . . . . .	106
31.2 Mini Extension: causal chains . . . . .	109
31.3 Mini Extensions: certainty and trust . . . . .	109
31.4 How does classic QuIP extend this basic idea? . . . . .	109
31.5 That's still not enough . . . . .	113
31.6 Solution: propositional claims about contributions, and assume additive causation . . . . .	114
31.7 What these extensions still don't do . . . . .	116
<b>32 Types of variable</b>	<b>120</b>
<b>33 Causal thinking is essentially contrastive thinking</b>	<b>121</b>
<b>34 Lo/hi Variables, types of variable, and contrasts</b>	<b>122</b>
<b>35 “For each .....” variables</b>	<b>124</b>
<b>36 Coding specific influences – individually and in combination</b>	<b>125</b>
36.1 Summary of the argument so far . . . . .	125
36.2 Encoding causal influences between lo/hi variables . . . . .	127
36.3 Breaking down the problem . . . . .	128
36.4 Some “mono functions”: functions with a single influence variable	129
36.5 Some “multi functions”: functions for packages with more than one influencing variable: . . . . .	133
36.6 “Package-free” combinations . . . . .	134
36.7 Contradictions? . . . . .	141
36.8 Building up more functions from these building blocks . . . . .	141
36.9 Coding influences within in the app . . . . .	142
<b>37 One function to rule them all?</b>	<b>143</b>
37.1 One universal type of function for coding causal claims with exclusively lo/hi variables. . . . .	143
<b>38 Coding a claim about the absence of a causal link</b>	<b>145</b>
<b>39 Strength / importance</b>	<b>146</b>
39.1 Including the idea of strength in our functional equation . . . . .	146
39.2 Direct ways of eliciting how important is one variable’s contribution to another. (Fiona) . . . . .	147
<b>40 Causal Inference</b>	<b>148</b>
40.1 Specifying the map . . . . .	148
<b>41 Upstream (Bayesian) inference</b>	<b>150</b>

<b>CONTENTS</b>	<b>5</b>
<b>42 Contour</b>	<b>151</b>
42.1 Flipping . . . . .	151
<b>43 Valence</b>	<b>152</b>
<b>44 Valence and direction</b>	<b>153</b>
<b>45 Effect: two kinds</b>	<b>154</b>
<b>46 Interventions and differences</b>	<b>155</b>
46.1 Shortcuts . . . . .	155
<b>47 Contribution</b>	<b>156</b>
47.1 Why reporting correlation is not helpful . . . . .	156
<b>48 Information about the <i>source</i> of our causal information</b>	<b>157</b>
48.1 Applying the “trust” attribute source-by-source . . . . .	157
48.2 Including information about sources . . . . .	157
<b>49 Clusters of similar maps</b>	<b>158</b>
<b>50 Maps of maps</b>	<b>159</b>
<b>51 Probability density functions</b>	<b>160</b>
 <b>III Simplification</b>	 <b>162</b>
<b>52 Simplifying causal maps: aggregation and filtering</b>	<b>163</b>
52.1 Why simplify a causal map? . . . . .	163
52.2 The tables available . . . . .	164
52.3 The steps taken by the app when simplifying a network according to user commands and/or automatically if requested. . . . .	164
<b>53 Aggregating and filtering beliefs</b>	<b>167</b>
<b>54 Aggregation and filtering based on face value</b>	<b>168</b>
<b>55 Aggregation and filtering based on particular research questions</b>	<b>169</b>
<b>56 Aggregation and filtering based on metrics</b>	<b>170</b>
56.1 Network metrics . . . . .	170
<b>57 Citation intensity</b>	<b>171</b>
<b>58 Conspicuous absence</b>	<b>172</b>
<b>59 Not forwards</b>	<b>174</b>
<b>60 Homogeneity of paths</b>	<b>175</b>
60.1 (non-) solution 1) . . . . .	176
60.2 Solution 2) . . . . .	176

<b>CONTENTS</b>	<b>6</b>
-----------------	----------

60.3 Solution 3) . . . . .	176
60.4 Solution 4) . . . . .	176
60.5 Solution 5) . . . . .	176
<b>IV Visualising</b>	<b>177</b>
<b>61 Visualising and formatting causal maps</b>	<b>178</b>
61.1 Conditional formatting . . . . .	178
61.2 Hard-coded formatting . . . . .	178
61.3 Constructing labels and tooltips . . . . .	179
61.4 Focus . . . . .	179
<b>62 Coding using the UI (outdated)</b>	<b>180</b>
62.1 Show that the influence of one variable on another has a particular contour. . . . .	180
62.2 Show that variable X is part of the definition of variable Y . . . . .	180
62.3 Show that variables X and Y are linked by definition . . . . .	180
62.4 Show that a set of influence variables influence the consequence variable <i>collectively</i> . . . . .	181
<b>V Summary</b>	<b>182</b>
<b>63 Summary of the rules for inference in causal maps, aka “Soft Arithmetic”</b>	<b>183</b>
63.1 The inference rules for causal maps . . . . .	183
63.2 The mini-map coding rule . . . . .	183
63.3 Focus on “lo/hi” variables and functions between them . . . . .	184
63.4 Recording the actual values of variables . . . . .	184
63.5 The rule for conceptual links . . . . .	185
63.6 The rules for coding different types of influence; single influence variable . . . . .	185
63.7 The rules for coding different types of influence; packages of multiple influence variables . . . . .	186
63.8 INUS . . . . .	187
63.9 Coding the strength of the influence of a package . . . . .	188
63.10 Causal claim coding form . . . . .	189
63.11 Problems . . . . .	189
<b>VI The why question</b>	<b>192</b>
<b>64 The “why question” as a generic method in social research</b>	<b>193</b>
64.1 Steps to asking “the why question” . . . . .	193
64.2 When would you <i>not</i> want to use this approach? . . . . .	194
<b>VII The why question</b>	<b>196</b>
<b>65 The “why question” as a generic method in social research</b>	<b>197</b>

<b>CONTENTS</b>	<b>7</b>
65.1 Steps to asking “the why question” . . . . .	197
65.2 When would you <i>not</i> want to use this approach? . . . . .	198
<b>66 The “<i>why</i> question” – asking about changes</b>	<b>199</b>
66.1 “Changes” . . . . .	199
 <b>VIIIsing the causal map app</b>	<b>200</b>
<b>67 Using the causal map app – already partly outdated!</b>	<b>201</b>
67.1 About the tables . . . . .	201
67.2 Left-hand tabs for producing and coding your diagram . . . . .	201
67.3 Right-hand tabs for outputs: viewing diagram etc . . . . .	207
<b>68 Using the causal map app for real-time, collaborative theory construction</b>	<b>209</b>
 <b>IX Case studies</b>	<b>211</b>
<b>69 Case Study: the Strawberry Line</b>	<b>212</b>
<b>70 Case study: Global Young Academy. Tracing the paths of GYA’s impact</b>	<b>214</b>
70.1 More details . . . . .	216
<b>71 Appendix: previous work</b>	<b>222</b>
71.1 From “Theorymaker” to “Causal Mapping” and from “Theories of Change” to “Causal Maps” . . . . .	222
71.2 Articles . . . . .	222
71.3 Presentations . . . . .	222
71.4 Poster . . . . .	223
71.5 Longer blog posts . . . . .	223
71.6 Resources and apps . . . . .	223
<b>72 References</b>	<b>224</b>

## Section 1

### What's in this guide?

A rough guide to the nascent theoretical basis behind the causal map app.

We'll look at:

- what is causal mapping?
- the need for the app
- a guide to the app
- tips on coding

There will also be general materials on harder issues in the coding and meaning of causal networks...

In the first, theoretical, part of this guide I'll talk about causal maps in general, what the different types have in common, how we can encode causal information in them and how to extract it again.

In the second, practical, part I'll make some concrete suggestions for best practices in creating and presenting causal maps. It will be based on our app which we hope encapsulates this approach, but it should make sense however you are creating your maps.

Why a new tool? There are already plenty of tools for storing and presenting causal structures, primarily of two types

- graphical e.g. Kumo
- theoretical e.g. dagitty.

What we lack are tools of either type which help us to construct

- models which include other people's models. These are really important if we want to model the behaviour of people, groups or systems which have their own implicit or explicit Theories of Change
- different versions of the same or overlapping models, e.g.
  - different stakeholders' (maybe partial) information on the same causal topic
  - different editions or versions of a Theory of Change.

**Part I**

**Introduction**

## Section 2

# Manifesto: piecing together fragments of causal information

Gone are the days when we could think of data or information as primarily about *numbers*. Many of us who are involved in understanding the social world and evaluating interventions within it spend much of our day understanding, presenting and manipulating *causal structures* and even *models of other people's causal structures*.

The fundamental, radical points are these:

- **Causal information is primary information.** It isn't something which exists only virtually as a potential conclusion on the basis of observations of non-causal variables. There is a fact of the matter about what causes what, just as there is a fact of the matter about the number of people on a training course.
- Parallel to that, **humans' perception of causation is primary**, as primary (and fallible) as our perception, say, of colour. All the things which we know, or think we know, about our world – from the colour of that dress to the way the wind shakes the trees – have already been through a lot of cognitive processing, and none of it is “secondary”. So when we ask stakeholders the “why question” (what causes what in this domain), we are not asking them, first and foremost, about what they deduce from their (non-causal) observations in the way we might as scientists or researchers. That would be a very shaky method; we might be better just to ask them to tell us about their non-causal observations and we could try deducing causality themselves. No, we are asking them about what causes what based on their underlying understanding of the causal structure of their world, which they have pieced together in a number of different ways. That underlying understanding might be somewhat erroneous, but it's certainly a lot better than ours.

As James Copestake (Copestake, Morsink, and Remnant 2019) says:

“.... attribution claims underpinning the QuIP do not require a control group, nor indeed variation in exposure to the intervention across the sample of respondents interviewed. Rather, causal

claims rely on the integrity of ‘within-case’ statements made by respondents themselves”.

## Section 3

# Causal maps: a unifying idea

Social scientists and programme evaluators often use a particular kind of diagram to represent causal relationships: boxes joined by arrows, to show what causes what. In this post we revive the term “causal map” as a broad name for this kind of diagram. Causal maps can be found throughout a whole range of traditions, from strictly quantitative to thoroughly qualitative. This fact alone makes the concept interesting as a unifying idea.

Causal maps are not peripheral things. They are central to much research and evaluation: a way of expressing the main findings, what we really wanted to know: does X causally influence Y? Did B contribute to C? What else has to happen? How sure can we be?

In this post we will briefly look at some different, overlapping approaches (which we will call “paradigms”) for drawing and understanding different kinds of causal map. Some of these paradigms are well developed, such as those coming from fuzzy logic and causal statistics, some are less clearly defined, such as the idea of a “program theory” in evaluation; but, as causal maps, they all have a few key points in common.

This matters to us because we’d like our causal map app to be useful for people working with maps from a wide range of different paradigms.

A rough definition:

A **causal map** is a model of causal relationships, in the form of a directed graph in which the items (nodes, elements) are linked by arrows, together with translation rules which tell us how to interpret the arrows, namely that an item with one or more arrows pointing to it is in some sense *causally influenced by* the item(s) at the start of those arrow(s).<sup>1</sup>

This definition says that a causal map comes with explicit or implied translation rules which let us go from it to a set of causal claims and back again. These rules are a bit like a “legend” at the side of the map. The key thing about the

---

<sup>1</sup>In most cases the model is literally drawn in the form of a network, but other ways of representing the same ideas are possible too, for example a set of narrative claims.

definition is that the fundamental translation rule is expressed at the level of the individual causal pieces, which we call “mini-maps”.

A (causal) **mini-map** is one item together with all the items directly pointing to it, i.e. directly influencing it.

The mini-map rule helps us translate “drought  $\rightarrow$  hunger” to the equivalent causal claim (“drought causally influences hunger”), and back again. Larger causal maps are built up from bunches of mini-maps.



Fig 1. A mini-map

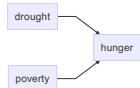


Fig 2. Another mini-map

Different kinds of causal maps can have radically different translation rules: what precise flavour of causation is meant? The causal arrow might mean “...completely determines...” or “...is sufficient for...”. Or arrows might have numbers associated with them, and the arrow might be interpreted as something like “...increases by a factor of .5 the probability of ...”. In most causal maps which programme evaluators have to deal with, the meaning of these arrows can be mixed and unclear, often meaning just “... causally influences in some way ...”

The *elements* within different kinds of causal map (like “drought”) differ correspondingly. In some of the cases above, the elements might be *propositions* like “drought happened”; in others, they are continuous variables; other kinds are possible too.

But a causal map's translation rules have to be explicitly causal. As Pearl points out, a genuinely causal arrow can be understood as something like “if you *do* X, Y will happen (perhaps, with probability  $p$ )”, whereas, for example, Bayesian belief networks are not causal maps because the arrows say “if you *observe* X, you will observe Y (perhaps with the probability  $p$ )”. Causation is not correlation.

To make them useful, causal maps nearly always come with additional *combination* rules which tell us how we can piece together a larger map out of smaller ones, and how to understand it on the basis of our understanding of its constituent parts. We can use these rules for example to make deductions about maps or to derive a simpler version of a larger map. As both Realist theorists and the Pearlian school point out, our real-life knowledge is composed of relatively portable, transferrable small units like mini-maps.

For example, if we know that

$$\text{drought} \rightarrow \text{hunger}$$

and

$$\text{hunger} \rightarrow \text{revolution}$$

can we deduce

$$\text{drought} \rightarrow \text{revolution?}$$

If we are told to interpret “ $\rightarrow$ ” as “completely determines”, this is arguably an acceptable conclusion; likewise with “may have some causal influence on”. But if we interpret it as “increases the probability of X by 50%”, this conclusion doesn't work, because we are trying to describe a partial influence which we expect to *attenuate* as we go down the causal chain. Concretely, we'd then expect the influence of drought on hunger to be less than 50%.

So, the explicit or implicit rules which come with a causal map also help us to make *deductions* with maps.

### 3.1 Causal map *paradigms*

We don't have to set up these kinds of rule for understanding causal arrows every time we draw a causal map. Instead, we make use of an established *paradigm*: we only need to say “this is a structural equation model” or “this is a fuzzy cognitive map”, and the rules are understood. A causal map paradigm is just a frequently-used set of rules for understanding and manipulating a whole genre of causal maps.

Particular kinds of causal map may have any number of additional properties and characteristics. For example, the arrows might include information about the degree of certainty we have in that causal link. And the corresponding

mini-map rule and combination rules might help us to do deductions with that kind of information. For example, if we are only partially certain that X causes Y and only partially certain that Y causes Z, the rules should tell us that we are even less certain that X causes Z.

Many causal maps are not just timeless statements of connections; they also include factual information about how things actually are. For example, this mini-map:

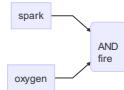
The horse kicked me  $\rightarrow$  I screamed

probably contains not only the background information that horses kicking people makes them scream, but also the information that both things actually happened. We call these maps “propositional” because they also contain this kind of factual information.

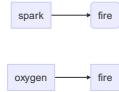
### 3.2 Mini-maps with more than one influencing item

Many causal map paradigms assume that causal influences are always between a single cause and a single effect. This is a very useful simplification. But sometimes we need to take into account causal connections in which something is affected by sets of more than one influence. For example, what if we want to draw a causal map to show that you need both oxygen *and* a spark to cause a fire? This can't be assembled from two separate diagrams:

We need a diagram in which “fire” has two different things pointing to it at once, and we need to note down how those two things interact, e.g. by writing AND where the arrows join.



This can't be expressed like this:



### 3.3 Semi-formal definition of a causal map paradigm

A causal map paradigm is a set of rules for constructing and understanding causal maps. It should contain as a minimum:

A **mini-map rule** defining what counts as a “mini-map” within this paradigm (roughly, a “consequence” element with one or more “influence” elements pointing to it); it may have other attributes too like factual information about the actual state of things, or the strength of the influences, the amount of trust in the information, etc. The rule has two parts saying (syntax): any such mini-map counts as a causal map and (semantics): how to translate such a mini-map into a particular kind of causal claim, and back.

**Combination rules** showing how to build up larger maps. The rules each have two parts saying (syntax): mini-maps joined in such-and-such a way also count as causal maps and (semantics): what does the composite map mean in terms of the meanings of its elements?

There are some important combination rules for joining two mini-maps. These different ways of combining elementary maps into composite maps embody some interesting social science questions.

- joining two maps on a shared item (what counts as the same thing? is “hunger” according to one person’s report the same as “starvation” in another?)
- joining two maps on a shared consequence item (this is surprisingly difficult in the general case which includes problems of multi-causation or overdetermination). Many paradigms make the assumptions that different influences can in some sense just be added together, so that the influence of X on Z is always independent of the influence of Y on Z. But other paradigms (like perhaps QCA) specialise in combining influences which are not independent of one another
- joining two maps on a shared influence item (usually unproblematic)
- merging several co-terminal arrows from different maps into one (what does it mean when we have overlapping information from different sources about the same causal influence? how do we combine them if they agree or if they disagree?)

- whether or not to allow joining two mini-maps in such a way that a loop is created, and if so, what does this mean? Fuzzy Cognitive Maps deals with this kind of composite map.

Plus, individual paradigms will have their own rules about how to combine any additional information which they allow for (e.g. level of trust).

### 3.4 Hybrid sets of rules

In real life, programme evaluators often have to deal with maps set up with hybrid rules. We might have to combine, say, information from publications in the field with data from a randomised controlled trial and information from stakeholder interviews into a larger “programme theory”. In this case we might have to resort to a mixed toolbox containing a larger set of rules.

### 3.5 Causal *multi-maps*?

In the approaches listed above, the information underlying them can usually be considered to come from just one source: one book, or one experiment, or one survey, or one expert; or the question “what is the source for that” isn’t even raised at all. We can call these kinds of maps “causal *mono-maps*”. They are particularly interesting for us at Bath SDR because they are at the heart of the QuIP methodology. Our causal map app is specialised in dealing with them.

A causal *multi-map* is a kind of causal map in which information about the source of each causal link (or more generally of each mini-map with several influencing items) is firmly attached to it and taken very seriously. The rules for this kind of map need to take this information into account, as in the following example.

If the homoeopath tells us that the right potency of gold is a cause of a well functioning liver, and the doctor tells us that a well functioning liver is a cause of clean blood, does that mean we know that the right potency of gold is a cause of clean blood? No, we can’t necessarily deduce from:

H believes that X causally influences Y

plus

D believes that Y causally influences Z

to

X causally influences Z

.... and we can’t even necessarily deduce this:

Some people believe that X causally influences Z

There are lots of similar challenges in joining together fragments of causal information. If a reliable source tells us they are almost certain that the link between X and Y is strong, whereas a hundred less reliable sources tell us they believe that there is no such link, how do we combine that information?

Is there a plausible set of rules for amalgamating fragments of causal information? Can those rules be adapted to cope with the different kinds of paradigms mentioned above?

We haven't found many formal approaches to this solve this problem directly (Markiczy and Goldberg (1995)).

The term "**causal mapping**" has often been used to acknowledge the possibility that causal information may come from different sources and explicitly opens the challenge of how to meaningfully aggregate that information.

### 3.6 A brief overview of different causal map paradigms

#### 3.6.1 Approaches already called "Causal Mapping"

There are a number of approaches which have been explicitly used the terms "causal maps" or "causal mapping" Laukkanen and Wang (2016), Laukkanen and Eriksson (2013), Bryson et al. (2004), Nadkarni and Shenoy (2004).

Causal mapping in this sense is loosely based on "Concept mapping", which has itself been used widely in some areas of evaluation. A special edition Trochim (2017) 2017 presents it as a relatively standardized area with standard steps (preparation, generation, structuring, representation, interpretation and utilization).

"Causal (aka cognitive) mapping (CM) first emerged in political science (Axelrod, 1976) and organizational studies (Bougon et al., 1977) as an innovative method for operationalizing and analysing the causal knowledge and beliefs of social actors." It is mainly used in project management.

These are amongst the few addressing problems of merging and aggregating maps from different sources

A rare and early exception is Markiczy and Goldberg (1995); they even derive a Distance Ratio method for measuring similarity between pairs of maps.

Scavarda et al Scavarda et al. (2006) discuss a method for creating a combining causal maps from multiple sources using a Delphi-like method. Their causal links have a default interpretation (if-then, although the meaning and restrictions of this are not further explored) and always have just one influence and one consequence variable. The method includes an algorithmic stage which uses hierarchical cluster analysis.

#### 3.6.2 (Comparative) Cognitive Mapping

#### 3.6.3 Theories of Change for a project or programme,

even (in a very restricted sense) Logical Frameworks

Some assume a sufficient connection, embedded in the mantras “if this then that” and “development hypothesis”.

### **3.6.4 Programme theories in theory-based evaluation**

#### **3.6.5 Fuzzy Cognitive Maps**

#### **3.6.6 Contribution analysis**

#### **3.6.7 Systems Diagrams**

#### **3.6.8 DAGs as promoted by Judea Pearl**

Judea Pearl (Pearl 2000) and colleagues - mostly statistics-based, though the theory is generalised to cover non-parametric cases.

Different kinds of causal maps have been well studied by statisticians, and recently Judea Pearl and colleagues have made giant strides to showing their utility as part of an approach to statistics which is better adapted than traditional statistics to dealing directly with causality rather than only with correlation.

#### **3.6.9 Structural Equation Models**

#### **3.6.10 Bayesian belief networks**

probably not, see Pearl (1986)

#### **3.6.11 (arguably) diagrams used in Realist Evaluation**

#### **3.6.12 (sometimes) diagrams used in Outcome Harvesting**

#### **3.6.13 Causal Maps as constructed in QuIP**

#### **3.6.14 Mental models**

Moon et al. (2019) from the field of conservation, with a lot of emphasis on complexity and systems. Also ask about what respondents say *should* be the case. Use 1, 2, or 3 for strength.

Elicits maps of the same domain from different stakeholders and compare in a narrative way.

“The progressive emergence of a shared vision can lead to a revision of assumptions (double-loop learning) and exploration of underlying values and beliefs (triple-loop learning)”

This approach is only monocausal/additive; it does not allow for multiple, interacting influences on a downstream variable.

There is a focus on different individuals’ different maps. They can be elicited with interviews, by drawing cognitive maps, Bayesian belief networks.

#### **3.6.15 Knowledge graphs, semantic networks**

Sometimes the phrase “knowledge graph” is used for any representation of knowledge in the form of a graph. In that sense, our causal maps are knowledge

graphs. See also “ontologies” in philosophy and computer science — tension between formal and practical interests (e.g. Google’s own “knowledge graph”).

The term “semantic network” is used for networks in which there is a wide variety of types of relations, whereas a “knowledge graph” has a very restricted set of relations.

The task of calculating the values of downstream nodes is very similar to calculating downstream activity levels in a neural network. One difference is that in a neural network, the contributions of each parent neuron to a child neuron can be simply summed (and then usually transformed by a sigmoid-type function to bring the activity levels back in the range 0 to 1). In other words, there is a pre-engineered solution to the question of how to combine the influences.

### **3.6.16 NCA (Necessary Condition Analysis)**

This is one of many approaches which are not in essence graphical (because they focus on individual causal combinations rather than constructing larger edifices from those combinations Dul (2016) .) However they do address some relevant issues.

### **3.6.17 Influence diagrams**

### **3.6.18 Paradigms which include the aspect of belief**

Modal logic of belief, Dempster-Shafer theory

### **3.6.19 “System Effects”**

<https://www.lukecraven.com/system-effects/>

### **3.6.20 QuIP maps**

The raw output of a QuIP-type coding exercise is a set of directed paths or arrows between a set of nodes or factors, in which the arrows might have additional attributes like respondent characteristics, question number / domain etc. One or more arrows B, maybe also C, D, .... coming into a node E encode a causal claim about how B (and maybe C and D....) influence E.

### **3.6.21 Maps of maps**

There are causal maps which incorporate elements which are themselves embodiments of causal maps, as we need to do when trying to model stakeholder behaviour.

## **3.7 Summary**

A causal map is a directed graph which is intended to model causal relationships, in which the items (nodes, elements) are linked by arrows which mean that the item at the start of the arrow causally influences the item at the end of the arrow. This is a broad definition which covers many different existing

“paradigms” of causal modelling. In general, the individual causal connections in a causal map may be based on information from more than one source.

The **causal map app** is a tool which helps social scientists and programme evaluators to extract causal claims from texts (e.g. interview data) from different sources, and to combine and present it visually.

## Section 4

### Causal mapping

As programme evaluators, we often have to deal with heaps of causal claims, for example that X causes Y, or causally contributes to it. At Bath SDR, when we work with our clients to conduct a QuIP evaluation, the field researchers produce a pile of interviews in which key informants have been giving their opinions about what causes, or contributes to, what: a heap of causal claims. How to process this heap?

In this brief post we want first to revive an old name (“causal mapping”) for the task of dealing with this kind of heap, and also to talk about tools for doing it.

Programme evaluators (and social scientists in general) are often confronted with this kind of task. Maybe you have some open questions at the end of a questionnaire in which people are making causal claims. Perhaps you have to do a “meta-evaluation” in which you combine information from different more or less authoritative documents. Sometimes you might conduct a series of interviews in which you directly ask people questions like “what causes what” or “what contributed to this event”. You might even try to gather causal claims, and merge them straight away them into an overall picture, as a participatory process with a group.

Information about causes, or at least about causal contributions, or at least about people’s informed opinions about causal contributions; all of this is gold to a programme evaluator or social scientist. We want and need to be able to collect, store, process, combine, analyse and display this kind of information, and do it easily and well. In the evaluation literature there is less discussion of these kinds of task compared to the amount of discussion around the questions of how one should collect, store, process, combine, analyse and display *numerical* or even so-called *qualitative* information as usually understood.

Collecting, storing, processing, combining, analysing and displaying causal data, the set of tasks which we collectively call “causal mapping”, are different from the related tasks involved in dealing with ordinary numerical or textual data. To take one example: storage. We can store sets of numerical data in structures like spreadsheets. We can store a set of interview or other text data as a (physical or electronic) folder full of documents. Causal mapping may itself involve extracting causal claims from statistical or narrative

data, but how do we then store the extracted causal connections? The obvious candidate is as a “directed graph”: a set of boxes joined by arrows, to show what causes what (or what causally influences what). Each individual causal claim is represented by one arrow in the diagram (or in some cases, by more than one arrow).

Causal maps are obviously good for the tasks of *displaying* sets of causal claims, but we aren’t used to thinking of diagrams or networks also as a way of *storing* information. But they are; and there is a whole range of modern software tools such as graph databases which do just that, in which the basic unit of information is not a number or a sentence but a link between two or more items.

Causal maps are also useful as *summaries* of causal information, central to much research and evaluation: a way of expressing the main findings, what we really wanted to know: does X causally influence Y? Did B contribute to C? What else has to happen? How sure can we be?

Many of the different kinds of diagrams in use in evaluation and social science, from the strictly quantitative to the thoroughly qualitative, can be considered to be causal maps. For example, “theory of change” is a very popular name at the moment for a particular kind of causal map in which valued outcomes and the means to influence them are specifically included. “Structural Equation Models” as used by some evaluators in statistically-based evaluation can also be considered to be causal maps.

The fact that causal mapping is used in very different kinds of evaluation makes the concept particularly interesting as a way to unify some diverse approaches. The phrase “Causal Mapping” goes back at least to Axelrod (1976), and the idea of wanting to understand the behaviour of actors in terms of internal “maps” of the world which they carry around with them goes back to Kurt Lewin and the field theorists; in evaluation theory, “Causal Mapping” appears from time to time in both qualitative and quantitative writing about programme evaluation.

Anyone who has ever tried to store and combine fragments of causal information, for example to summarise the causal claims in a set of interviews or documents, will know that it is difficult without specialised tools. Often, the long-suffering evaluator ends up with a frustrating mashup of spreadsheets, scraps of paper and sellotape. The task becomes particularly difficult when it is important to keep information about the source of a causal claim firmly attached to it, in order to, for example, to be able to show later only the claims made by women, or older people.

(Fiona’s sellotape picture)

At Bath SDR, we’ve been working together with Steve Powell to develop the **causal map app**, a tool which helps social scientists and programme evaluators to extract causal claims from text sources like interview data, and to combine and present it visually. We are using it right now to process data from QuIP studies. Beyond that, we are really excited that having better tools for causal mapping will make it easier to solve some of the central problems facing programme evaluators. We hope to have a publicly available (and free!) version available for interested users in 2020.

## Section 5

# The causal map app in four pictures

First, you upload or paste the short text passages or “statements” which you want to analyse for the causal claims which they contain.

Next, you view each statement one by one, highlighting passages which contain causal claims. For each claim, you decide what is the cause and what is the

“  
3: The weather was very cold this year, and it affected the land, it affected the cultivation and irrigation. For this reason the harvest will be less this year . This year we may have less crops: grain and potatoes, and other vegetables. We have used fertilizers, we buy seeds, purchase it also with money. Our harvest is only sufficient for household consumptions.



The screenshot shows a statement in a blue box: "3: The weather was very cold this year, and it affected the land, it affected the cultivation and irrigation. For this reason the harvest will be less this year . This year we may have less crops: grain and potatoes, and other vegetables. We have used fertilizers, we buy seeds, purchase it also with money. Our harvest is only sufficient for household consumptions." Below the statement is a toolbar with icons for copy, cut, and delete. Below the toolbar is a row of colored boxes containing words: "cold weather" (blue), "poor cultivation and irrigation" (green), and "fertilizers" (orange). To the right of these boxes are icons for "X" (purple), "B" (orange), and a "next" arrow (grey).

effect.

Update		text	code	st
1		It has improved on average, but not that much. The road was not better, no the roads are open, life is improving. But we are far from the centre. We have clubs, we have medical point, but we do not have experts.	LY-1	
2		The schools have good conditions, but we do not have English teachers. We buy necessary materials. Early child development is functioning. For children starting from 6 month, once in a week.	LY-1	
3		The weather was very cold this year, and it affected the land, it affected the cultivation and irrigation. For this reason the harvest will be less this year. This year we may have less crops: grain and potatoes, and other vegetables. We have used fertilizers, we buy seeds, purchase it also with money. Our harvest is only sufficient for household consumptions.	LY-1	
4		We mainly consume this food, and also purchase from the shops. We purchase potato,	LY-1	

Figure 5.1: 1573059919382



Figure 5.2: 1573060610020

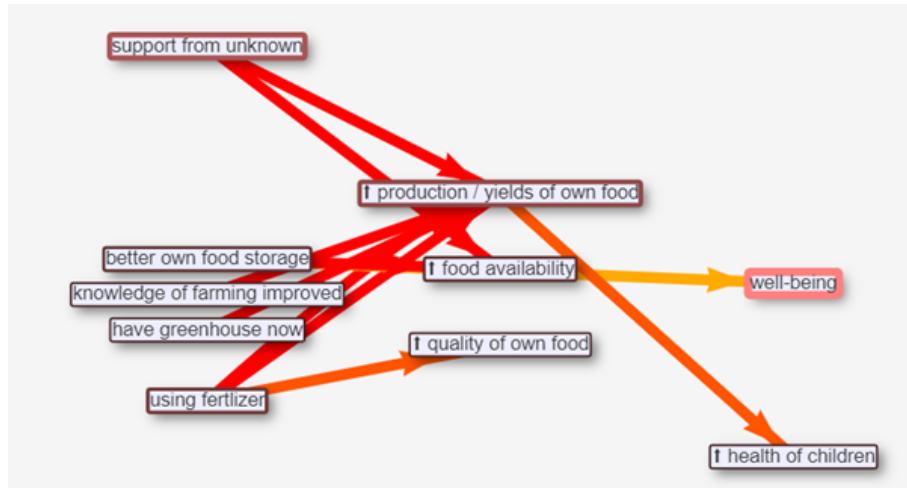


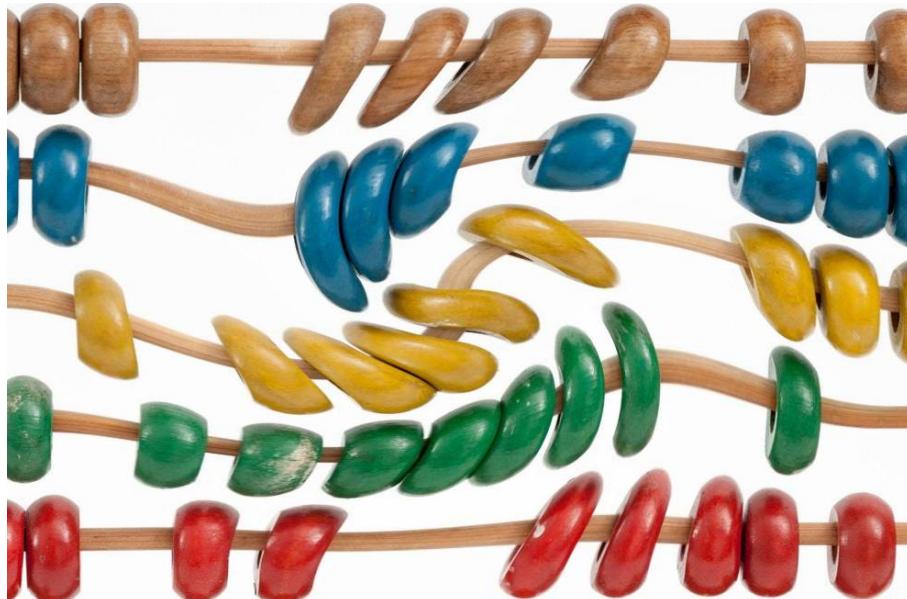
Figure 5.3: 1573061020009

Each causal claim becomes part of a small, “live” causal map in the right-hand window which grows as you work.

Last, you can view the whole map built up from all the causal claims you have been reviewing. You can search, change, edit and simplify the codes you have been using, as well as getting an overall picture: what caused what in these stories? You can also conduct more sophisticated analyses and produce graphics and reports, answering questions like: *did the stories which the women told differ from those which the men told?*

## Section 6

We need a “soft arithmetic” for causal maps



The forthcoming sections are the main theoretical contribution of this guide. They are an attempt to put together a kind of “soft arithmetic” for causal maps: a set of rules which tell us how to combine different pieces of causal information and how to make deductions with them. To do so, we have to address a wide and challenging range of issues. Many of these issues are familiar to practitioners like QuIP coders, indeed to anyone who tries to piece together fragments of causal information, but in most cases there is no consensus on how they are to be answered. I’ll be arguing that **if we want to try to address even the most basic needs of evaluation users, we are compelled to try to make some use of some sort of numbers when encoding causal information** – but in such a way as to keep things as fuzzy as they really are, rather than trying to demand or claim an unrealistic level of precision.

## 6.1 An ambitious project

Yes, this attempt to find a common logic behind the different approaches – from Theories of Change to Structural Equation Models – that come under the rubric of causal maps – see the list here – is ambitious.

## 6.2 Users of a causal map expect to be able to deduce some kind of comparative information from it

On the basis of a causal map, we must in principle, at least some of the time and with suitable qualifications and caveats, to be able to ask typical (evaluative?) questions like these:

Was the influence of C on E in some sense positive, or upwards, or increasing?

Was the influence of C on E not just theoretically present but of some meaningful size?

Is the influence of C on E bigger / more important than the influence of D on E?

Is C one of the biggest / most important influences on E?

Is the influence of C on E meaningfully bigger / more important than the influence of D on E?

... as well as, of course, questions about combinations, like:

Is the combined influence of C and D on E important?

Do both C and D act to increase E?

Will C affect E if D is high/present?

and so on.

In particular, we would like also to be able to ask them when E is not an immediate daughter of C but is further downstream of C.

(Going a step further, we’d like to be able to say things like “show me only the strongest influences in this map” or “show me a summary overview with only the most important things”. This kind of summarising and aggregation presupposes the ability to make statements like those above.)

I’d claim that these are among the central reasons clients commission QuIP reports, (or indeed most other kinds of evaluation), in summary:

- to get overview findings about the influence of C (an intervention) on some valued downstream factor like well-being — and in comparison with other interventions;
- and to get more in-depth details about the whole causal map, what causes what, what helps the intervention, what hinders it, and so on.

... again, we can have caveats and discussions about who decides, who participates etc; and we should tirelessly emphasise the high level of fuzziness and uncertainty involved. Often we will answer questions like those above with “sorry but we can't really be sure”. But still, if we *always* answer “there were a lot of links between C and E, but we have no further information”, we would not be doing our job. It would be like a telecoms agency delivering only metadata about the communications history of a criminal – certainly not useless, but a thousandth as useful as knowing what was actually said, whether in detail or in summary. Also we would not be doing the main job of an evaluator, according to Michael Scriven (2012) which is not just to describe what happened but to (help) judge is that *good enough*.

By all means we can let our evaluation stakeholders play with the encoded raw data and let them “draw their own conclusions”. But in order to draw their conclusions they have to know at least implicitly *how to draw conclusions from causal maps*. They might have better information than we do about the *content*, but our job as experts in causal maps is to advise them specifically on *how to draw conclusions*, how to make decisions about which parts to focus on, how to summarise them, and which parts to filter out.

### 6.3 Asking and answering those kinds of “typical questions” of a causal map boils down to assigning some kinds of numbers to its elements

The point is that these statements all presuppose, make use of, a kind of “soft arithmetic” for causal maps. To be able to make any kind of “bigger than/smaller than” claims about causal maps, we need the individual arrows to have, as a minimum, a (probably hidden) property like, in some sense, “strength” and “direction” as well as a way of comparing those properties. That is, *we are already doing* a kind of vague arithmetic. To be able to make double comparisons like “The influence of B on E is bigger than the influence of C on E” we are essentially implying that the arrows have some kind of numbers attached to them. That is in a sense the definition of what numbers *are*: attributes which justify comparisons. It remains to be seen what kinds of numbers those would be.

But, you say, you really don't want to actually put numbers on the arrows or the variables on your maps, you really don't want to claim that this is an effect of size 7.51, or -102, etc.

*I* don't want to put numbers on the maps either. But numbers in some form have to be there in the background, however flexible and fuzzy and inaccurate, or we couldn't make useful comparisons.

Maybe you often try to withhold judgement when asked to summarise a causal map. But if you see a hundred thick and well-evidenced arrows from smoking to cancer, and only a few from other causes, you’d be wrong to stay silent. The point is that at least in extreme cases like that, you’re implicitly doing soft arithmetic; you’re using some ideas about how to make a summary on the basis of the amount of evidence and how to combine that evidence. But what are the implicit rules of this soft arithmetic which you must be using?

#### 6.4 Aren’t there strategies to encode causal information without using any kind of number?

You might say “ah, but I have a strategy for encoding causal information which does not involve any kind of number or even any idea of strength or gradation – for example, C just contributes positively somehow to E, basta.”

By all means, there might be some such strategies, which have various severe limitations. I’ve addressed the issue here.

---

#### A technical note

The enormous advantage of quantitative approaches in social science, using ordinary integers and rational numbers, is that this arithmetic is pretty straightforward and mostly very well understood. It jumps over many of the issues which we have to tackle in the forthcoming sections.

But we have some advantages over the statisticians too. For one thing, the special status of *causal maps* within quantitative social science, and the fact that they are a completely different kind of creature from mere correlative structures, Pearl (2000), is still controversial in statistics.

Also, their central problem is “how can we get from these mere correlations to actual causal knowledge” whereas our causal maps already consist of (putative) causal knowledge; our problem is validating it, manipulating it, combining it, summarising it.

What I am trying to do in the next sections is apply a modest subset of Pearl’s ideas to the kind of ill-defined and fuzzy data we get in most actual social science, rather than the kind of quantitative data he (mostly) presupposes.

## Section 7

# We need rules about how to encode causal information in a causal map (and decode it again)

Causal maps are supposed to encode causal information. But how?

How should we encode — capture, write down, picture — causal information? How should we convert causal information — whether gathered from narratives or experiments or experts — into a causal map?

The reverse question:

How should we interpret a causal map when we see one?

These section will try to answer these questions in a reasonably general way, i.e. not just for specific kinds of map.

---

We learned in school how to write down and manipulate numerical information. But not causal information. We need to agree, and teach one another, how to write down and manipulate causal information too.

For the moment, we are not at all interested in which particular medium — for example, sentences as opposed to diagrams — we want to use to encode causal information, because we want to understand the rules which should govern the use of the symbols in *any* such encoding. (Spoiler: in fact we will be using diagrams containing arrows between boxes, basically, for this task.)

A social researcher or an evaluator might be set the task of encoding the causal information in some written or spoken information, say from an interview, and for that they will need symbols or conventions which make the causal information clearer.

We'd like to perhaps be able to use different symbols so we can specify these links more precisely, and add a corresponding legend to the map. But what symbols do we need? Some kinds of causal map already have their own sets of symbols and conventions, like SEMs. Do all the different kinds of causal map, like those in the list, all need to have their own legends? Can some be shared? The problem of what symbols to use is certainly a problem for the first three.

If we have to convert fragments of causal information expressed in other ways, like narrative testimonies from stakeholders about some domain of interest, into causal maps — a task we will refer to as “coding” — we need to know how to do that, what symbols and conventions we need to use to encode as much of the relevant causal information as we can.

So we have a challenge to establish rules, conventions, for constructing causal maps which makes their meaning transparent.

## 7.1 How do causal claims work within ordinary language?

Of course, people speak and write sentences which contain causal information all the time. But it isn't easy to see exactly what information is contained in such sentences, because in natural language we have a very messy and heterogeneous patchwork of ways of talking about causation: dozens of half-complete systems for encoding different variants of causal information, some of them suffused with ideas like blame and responsibility. Just as language contains overlapping, only partially consistent and mostly incomplete, ways of talking about time, and about space, and about quantity. Mathematicians and philosophers have nevertheless managed to systematise them in ways which are mostly satisfactory.

Causal language is not one system. But in order to be able to work with causal maps, we need to agree on a single set of rules about encoding and decoding causal information in them; and they should work as a good enough approximation to the ones we find naturally occurring in language.

## 7.2 Understanding the elements of a causal map by agreeing how to make deductions with them

Causal arrows are not conduits for a mysterious force caused “causation”. What flows down the arrows, so to speak, are actual forces - magnetism, peer pressure, greed, whatever. The arrows are the conduits for these *actual* forces: the form, not the content.

We will not get anywhere if we try to understand the meaning of a causal arrow, or how to use it, by using synonyms for “cause”.

Instead we will learn its “meaning” in precisely the way we learn the “meaning” of arithmetical symbols like + for addition: by learning the rules by which we may draw inferences between sentences containing such symbols.<sup>1</sup>

---

<sup>1</sup>If we were doing formal logic, we could say that we want to establish the *axioms* and *theorems* for sentences (or diagrams) involving the causal arrow.

*SECTION 7. WE NEED RULES ABOUT HOW TO ENCODE CAUSAL INFORMATION IN A CAUSAL MAP (AND DECODE IT AGAIN)*      32

I claim that this problem about *meaning* can be solved by agreeing on how to *reason* with causal maps, to do causal inference. The causal map app needs to have the right rules for causal inference in order to be able to bring the maps alive — so that we can ask them questions like “does variable B have more influence on E than variable C?” or give them instructions like “Hide all the variables which only have a small influence on E.”

When we learn the meaning of + in school, words don’t help us much, we learn by making deductions; we learn that from

$$x = 2 + 3$$

we can deduce, for example,

$$x = 5$$

according to certain rules. We could call them “inference rules”.

But in this guide, we aren’t dealing with equations, we are dealing with causal maps. In order to be able to explicate what the elements of causal maps mean, we need to agree on inference rules for causal maps.

As far as possible, we will try also to identify symbols and conventions which *fuzzily* capture the fuzziness, incompleteness, ambiguity etc of the way people talk about causal relationships and of the supposed causal relationships themselves.

## Section 8

# Soft arithmetic is also the answer to understanding causal maps

In section 6 I claimed that in order for our causal maps to be of any use to anyone, we need to, at least sometimes, attempt to make comparisons between different elements, and comparisons of comparisons, like “B is more important for E than C is”. This amounts to being able to assign some kind of numerical quantities to elements of the maps, so that we can make deductions with our maps following **rules which I call “soft arithmetic”**. In lots of occasions we won’t be sure enough to be able to make any such statements, but in others the evidence is so strong that we wouldn’t be doing our job if we didn’t.

Then in section 7 I claimed that we need rules for understanding causal maps: how do we turn causal information from stakeholders into features in a diagram, and in reverse, how do we read a diagram to decode the causal information. I argued that the best way to agree or understand these **translation rules for causal maps** is to show how to make deductions with them. Just as we don’t waste time in primary school explaining what “+” means; instead we show how to *use* it.

So it won’t be a surprise if I now point out the obvious: the two claims are one and the same, and the two sets of rules are one and the same. **The rules we will look at in the following sections will tell us both how to do “soft calculations” with causal maps and tell us what the elements mean, so that we can encode causal information into maps and back again.**

## Section 9

# Metalanguage: some words for talking about causal maps

We will need some metalanguage to talk about causal maps. Metalanguage is not part of the maps, it is a part of English which we use to talk *about* the maps. I've spent quite a lot of time trying to get the terminology right, I'm not making up words for the sake of it.

### 9.1 “The cause”, “a cause”, “(causal) influence”

I prefer not to use phrases like “the cause” or even “a cause”. For one thing, these are too mixed up with our human concerns like blame, and legal responsibility, and moral judgements. For another, these words are too specialised for binary, true/false variables. Also, they are too monolithic. For example, C is certainly not “the cause” of E below.

For similar reasons, in the diagram below I prefer to say that C *influences* or *causally influences* E. Indeed we can and should also say that the *package* of B and C together causally influences E. This makes it clear that we are talking about causation, not for example correlation, but that we are not of course claiming that E is completely determined by C or B.

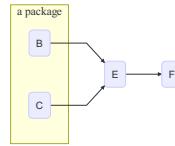
### 9.2 Nodes, variables, vertices ...

From a network perspective, the boxes can be referred to as nodes or vertexes. I will refer to these mostly as **variables**. However, calling them “variables” means that they are things which can vary, be different, for example the temperature outside now, or the extent of last season's rains; sometimes, variables don't seem like quite the right fit when coding causal information which talks about *events*. See later.

The lines are referred to as (directed) edges, paths or links. I will call them **arrows**.

### 9.3 Influence variable, consequence variable, package of variables

In a map like this, we call the variables on the left the *influence variables* of E, and E is the *consequence variable* of B and C. We can also say that B is a *parent* of E and E is a *child* of B. And we can say that both E and F are *downstream* of B and C, and that they are *descendants* of B and C, and so on. We can also say that B, C etc are members of a causal *package* – provided they are indeed coded together as such, see xx.



### 9.4 Mechanism, theory ....

It can be useful to distinguish between on the one hand “theories”, “theories of change”, “causal maps”, “representations”etc on the one hand and the aspects of the world which these things are true of on the other hand, which we can usually refer to as “mechanisms”.

## **Part II**

### **The rules of Soft Arithmetic for causal maps**

## Section 10

### The mini-map coding rule

Here is the first rule of Soft Arithmetic for causal maps.

Information like “the influence variables B, C and D all have some kind of causal influence on the consequence variable E” can be coded with a **mini-map** in which one or more variables (the “influence variables”) are shown with arrows leading to another (“the consequence variable”). The information and the map are equivalent.

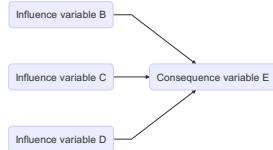


Fig xx. A “mini-map”: one or more variables are shown with arrows leading to one other variable.

In this extreme case, there is no information at all about the *content* of the causal statement from something as vague as “B and C and D all somehow contribute positively to E” to more detailed claims. Later we will see different ways in which we might encode that content. But we don’t in principle need any specific way to do that, we can just write what we know on the downstream box:

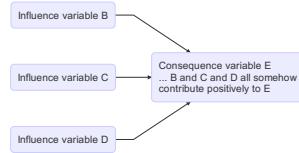


Fig xx: A mini-map in which the information about the *content* of the causal influence has been written on E, preceded by three dots . . . . The three dots indicate that what comes after is not formally part of our Soft Arithmetic.

Mini-maps are the atoms of causal maps. You can build up any causal network from them.

There are alternative ways of defining the “atoms” of causal maps, for example as multiple causal links forming a causal chain as QuIP does, or used to. Or you could also allow for more than one downstream variable within an “atom”. Often just single links between one influence variable and one consequence variable are treated as the atomic units – for example, Pearl does this. But this begs the question of what to do about causal influences which essentially involve two or more influence variables working together, for example synergistically.

Mini-maps, defined like as above, are our atomic units.

Many of the example mini-maps I will show contain several influence variables. In 90% of practical applications, 90% of mini-maps will just contain a single influence variable and a single consequence variable. But we don’t want to get stuck when we need two or more influence variables.

It is really important to understand that a mini-map can encode *any* combined influence of a causal package of variables on another, not just the kind of cases we expect when we come from a background of linear, addititative models. From that background, we might have to remind ourselves “aha, there might be interactions between the members of the causal package” but if you say this you are only showing that you have not yet shaken off the shackles of linear thinking, because the *usual* case is that when B and C are part of a causal packet, the influence of B on E depends on the value of C and vice versa. Independence is a *special case*.

## 10.1 Meaning depends also on elicitation context

Markiczy and Goldberg (1995) point out that the context in which the causal claims are elicited partly determines their meaning. It makes a difference in particular whether the respondents are asked “try to make sure you have included all the relevant factors” or not. It makes a difference whether a predefined set of concepts are used or an open-ended approach as with QuIP. Does the absence of an arrow mean that the respondent believes there is no such

causal link or merely that they hadn't thought to mention it? This question is crucial to how we merge data from different respondents, see later xx.

---

#### Technical note

You could say,

aha but the sentence above contains the words “causal influence” so you have explained one mystery with another. If you take that part out, the diagram could be about anything, and the arrows might mean “is larger than” or “is a child of”, or lots of things.

That's true, but the whole point is that we will specify not just this inference rule for causal maps, but enough different rules that only causality is left as a possible interpretation of their meaning. Or, to put it differently, if we have a child who can use + perfectly, we aren't bothered if they can explain it in words or not; if you understand the inference rules, you understand enough. If you want to test whether someone understands how to code and interpret QuIP information correctly, you can ask them to make various inferences with the maps.

Again: we are not going to say what “causal influence” means. We are going to show how it works.

Some of these rules will seem pretty trivial and obvious. They should. But they are necessary for building up a complete and consistent system. In any case I need them to make an app which actually works, and we need to spell them out so we can agree how to code with the app.

It's crucial that a mini-map codes information about causality, not co-incidence. So the causal map “ $C \rightarrow E$ ” should **not** be interpreted along the pattern of “if you observe (a high level of) C you are more likely to observe (a high level of) E”, though that may or may not be a corollary of the causal information. The strongest and most correct interpretation is “if you intervene in the system and manipulate C, which may involve breaking any causal links from other factors to C itself, then this manipulation will produce a corresponding effect in E”.

## Section 11

# The mini-map coding rule – functional form

The mini-map coding rule said that Information like “the influence variables B, C and D all have some kind of causal influence on the consequence variable E” can be coded with a mini-map in which one or more variables are shown with arrows leading to another.

This information can also be expressed in a functional form:  $E = f(B, C, D)$  which says, given the values of B, C and D, the influence rule or function tells us how to work out the value of E.

More generally, the function will in real-life cases not completely determine the value of E. Other, mostly unknown, factors are involved. With classical, linear functions we can just **add** the influence of  $f$  to the prior value of E. (Or conversely we can express the value of E as being a main prediction made by the function  $f$  together with additional “error” influences which can likewise be **added** to the prediction.) As my approach will be broader and will cover non-parametric and non-numerical cases, we cannot rely on this convenient convention. So we should more generally express the function like this:

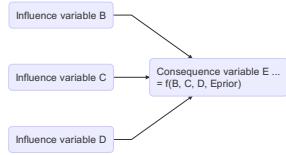
$$E_{posterior} = f(B, C, D, E_{prior})$$

. What this says is that  $f$  shifts the value of E away from our previous or prior best guess about E to a value determined not only by the influence variables but also that prior value.

This generalisation might seem very abstract but we will see that it is necessary to adequately present some quite basic functions like necessary and sufficient conditions.

The function statement, the written information and the map are all equivalent.

## SECTION 11. THE MINI-MAP CODING RULE – FUNCTIONAL FORM1



---

### Technical note

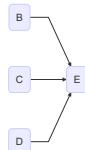
It is important to note that the = sign in these functional expressions is not the usual equality sign. For one thing, it is not symmetric. See Pearl (xx).

## Section 12

### The extras rule: adding extra information, in particular about the levels or values of the variables

Given a map and some extra information about the values of the variables and/or the nature of the influence, we can add that information, in ordinary English, to the arrows and variables

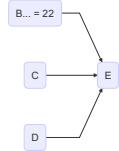
From this:



and, the English sentence asserted in the same context

“Variable B = 22 “

We can derive this (and vice versa):



In fact, we already did this when we added a plain-English description of the function to the label of the consequence variable in the mini-map coding rule

The three dots ... indicate that what comes after is not formally part of our Soft Arithmetic.

The same goes for any other kind of information in English (or any other natural language, of course) which tells us something about the state, or value, or level of one or more variables like or “Variable B is high” or “the level of Variable B is 22” or “Variable B is on” or even “Person B’s emotion is angry rather than sad”.

## 12.1 Interpretation

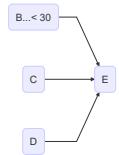
It just says, you can move information about the levels of the variables formulated in English sentences into similar statements on the variables themselves within the map, and back again. The extra information is *not part of soft arithmetic* in the narrower sense. We don’t (yet) have soft arithmetic rules for manipulating it, but we can use ordinary common sense.

## 12.2 Corollary: Ordinary reasoning

From the rules above it follows that you do not need an extra rule to do ordinary reasoning with causal maps. So from the above map you can also deduce the map below. (To check, remember that we could use the Extras Rule to convert the map into the map without the extra information plus the English sentence “B = 22” and from this we can use ordinary reasoning to get the English sentence “B < 30” which we can then combine with the map without the extra information to give the version below.)

*SECTION 12. THE EXTRAS RULE: ADDING EXTRA INFORMATION,  
IN PARTICULAR ABOUT THE LEVELS OR VALUES OF THE  
VARIABLES*

44

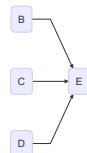


## Section 13

### The juxtaposition rule

Given two maps, derive the map formed by juxtaposing their contents into a single map. The context in which they are valid is the intersection of the contexts in which the original maps were valid.

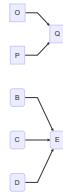
The next rule says that you can combine separate mini-maps into larger maps and vice-versa. So from this:



and, say, this



you can deduce this (and vice-versa):



#### 13.0.0.1 Interpretation

It seems so trivial it is hard to put into words. In short: you can combine two causal maps into one. If you know that these pink things influence those green things, and you know that these red things influence these blue things, then you know that these pink things influence those green things *and* that these red things influence these blue things<sup>1</sup>.

This all gets more interesting when the maps contain one or more common variables. This is the subject of the next set of rules.

---

<sup>1</sup>though this motivation via knowing about stuff is a cheap sell and not strictly true — it is a psychological claim and really we are not talking about psychology

## Section 14

# Rules for joining maps

Very often we will have different fragments of information which share common variables, which we can join up to create more comprehensive maps. In this section we will look at four different ways to snap together causal maps. These four are enough to build up arbitrarily complicated causal maps.

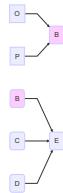
This section introduces the four cases from a purely *structural* perspective. The rules in this section will probably seem trivial. They basically all say “yes, of course you can join together maps on their common variables”.

In the subsequent sections we will look at the more tricky question of how the functions which contain the *causal content* of the mini-maps can be combined.

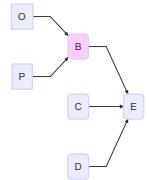
### 14.1 The chaining rule

Given two causal maps with one common variable which is an **influence variable in one map and a consequence variable in the other**, you can derive the single map which chains them together, *providing this does not create a loop*.

So from this:



you can deduce this (and vice-versa):



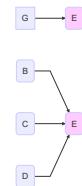
This illustration uses a mini-map with two influence variables combining with another mini-map with three influence variables, but the number of influence variables is not important. The same applies to the other rules in this section.

We will deal with loops or cycles later.

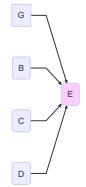
## 14.2 The shared consequence rule

Given two causal maps with one common variable which is a **consequence variable** in both maps, you can derive the single map which combines them.

From this ...



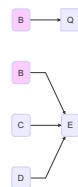
... we can derive this:



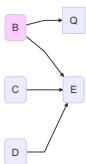
### 14.3 The shared influence rule

Given two causal maps with one common variable which is an **influence variable** in both maps, you can derive the single map which combines them.

From this:



derive this:



### 14.4 The shared arrow rule

Finally:

Given two maps which contain a co-terminal arrow, i.e. with a common influence variable and a common consequence variable, we can combine the maps on these two variables.

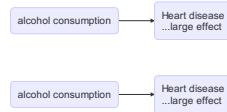
So if we hear:

heart disease has a causal effect on alcohol consumption

and again

heart disease has a causal effect on alcohol consumption

like this:



..... we can combine the mini-maps, like this:



### Technical note

We could present these rules more formally, to provide a recursive definition of what we mean by “causal map”, like this:

- If M is a mini-map (created by the mini-map coding rule), it is a causal map.

- If C is a causal map and M is a mini-map, the map constructed by joining C to M by any of the rules in this section is a causal map.
- Nothing else is a causal map.

## Section 15

### Rules for joining maps: what counts as “the same” variable?

We already saw that we will very often have different fragments of information which share common variables, which we can join up to create more comprehensive maps. In the previous section we looked at four different ways to snap together causal maps. They basically all say “yes, of course you can join together maps on their common variables”. However this does open up the question, critical for people doing actual coding, of *what counts as a common variable*.

We also need to deal with different possible formulations of variable names using what you could call different *parts of speech* like “more investment” / “amount of investment” / “investment increased 10%” / “investment will increase 10%”.

[Need to add more here!]

## Section 16

# The chaining rule, functional form (zooming out / black-box rule)

Given two mini-maps containing the paths  $E = f(B)$  and  $O = g(E)$  respectively, the **direct influence** of B on O is  $O = g(f(B))$ .

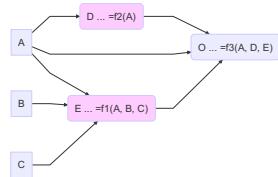
This rule tells us how to calculate downstream effects in a map which has been created using the chaining rule.

It is an elementary application of the “substitution rule” from secondary school maths.

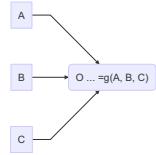
This problem is familiar from statistics. We can ask, if B affects E and E affects O, what is the **direct effect** of B on O? If we know  $f$  and  $g$ , we can trivially calculate  $g(f(B))$  simply by applying  $g$  to the result of  $f$ .

This rule, although it seems trivial, has wide practical applications. For one thing, it allows us to simplify causal maps by “zooming out” and omitting some of the detail.

So if you know the information in this reasonably complicated network:



.... you can in some sense “zoom out”, create a black box, subsuming the functional content of the pink boxes, to conclude this:



... where the function  $g$  can in principle be constructed if you know  $f_1$ ,  $f_2$  and  $f_3$ .

## 16.1 Two special cases

What's quite weird about this rule is that there are two extreme extensions of it which are in a sense opposites. These special cases are important because they are possible interpretations of the arrows in an arbitrary causal map in which no other information is given about the function involved.

### 16.1.1 Bare influence

Consider the very simple function  $H$  such that  $E = H(B)$  which just means, B has *some* influence, on E. Then from

$$E = H(B)$$

and

$$O = H(E)$$

we can conclude, using the rule, not only

$$O = H(H(B))$$

but we can also argue

$$O = H(B)$$

.. this does not follow from the rule but it should be possible to prove it: if B sometimes has some effect on E, and E sometimes has some effect on O, then B sometimes has some effect on O.

We can generalise this argument to cases with more than one influence variable. So  $E = H(C, D)$  means that there are some values of D such that making some difference to C makes a difference to E (it might be that most of the time, tweaking C does nothing to E), and similarly there are some, any, values of C such that tweaking D makes a difference to E, and so on.

### 16.1.2 Total control

Whereas in the second extreme version, the interpretation of the function  $H$  is “the influence variables all have *total causal control* of the consequence variable”. There is no room for E to do anything not dictated by B (and possibly, simultaneously by C, D etc).

So, if

B completely controls E

and

E completely controls O

we can conclude that

B completely controls O

---

## 16.2 Technical note

It is a bit naive to assume that if we know  $f(x)$  and we know  $g(x)$  we can always easily calculate  $g(f(x))$ . In fact, the mere fact that we know  $f(x)$  does not mean that it is easily calculable either. There are important caveats about computability and decidability which have practical consequences.

The zooming-out rule brings us to the edge of what seems like a debate about philosophical realism. For example, we could assert that in all practical applications, we can assume the reverse of the zooming-out rule: that with a more detailed investigation we could “zoom in” to any atomic relationship and find other, previously hidden, variables mediating the relationships. Or:

What counts as a “black box”, what counts as “revealing the underlying mechanism” is relative and context-dependent. You can always zoom in a bit further or zoom out a bit further.

Arguably this is counter to some interpretations of Realist approaches in evaluation and social science according to which there is a hard-and-fast distinction between the two: there is a superficial, black-boxy way of looking at things and a correct, Realist, zoomed-in view; and according to which there is a “fact of the matter” about mechanisms, which are real things waiting to be discovered, not constructions. (Of course there are other differences too between the two views, according to the Realist approach.)

In fact, the approach I am presenting here is not extensionalist. I do not claim that the zoomed-out causal map is equivalent to the zoomed-in version, even though by definition the designated upstream variables produce the same

*SECTION 16. THE CHAINING RULE, FUNCTIONAL FORM  
(ZOOMING OUT / BLACK-BOX RULE)*

56

effects on the designated downstream variables. For one thing, as Pearl points out, a genuinely extensionalist large cognitive map would be computationally impossible. No brain and no device could construct a single global causal map from all possible inputs to all possible outputs. Mini-maps and small aggregations of them are in fact the key to acquiring and using knowledge.

## Section 17

# The shared consequence rule, functional version

Suppose we have coded the fact that something (call it B) affects E according to one function  $f$  (e.g. it contributes positively to it), and we have also coded the fact that something else (call it C) also affects E according to another function  $g$ . The shared consequence rule says we can snap together these two pieces of information. That seems innocuous.

But also, according to the equally innocuous-seeming “mini-map coding rule”, the new map says that there is a *single* rule which (at least partially) determines E given information about B and C. Suddenly this is a big deal. What is that rule? How can we construct the correct function from  $f$  and  $g$ ?

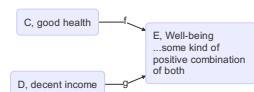
For example, if we have the information:

you can't have well-being without good health

and

income contributes to well-being

we can encode them like this:



What is the new causal function implied by this new mini-map?

..... we have two single-variable packages which influence a variable “hunger” or “amount of hunger”, the first perhaps expressing a necessary condition, the second perhaps expressing a continuous relationship. How can we combine this information to give us a single function of the form

$$\text{hunger} = h(\text{drought}, \text{poverty})$$

What is  $h$ ?

In the special case of the “quantitative” paradigm of (rational numbers and linear effects), creating this function is assumed to be easy. The rule is that we simply add the two effects, and assume there is no interaction between them unless we are told so. We also assume that the effect of B on E does not depend on E.

But in social science we often cannot fulfil these stringent assumptions. For example, we may be working with binary/Boolean “false/true” variables, and the effect of B on E can depend on E, for example in the case of necessary conditions, see later xx.

Later we will make a lot of use of variables and functions which range between 0 and 1, so-called “lo/hi” variables, as in fuzzy cognitive maps. Here, if we repeatedly applied a function which increments E by an equal amount, the value of E would sooner or later exceed 1. So the increment has to get smaller as E gets larger, or at least get curtailed at 1. So we have to use a different way to combine functions.

In this section I suggest that there is no perfect solution to calculating  $h$ . However I do later make a suggestion.

We could ask whether the constituent maps are considered to complement or contradict one another; and I recommend understanding them as contradictory by default. That means that if someone says “drought drives hunger” we assume that they *are* making an exclusive claim, and *would* object to the simultaneous claim “poverty drives hunger”; although if there is both drought and poverty, we would probably expect the level of hunger to be higher than if there was only drought or only poverty.

Some kinds of influence like a necessary condition by their nature are liable to contradict others. So if C is necessary for E, B cannot be sufficient for it.

One alternative would be to assume that only when the maps which we are merging come from the same source<sup>1</sup> may we assume that the information is complementary. When they come from different sources, it is less clear how we should combine the information but perhaps we would prefer to default to treating them as contradictory.

---

<sup>1</sup>We will discuss the concept of a source in section xx

Sometimes we might feel that the information in the influencing packages contains a complete or a partial contradiction, in other cases not. Our common sense pulls us in different ways about what to do in different cases.

We could try

$$E = f(B) + g(C)$$

But even if arithmetical addition makes sense in this case, there is no reason to suppose that we can just “add” the influences. Why shouldn’t they interact so that the influence of  $f$  and  $g$  together is much stronger than either alone? On the other hand, why shouldn’t they work against one another? The right answer might equally be multiplication or anything else:

$$E = f(B) * g(C)$$

Or, perhaps the effect of  $g$  is higher when  $f$  is low, or vice versa? The problem is that we are missing the additional causal information about how to combine these two (sets of) influences.

And yet, we can hardly insist that for every random two pieces of causal information there has to be another which tells us how to combine them. If a farmer tells us “drought leads to hunger” we can’t go back to them and say, aha, but how does this rule apply in the presence of information about how poverty leads to hunger? Or, worse, in the presence of *each and every* specific piece of information about how anything leads to anything else?

The whole point of the mini-map approach to coding, aka the graphical approach to causal inference, as Pearl insists, is that mini-maps are the portable, robust atoms of our causal understanding. They are rules which we carry around with us and can apply and combine whenever they become relevant. We can’t appeal to another whole set of rules which govern how to apply the first set of rules. There must be some default understanding of how we combine separate pieces of information about the influences on a variable.

## 17.1 Contradictory combinations

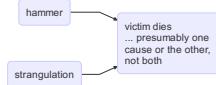
To take the less usual case of contradictory combinations first, if we hear:

the victim was murdered with a hammer

and

the victim was murdered by strangling

we might encode them like this:



... but this might feel weird. We would perhaps prefer to work out which putative cause was actually effective. Or we might suggest XOR as the relevant function.

In the absence of further information, we might look at the two alternatives

$$E = f(C)$$

$$E = g(C)$$

and simply take the average of the two, for all values of C. Again, we could weight these averages.

## 17.2 But which case is which?

In a later section I present a default rule for combining complementary causal maps. And above I started to sketch one for combining contradictory causal maps. The big problem is that we don't necessarily know which one to use. This has to do with whether we think of "G -> E" as meaning

G (and possibly other things) affects E

or

only G affects E

In the first case, the mini-maps *complement* one another, in the second case, they contradict one another. The problem is particularly hard because one causal claim might be considered to be exclusive (to contradict others) in one context but not in another. For example in this case, the second claim would probably be taken to contradict the first, but not necessarily vice-versa:

Heart attacks are caused by chakra imbalance

Heart attacks are caused by a blockage of arteries

... and on the other hand, in this case

Heart attacks are caused by a blockage of arteries

Stress contributes to heart attacks

the person asserting the first would most likely not contradict the second claim but see them as complementary in some sense, and vice-versa.

These are all real problems. If we want to be able to “zoom out of” and simplify our causal maps automatically, we have to do one of the following:

- wherever necessary, code information like “be careful, these are contradictory fragments, don’t combine them with the default combination rule” during the atomic coding of individual causal packages.
  - assume that all causal fragments combine in complementary fashion and simply refuse to code any which are not
  - resign ourselves to the gruelling process of reviewing “by hand” each and every variable by variable every time we want to view a causal map or create a new view (filter, merge, combine, zoom out) of a causal.
- 

In conclusion I will repeat this (somewhat arbitrarily) decision:

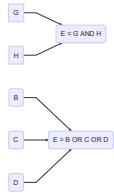
When maps are combined using the “same consequence” rule, we will **assume the constituent maps are contradictory rather than complementary** unless we are told otherwise.

Nadkarni and Shenoy (2004) makes the same assumption. Most approaches to causal maps implicitly assume some default approach, though mostly without specifically mentioning or arguing for it.

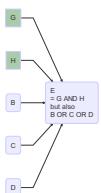
---

### 17.3 Extension to causal *packages*

Above we have only considered mini-maps with single influence variables. In general, mini-maps show how several influence variables together influence the consequence variable as a *causal package*. The logic is the same, but if we combine two causal packages, we have to remember the packages and which variable belongs to which package. We can’t just smash them all together (unless we make very strong linearity assumptions). This means that in this example, there are two causal packages, one which combine to influence E via logical AND, the other via logical OR (but the same applies to any kind of functional package).



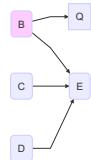
Here the variables have been coloured to visually record which variable belongs to which package (and in general, there is nothing to stop the same variable belonging to *both* packages). Our app has to remember this too (and indeed it does).



## Section 18

# The shared influence rule, functional version

What about when we combine maps according to the shared influence rule?



There is nothing to say because there is no problem; this way of joining maps together does not result in a new functional combination which needs calculating.

One might be tempted to think that “driving” both Q and E would somehow “exhaust” B. But if we needed to take into account some factors which would indeed affect the value of B, we would need a corresponding mini-map in which B was the consequence variable.

## Section 19

### The shared arrow rule, functional form.

When joining two causal maps containing one shared influence variable C and one shared consequence variable E, the first encoding an influence  $f$ , and the other encoding an influence  $g$ , the combined influence can be calculated in the same way as in the [previous section][#shared-influence-function].

---

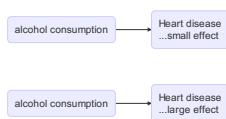
An example of contradictory maps:

If we hear

heart disease increases strongly with alcohol consumption

and

heart disease is hardly affected by alcohol consumption



..In the absence of any other information, we perhaps suggest some kind of average influence.



The situation seems simpler when the information is not contradictory:

If we hear:

heart disease increases strongly with alcohol consumption

and

heart disease increases strongly with alcohol consumption

... we are usually only too happy to combine the mini-maps. The combined function can be calculated as in the previous section.



## Section 20

# The rule for merging arrows. Weight of evidence.

We already saw that the rules for joining causal maps can result in a map in which there is more than one arrow from a variable C to a variable E. And we discussed how to calculate a single function which combines these influences. It can be more convenient to replace this set of arrows with a single one, encoding the same combined influence, and also encoding a new attribute which we can call “evidence”, “number of mentions” or similar.

---

So if we hear:

heart disease increases strongly with alcohol consumption

and again exactly the same thing:

heart disease increases strongly with alcohol consumption



... and have already combined this into a single map:



... we can show this as a single arrow:



We will also consider that this link is now in some sense has more evidence.

Later we will see that it would be usual also to show this visually e.g. by using a thicker arrow.



## 20.1 Problems with evidence

We would probably also want to in some sense increase the “evidence” for a link if we had similar but not identical pieces of information, e.g. that the effect of C on E was on the one hand “large” but on the other hand “very large”. But what if we are combining fragments of information which are not even compatible?

What if we hear:

heart disease **increases** with alcohol consumption

but then:

heart disease **decreases** with alcohol consumption

We could still record stronger “evidence” for the combined link, and by all means it is true that there is a lot of evidence about some kind of link, but that would be very problematic. What if the combined function was calculated to be in some sense zero?

Case A: there are 50 studies which suggest a strong positive link and 50 studies which suggest a strong negative link. We could combine these into a single map (see section xx) claiming a *zero* link but evidence of 100.

Case B: there are 100 studies which *all* suggest a *zero* link. We could combine these into a single map with a combined function claiming a zero link but evidence of 100. It’s the same map as Case A! That must be wrong.

## Section 21

### The chaining rule with loops

A special case of the mini-map coding rule is when the consequence variable is also one of the influencing variables.

A special case of the join rule is when the resulting mini-map contains a loop or cycle...

[add something here!]

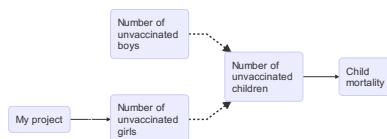
## Section 22

# The rule for conceptual links

Conceptual links like “B is part of the definition of E” can be coded analogously to causal links but should be clearly distinguished from them.

The number of unvaccinated children is a conceptual combination of the number of unvaccinated girls and the number of vaccinated boys, but it isn’t caused by them.

The Soft Arithmetic of causal maps which include conceptual connections too is really neat. So for example if my project increases the number of girls who are vaccinated, and we also know that total child mortality is partially caused by the number of unvaccinated children (showing conceptual connections as dotted arrows):

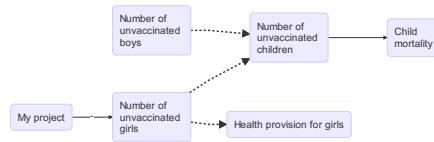


..... then the causal influence travels along the lines in the direction of the arrows, oblivious to the fact that some of them are dotted, and we can infer that my project likely helped to suppress child mortality. Nevertheless the dotted lines are not causal connections.

Where necessary, you should specify the function involved – in this case, the arithmetical sum.

## 22.1 Variations

In the simplest case, conceptual definitions are tree-like. But in real life we get real headaches with concepts which mutually define one another. Or with concepts which are part of various other concepts, like in the next figure. If we are going to track the influence of “my project”, we don’t want to unfairly exaggerate it because is part of two broader concepts? Or do we?



## 22.2 Problems

Understanding conceptual links is crucial for coding. e.g.

- Source P says the project “improved older people’s health”
- Source Q says the project “improved people’s health”

## Section 23

# Combining “extra” information like the values or levels of variables

What if the mini-maps we want to combine agree in their causal information but differ in the extra information like the values of variables:

the heavy rains last week led to 10m floods in Newville

and

the heavy rains last week led to 5m floods in Newville

we would not be able to encode them in the same way but might do it like this:



[add more!]

## Section 24

# Context

All maps actually have a context attached to them in which they are claimed, stated, posited, uttered, etc.

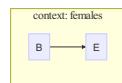
I am not going to deal much with context at this point, i.e. I will assume that all the maps come from the same unspecified context, but context is really important.

### 24.1 Intersections

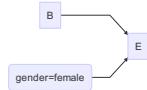
In general, the combined map is only true in the *intersection* of the contexts. (So if one map is true for females, and the other is true only for young people, the combined map is true only for girls).

### 24.2 Replacing contexts with variables

If one or more variables B, C,.... influence a variable E in a context K, we can recast this so that the information about the context is replaced by an additional variable K which governs the influence of B, C etc on E?



That would be great, because this would just be another mini-map:



... and we would have reduced the idea of “context” to a variable within an expanded mini-map. The difference is that we (typically) have no information about the relationship between B, C when gender is *not* female.

---

#### Technical note

There is no such thing as absolutely specifying a context. The only way to do it meaningfully is simply to mention any relevant differences from the current context. See Pearl xx and Mayne xx.

## Section 25

# Maybe the nodes are not variables, maybe they are ... propositions, events, schemas?

Mostly in this guide I assume that the component parts of mini-maps are *variables*: things which could be different, take different values. On balance, variables are the best choice for this job.

But there are other solutions which come close and which are worth a quick mention.

### 25.1 Events

When we are coding retrospective narratives (“this happened, then that happened”): couldn’t we treat these as events rather than variables? And what would the difference be? (The QuIP approach of eliciting causal information in the context of questions about “changes” invites the respondent to provide this kind of narrative-style information.)

In event-based maps we would have nothing like the same kind of predictability which we claim for variable-based maps. At best we could claim that a causal chain was plausible in retrospect.

The thing about event-based historical scenarios is that the cause does not just create a causal pulse on the next variable but also *co-creates that next variable*.

[add more here!]

### 25.2 Propositions

Our nodes could be propositions in the sense of propositional logic. One big difference to a variable-based approach concerns negative statements. If we have a Boolean, false/true *variable* like “Project is launched, false/true”, this suggests that we are supposed to be equally good at recognising each value of the two values of the variable. Whereas *propositions* are arguably asymmetric in terms of evidence: the evidence for NOT-C is just the absence of the evidence

for C, which can be problematic if the search is unbounded, because we don't know if really NOT-C or just that the evidence hasn't arrived yet.

Some event-based narratives, especially non-scientific ones, can be better understood as links between propositions than about links between variables.

### 25.3 Schemas / Schemata

Most human decision-making is “fast” Kahneman (2011) and associative – the kind of thinking which is better understood in terms of *cognitive schemas*. Causal maps with variables or propositions for their nodes are better at dealing with “slow”, deliberate reasoning. We could in principle use something like causal maps to model the links between *schemata* - for example, an arrow could show that schema E is more likely to be activated when schema B is activated. The trouble is, it is really difficult to accurately elicit this network from respondents.

What is great about schema theory is that it provides a model of how schemas can evolve. This kind of training has also been proposed for FCMs Koulouriotis, Diakoulakis, and Emiris (2001).

## Section 26

# Variable-based and propositional-based approaches to causal diagrams: chalk and cheese??

Causation: the relation between mosquitoes and mosquito bites  
(Michael Scriven)

I want to present two different approaches for coding and combining causal links: the “proposition-based” aka “mosquito-bite” approach, versus the “variable-based” approach. The second is more general, more fiddly and less intuitive. The first is simple, plausible, and limited. I think some of the difficulties we’ve had in discussing how to do QuIP coding come back to this distinction.

Classic QuIP coding – as I think it was, say, a year ago (see the Appendix to Copestake, Morsink, and Remnant (2019)) e.g. before any innovations like the coding of multiple influence items or positive versus negative links – is very close to the first, though it does have a couple of elements of the second; and recently we/you have been trying to extend it a bit further and more consistently.

**For any given QuIP project, I think you have to choose. Stick to simple and intuitive “proposition-based” coding, which is what you are doing now, and do without the snazzy extras like multiple causes, necessary conditions, adding up influences, etc. Or switch to “variable-based” coding, which is more powerful but perhaps often not what you want. My app can do either, and you can adapt Excel to do either. But I don’t think you can mix the approaches within one project.**

It is hard to just add a couple of features to a proposition-based approach. It’s one or the other. It’s difficult because of the logic, not because of the tech.

But James has an interesting suggestion: code individual responses in a “proposition-based” way, but with the option to generate variable-based theories when aggregating these propositions, for example perhaps deducing

where there are necessary conditions. But at the bottom level, when coding individual responses, you would still be unable to code multiple causation, necessary conditions, etc.

Logical Frameworks (Logframes) lean heavily on the proposition-based approach too.

In the *proposition-based* approach the items are propositions like “I lost my job” or “the weather is getting worse”. The propositions can stand alone but most often we want them linked together, like “I lost my job → I sold my cow”.

One natural way to get statements for this approach is to ask “what’s changed here” and then ask “what caused that?”. The answers are specific statements about what led to what – what actually happened – in a particular context.

The propositional approach has many advantages: above all, it is a part of a very natural way to ask about and record causal influences and which gathers evidence with high plausibility. I also suspect that its simple “causal chain” model is easier to translate for different settings and cultural contexts. However it is nothing like as general as a variable-based approach and has trouble dealing with common cases like multiple influence variables.

Terminology: I will follow the terminology in the Glossary in BSDR (2017), together with our recent additional Glossary (also, I will call the nodes of the causal maps “items”, which is a more neutral and appropriate term than “variables”).

## 26.1 The variable-based approach

This approach is what I’ve already been suggesting for QuIP coding; I’ve described it in some detail at [http://www.pogol.net/\\_causal\\_mapping/index.html](http://www.pogol.net/_causal_mapping/index.html). It’s just a more qualitative extension of mainstream causal modelling in social science such as we see in structural equation modelling (SEM) or as described by Judea Pearl; and it’s close to the sort of causal networks we see in Fuzzy Cognitive Mapping.

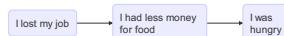
In the *variable-based* approach, the items can be e.g. *level of* unemployment, *level of* retention of cattle. These variables can be binary (yes/no) or continuous but the most important difference is that they are not true or false; they don’t make any claims. Specification of the actual / current / predicted value of the variables is an additional, optional step. The variable-based approach is more general; the proposition-based approach is a special case of it. A *proposition* like “I sold the cow” can be expressed using the variable-based approach (more long-windedly) as “The variable *did I sell the cow, yes/no* takes the value *yes*”; or even as “There is a variable called *did I sell the cow, yes/no*; and this variable takes the value *yes*”. But most often, we don’t need to bother with this kind of thing because we are encoding statements like “loss of job → sale of cow” and we don’t need the information about whether a particular respondent actually did the selling or not.

One natural way to get statements for this approach is to ask “what causes what around here?” or “What has to happen for a farmer like you to get a good harvest?”. The answers are general statements about what leads to what

in a particular context. That context can be quite narrow – just these farmers in this valley – but the statements are general: they don’t necessarily make any claims about what actually happened. In this approach, we encourage people to tell us their knowledge about the whole domain regardless of what happened to them.

## 26.2 More on the proposition-based approach

Below is an imaginary causal chain derived from some corresponding statement(s) from a single source. It is a good example of a “mosquito-bite” or “propositional” chain. Each item can be said to *cause* the next one, down the chain, like dominoes falling over. This idea of “X causes Y” is strong in this approach, a familiar idea in ordinary language. You can’t say “X causes Y” so easily with variables. It is transitive: causality flows down the chain, and if  $X \rightarrow Y$  and  $Y \rightarrow Z$ , then  $X \rightarrow Z$  too. It captures the idea that we can sometimes even see causation happening – the ball hits the window, which smashes. The mosquito bites the person, who feels the jab. The two parts of the causal link are even hard to separate: it is the ball which “smashes” but the glass which “is smashed”; two sides of the same coin. We feel strongly that we don’t need any more waffle, that this is just causation happening in front of our eyes. It a strong kind of evidence, it is “not circumstantial”.



In the mosquito-bite approach, every item contributes 100% to the item downstream of it (and to those preceding that item), and every item can be attributed 100% to the item upstream of it (and to those preceding that item). Is this a simplification? Of course, but *every* model/approach to causation is a simplification.

Propositional-based causal thinking is so sweet and neat that we’d all love it to be good enough at least for everyday thinking about how the world works and in particular for looking at the role played by a project.

My main thesis here is that it is difficult to tweak this “mosquito-bite / propositional” approach to allow for desirable features like strength of connection or multiple causation without ending up with a variable-based approach.

### 26.2.1 The items are propositions, not variables

The items (the boxes or nodes between which arrows are drawn) are **propositions** and not, for example, variables. The items themselves are specific empirical (but not causal) claims made in and about a specific context (such as “X happened” or “Y has got worse”) and are, broadly speaking, more or less true or false in that context.

The claims have a varying amount of subjectivity, from a fairly subjective “this has got worse for me” to a fairly intersubjective “the rains have got worse”; nevertheless they are grammatically all *claims*, which can in principle be more or less true or false.

This also means that the items have *no gradation*. They are always one thing or the other. There are no possibilities in between.

### 26.2.2 The items have been individually experienced

In general, the propositions are / have been **personally true for, or at least personally experienced by the source** – BSDR (2017) doesn’t say this explicitly, and focus group data asks for information on “people like me”. So, seeing that the absence of X leads to the absence of Y does not give me the right to claim that X leads to Y (even if I have seen X leading to Y in relevantly similar cases, but not my own case.)

### 26.2.3 The propositions have contrastive implications

I think that this way of thinking does at least concur with the a mainstream view of causation that a causal claim always *implies* a “counterfactual” contrast. If someone makes a claim of the form that “X caused Y”, but then says that actually even if X had not happened in that way, this wouldn’t have made any difference to Y, and that Y can happen or not happen irrespective of X”, we would conclude that the speaker does not share our understanding of causality, proposition-based or not, and we should **not** code the initial statement as a causal claim.

For example, if someone says that the reason my crops were good is because it is the Year of the Dolphin, and the reason my neighbour’s crops were poor is also because it is the Year of the Dolphin, if they have not named further explanatory factors we would probably not code this/these as causal claims at all.

This is quite another thing from saying that the people making it have to be in possession of corresponding evidence, and is compatible with a proposition-based as well as a variable-based approach. Mosquito-bite claims are based on a more fundamental experience of causation than comparing a state of affairs with a counterfactual. But they do have counterfactual implications.

### 26.2.4 Two-sidedness of propositions

With variables, all the values of the variable have to be clearly defined. So if we have a variable like “% unemployment” then we know what 100% means, we know what 20% means, etc. If we have a variable “Lost job, yes/no” then we

know what the “yes” situation looks like and we know what the “no” situation looks like.

A strength and weakness of propositions in ordinary language is that we might know what their “yes” side looks like but we don’t necessarily know what the negation looks like. So we don’t always know the negation of “I sold my cow”. Is it “I didn’t sell my cow, I kept it” or “I sold my pig” or “I lent my cow to my neighbour”. This kind of ambiguity makes these one-sided propositions useless for explanation. I think that QuIP coding does understand propositions in a two-sided way. That is what makes them amenable to contrastive explanations, see above.

#### 26.2.5 Aggregation

The “proposition-based approach” has, by design, less taste for aggregation and simplification, preferring to say “dig into the data yourself”. That’s fine. But classic QuIP coding does do one important kind of aggregation out-of-the-box, namely when coding two different items from two different sources as “the same”, using in QuIP-speak “the same tag”. This issue of deciding whether, say

- “I sold my cow”
- “I sold some cows”
- “I sold some livestock”
- “I got rid of a cow”

should be coded under the same “tag” and if so which one, is of course a challenge of all social research. The propositional approach has a somewhat harder time of it because there is no way to code the idea of “same thing, different degree”.

#### 26.2.6 The items can express changes, states or events

A strength of this kind of proposition-based analysis of causal statements is that it quite naturally captures the kinds of causal explanations which people actually give. One could argue that we shouldn’t even be worrying about what format the antecedent causes take (events, states or whatever), because they are just whatever explanations people give. But this would be problematic. The coder needs to understand the items within a claim, rather than just reproducing them verbatim – first in order to be able to judge whether a different claim by a different source involves *the same* item or not, and second when reflecting on the other statements made by a respondent changes our understanding of the original statement, and third, most importantly, to check that they really mean *causation* here.

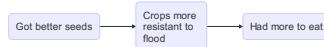
QuIP seems here to be *more* restrictive than the proposition-based approach, sometimes suggesting that only *changes* are relevant, though the definitions in the Glossary in BSDR (2017) are not completely consistent. So, a Driver is “An action or state (X or Z) behind outcomes (Y).” and Outcomes are “Changes (positive or negative) reported by respondents”; as “outcomes can also become

drivers of change”, this means that the apparent distinction between “actions or state” on the one hand and “changes” on the other hand is blurred (because any one item can be both).

Perhaps classic QuIP coding requires only that final (“no-child”) outcomes which have to be expressed as *changes* in the reference period, whereas other items can be expressed in a variety of ways (though still as propositions). But see below.

#### 26.2.6.1 Changes

Changes are familiar items in QuIP coding, for example .....



Each item can be seen as a comparison between states at at the beginning and end of a reference period.

#### 26.2.6.2 Events

Events are perhaps the archetypical kind of thing to appear in a proposition-based causal chain.



Though there is no hard-and-fast distinction, it can make sense to distinguish events from changes. In this case, the event can also be parsed as a change, from lower to higher food taxes. But ...

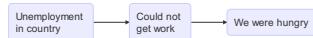


... in this case, the triggering event can hardly be described as a *change*. It is a clear example of a one-off event.

“Events” pass the contrastive test (see above): if the incursion hadn’t happened, we wouldn’t have lost (so many) cattle, but it did, and we did.

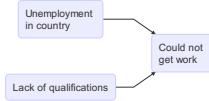
#### 26.2.6.3 States

Putative static causes like “unemployment” are things whose status may not be changing, but still pass the contrastive test. (Of course, causal claims about *changes* in an influencing factor like “*increasing* unemployment” are also possible.)



“Unemployment” perhaps remained at a high level throughout the reference period. Perhaps the source even explicitly mentioned “if we’d had jobs, we could just have worked overtime to compensate”, contrasting unemployment only with its absence, namely employment. Here, unemployment is not mentioned as something which changes over time; rather, it is contrasted with another situation which might have existed outside the reference period or simply as a desirable but non-existent situation.

The role of “unemployment” here should be distinguished from its role as an “enabling factor” in this example (which in any case can probably not be coded in the propositional approach):



#### 26.2.6.4 “Differences” as a general way of expressing events / states / changes

The distinctions between events, states and changes are not very clear and maybe not important. A response would be to replace these ideas with the broader idea of “Differences” or “deltas” (<https://www.linkedin.com/pulse/dear-world-bank-its-differences-changes-we-look-impact-steve-powell/>).

All the examples above can be covered by the idea of “Differences”; things being one way rather than some expected or possible contrast: a state of unemployment rather than a state of employment, a status quo in flood levels rather than the feared contrast of floods, an incursion of terrorists rather than none, and so on.

As we already mentioned, QuIP understands propositions in a two-sided way, i.e. their negations are clearly defined. This makes it possible to express Differences using propositions too.

#### 26.2.7 Golden threads

One big advantage of QuIP-style coding is that it has the concept of an entire causal chain which is not reducible to a bunch of individual links. Indeed, attribution codes are given to whole chains, *not* to individual links.

It’s of practical benefit that BSDR has now adapted the coding process to be built up from individual atomic maps, rather than chains. But this shouldn’t mean that chains have disappeared from the coding process.

It is essential for QuIP to be able to track from outcomes of interest back to project interventions within individual causal claims from individual respondents. I understand that this is gold for QuIP; we can call them “gold threads”. It is much more valuable than a case in which one respondent mentions a link from the project to some intermediate item X, and another mentions a link from X to an outcome of interest. This cannot be done if we just construct a causal map for all respondents from a pile of atomic links. The information about which chain each link belongs to has to be additionally retained, in order to be able to dig for QuIP gold.

Golden threads are valuable because of their direct plausibility; actual individuals spontaneously mentioning a causal chain from an intervention to an outcome of interest. Donors love this stuff.

On the other hand, it is still possible to use QuIP (and proposition-based approaches) for questions which are not focused on a particular project.

### 26.3 Statistics?

Isn't a variable-based approach just SEM? Well the extension of SEM-type thinking I'm talking about here has no parametric assumptions at all, so it is more general, weaker, and potentially more computationally expensive; it has to be able to deal with ideas like "a bit more than". And most importantly it isn't statistics, in the sense that it isn't conceiving of respondents' causal claims as some sample out of a larger set which approximate the truth. It says, even if we believe these statements are actually the truth – they are already the conclusions which statisticians try to reach from their empirical data – and after all they are based on people's actual causal knowledge, not just on a few observations – how can we code and combine them? In this sense it is no different from the propositional approach. But yes it is different in the sense that we are trying to encode and aggregate theories, models, with explicit counterfactual implications, rather than direct, one-off reports of "this happened and it made this happen"

### 26.4 Conclusion

The "proposition-based" aka "mosquito-bite" approach is simple, plausible, and limited. The "variable-based" approach is more general, more fiddly and less intuitive. I think some of the difficulties we've had in discussing how to do QuIP coding come back to this distinction. The two are quite different and can't be mixed in one set of codings. Perhaps, as James suggests, coding can be done with the propositional approach and then a variable-based theory built up from that. But this does not get round some of the important limitations of the propositional approach in the very first step – basic coding. I will look at these at in my next contribution.

## Section 27

# Limitations to the proposition-based approach

### 27.1 Problems *within* the proposition-based approach

#### 27.1.1 Problem with only coding personal observations rather than personal causal knowledge

Suppose a new road is built (this is not the intervention), and nearby farmers have worse yields. I think James says we can process the interview testimony of farmers living near the road (whose yields have decreased) as `new road in neighbourhood —> yields decreased` but not the testimony of more distant farmers (whose yields did not decrease), even though they are in touch with their peers and have personally experienced the “other half” of the counterfactual, and can also attest to the above statement. (Yet I think James says it is OK to code this kind of information when the farmers are reporting together in a focus group.)

From a variable-based perspective, this might seem bizarre. But from the perspective of, say, a journalist, or an evaluator conducting an ordinary summative evaluation of a project, it is a completely normal way of thinking: “we don’t want people’s theories, we just want the facts”.

This is a massive strength and weakness of the propositional approach. The thing is, I’d say that what counts as a simple fact about causation actually rests anyway on people’s underlying theories. We certainly don’t want farmers to tell us merely “I noticed the training. And I noticed my crops were better. That’s all I can say”. That reduces QuIP to a kind of soft self-report questionnaire with a dangerously low N. No, we want them to say “I experienced the training, I learned a lot, I could do things in new ways which make sense to me given what I know about farming, and I wasn’t surprised to see my crops were better”.

Evidence based on testimony from respondents who have “actually experienced” all the parts of a causal link (yes, I got the new seeds, and yes, I got better harvests, and yes I think the one is the cause of the other) usually carries a lot more weight for the reader than hypothetical statements. On the other hand, I think this oversells the importance of the experience of co-occurring events (which opens QuIP to criticisms about small sample size), as

if our respondents were just uninformed scientific observers of random events, and underplays the importance of respondents' more fundamental, underlying knowledge about this kind of causal link which is not only derived from any particular chain of events.

### **27.1.2 Problems with establishing common codes using difference of degree**

Classic QuIP does already take one very important step onwards from the most basic and literal coding of individual causal claims, namely by establishing common codes for similar items across sources. This is already a kind of aggregation. For example, if one source mentions “some health problems” and another mentions “severe health problems”.

In a variable-based approach this could be easily dealt with for example by saying that the effect of two different items from the two different sources have a different strengths; and the variable “health problems” can have different values or levels. But if “health problems” is just a proposition, the difference in degree gets lost; conservatively, “health problems” can only be interpreted as the weakest proposition which could fall under that title.

Fuzzy Sets Ragin and Pennings (2005) provide an interesting approach: the items (like “unemployment” or “democracy”) can carry on as monolithic ideals, but respondent *claims* about them can be moderated according to degree of accordance with that ideal, aka degree of membership of that fuzzy set. However I think this ends up being a kind of variable-based system.

### **27.1.3 Need for a distinction between “plus” and “minus” links?**

The rationale for including successive antecedents in a causal chain is that each causally explains the item it influences. In the variable-based paradigm, there is no initial room or need for the idea of “directionality” of links. Statements like “I slept badly → I failed the exam” and (a different respondent) “I studied hard → I passed the exam” fit well into the propositional approach. (Expressing a negative or “minus” link like “Sleeping badly → MINUS → Passing exam” fits better into a variable-based approach.)

So far, so good. The problem arises when we (understandably) want to be able to code “failing the exam” from one causal chain as the same item as “passing the exam” from another chain; if we do recode all “failing” items as “passing”, all the old links will have to have reversed directionality.

It’s also worth noting that if a revised QuIP is going to allow minus as well as plus links, the simple ways to summarise project impact on p. 4 and 5 of BSDR (2017) have to be adjusted.

On the other hand, we have to be careful. The proposition-based approach is perhaps more sensitive to ordinary language. Using variables, it is tempting to express “Getting more vitamins → Feeling more healthy” equally as “Getting fewer vitamins → Feeling less healthy” or “Getting more vitamins → Feeling less ill” or “Getting more vitamins -MINUS-> Feeling more ill”. Whereas there are real distinctions; illness is not merely the opposite of health.

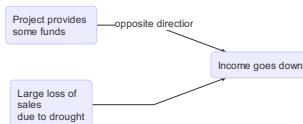
## 27.2 Limitations: Problems with *extending* the proposition-based approach

### 27.2.1 No way to code multiple causation

In the proposition-based approach, there is only one influencing item for each consequence item in a causal chain.

(However, the same source may mention “the same” item as part of different chains – when mentioning several chains leading to the same final outcome, or when chains share an intermediate or initial item.)

Allowing causal packages with more than one influencing item is not as simple as it sounds, not least because of the provision “items have to have been individually experienced”.



In this case, from a variable-based perspective we can see that there is a simple, continuous outcome. The project has a small, plus effect on it, which is outweighed by another larger, minus influence. The overall change (the only thing which the classic QuIP protocol can record) is a moderate reduction in income.

There is no way even of saying “the drought only led to a small drop in income; if the project had not happened, the drought would have led to a larger drop in income” because in the proposition approach there is no way to express this gradation of income.

The logic of “attribution” coding is tricky here. On the surface, this is a negative change, a negative outcome, which is also (partially) attributable to the project! This would make the project seem *bad*, whereas in fact the specific contribution of the project is (from a variable-based perspective) positive.

Also the requirement that “the items and causal mechanisms have to have been actually witnessed by the respondent” is harder to understand/implement here. It’s not quite clear what this means in the case of two opposing, if independent, forces. If I watch two ropes pulling on an object with equal force in opposite directions, do I really witness the causal effects of the two opposing forces in the same way that I might witness just one rope pulling unopposed on the same object?

I can’t just directly witness how two forces interact to make something else happen. I can only use my existing expertise and/or tweak the mechanism to

find out how it works and/or observe over time and/or observe my peers. And that means elaborating a theory, going beyond direct experience.

It is very difficult to extend the propositional approach to even these very simple, additive / subtractive cases.

### 27.2.2 Problems with expressing difference of degree

The cousin of the problem with coding similar outcomes with different degrees (see above) is that the propositional approach has no way to code statements where the effect of B on E is explicitly relativised, e.g. “my lack of skills maybe contributed to me losing my job, but it wasn’t that important”. This already implicitly introduces the idea of additional causes, see above.

### 27.2.3 Explicit causal knowledge

I think it is also important that what is to be coded is an explicitly *causal* claim. It is not enough that the respondent says “I noted that X changed. I also noticed that Y changed.” The QuIP protocol is designed to draw on respondents’ underlying causal world-view, and not primarily on observation of co-occurrence. The variable-based approach says this explicitly: we want to get at people’s theories. I don’t know how the propositional approach can distinguish between “I noted that X changed. I also noticed that Y changed.” and “I noted that X changed. I also noticed that Y changed; X caused Y.”

### 27.2.4 Vulnerability to luck / coincidence

Like most practical evaluation techniques, QuIP is vulnerable to over-interpretation of happenstance. For example, if a perfectly-conceived and prepared project is denied success by an incredibly unlucky meteorite strike at the last minute, QuIP is unable even to begin an evaluation, because there are, unluckily, no “positive changes” to put at the end of causal chains. QuIP is not alone in this: most narrative approaches to evaluation would have the same problem. Whereas a broader variable-based perspective<sup>1</sup> might ask to what extent the project was the right kind of tool to address the problem, regardless of the fact that it was denied success in an unforeseeable way at the last moment.

This is the difference between evaluating overall *success potential* (variable-based; most interesting for scientists and policy-makers) and *actual success / failure* (proposition-based; most interesting for journalists, law courts, and Ambassadors).

There is a big tension between plausibility and validity.

The Ambassador might well have trouble understanding why an evaluation of success potential might give a project full marks even though it was ultimately unsuccessful.

(The same problem works the other way round, when a poor project “achieves” positive changes via lucky pathways.)

---

<sup>1</sup>One could argue how “theory-based evaluation” fits in here

### 27.2.5 Unable to code necessary/sufficient conditions

Many other causal connections which we might want to code are essentially someone's more or less formal theory, going beyond the direct one-off connections which the propositional approach is so good at capturing. For example, you can't directly observe that X is necessary for Y. That claim rests on pre-existing causal knowledge, or is the result of more or less formal experimentation. So classic QuIP can code the observation "water made my crops grow" but not "water is necessary for my crops to grow" because that goes beyond a mere observation.

## 27.3 Weaknesses of the variable-based approach

*Events* are easy to capture with a proposition-based approach and harder to capture with a variables-based approach. So the assassination of Arch-Duke Ferdinand was a causal trigger for WWI, but it is hard to conceive of that assassination as a variable, (something that would still be around but not actuated in an alternative history in which he never went to Sarajevo).

## 27.4 What is to be done? Suggestions for QuIP

- 1) Just stick to propositional coding. It is simple and can code the majority of statements. We can aggregate items and links as now, e.g. "X was mentioned 20 times" but go no further in our interpretation of what that means. Accept that we won't be able to directly code multiple causation, necessary conditions, links with different strengths, statements about lack of change, etc. This approach has nevertheless worked well so far. It is natural, relatively easy to apply and explain, produces highly plausible evidence, and has a very natural and persuasive way of presenting the extent to which a project seems to have led to key outcomes, which is what commissioners want most. In particular it produces "golden threads": when one respondent explicitly mentions a causal chain from the project to an outcome.
- 2) As James suggests, coding can be done with the propositional approach and then a variable-based theory built up from that. But this does not get round the limitations of the propositional approach when doing basic coding.
- 3) Use a variable-based model, as already discussed, but to make more effort also to preserve the simplicity and plausibility of the classic model. For example, the existing method of attribution coding can put special emphasis on "gold threads" when and if they appear; plus, "gold threads" can to a large extent be identified automatically.

## Section 28

# Clarifying how to ask QuIP questions in contexts when the final outcomes are not necessarily “changes”

### 28.1 Classic QuIP: the focus on changes

The QuIP protocol usually begins by asking respondents questions about changes in pre-defined domains of interest like nutrition within a reference period. Sources mention items important for them in that domain and work backwards from each to provide chains of antecedent causes. QuIP is mostly very interested in the extent to which respondents spontaneously mention an intervention of interest amongst the causal antecedents — though it can be used for other purposes too.

Here, I'll look at how QuIP doesn't need to rely too heavily on the idea that the items we code have to be *changes*.

### 28.2 Even final outcomes don't have to be changes

The way the very questions are formulated in QuIP means that, in QuIP, final outcomes (those at the end of an individual respondent's causal chains) have to be expressed as changes (because that is what we ask people about). But that doesn't have to be the case. In some contexts it might be important to ask about other “Differences made” by a project.

**SECTION 28. CLARIFYING HOW TO ASK QUIP QUESTIONS IN CONTEXTS WHEN THE FINAL OUTCOMES ARE NOT NECESSARILY “CHANGES”**

92

**28.2.1 One-off events**



The commissioner of this evaluation might have been involved in, say, providing the preparation of employers and is very interested in improving the retention of first-time workers within the reference period. But the individual students interviewed as part of this evaluation are not, from their own individual perspectives, describing changes of any kind; they are describing an experience which is by definition a one-off. They are describing Differences made, contrasts with *hypothetical alternatives*, not with *prior states* experienced by the respondents themselves. “I really knew what to expect and the people knew what support to give, much better than what my cousin said it is like where she lives.”

Weddings, births, transition to first employment etc, and their specific characteristics are all stand-out examples of events which are obviously important in individual life cycles and are often mentioned in QuIP interviews but which, at least from the perspective of individuals, can only with difficulty be characterised as changes within a reference period.

It’s an ethical fallacy (common in evaluation circles) that desirable changes have to have a significant trace arbitrarily far into the future. Good things can simply peter out. It is important that for example refugees are treated with dignity during a project, whether or not that somehow has consequences at the endline of a project. So the “final outcomes” of interest don’t have to be in the present. We can ask “how was your experience in the maternity ward, how did it compare with what you had expected, with what other mothers had told you?”

**28.2.2 Maintaining the status quo**

QuIP can also contribute to the evaluation of projects which aim to maintain the status quo, where the aim is for things not to change.

**SECTION 28. CLARIFYING HOW TO ASK QUIP QUESTIONS IN CONTEXTS WHEN THE FINAL OUTCOMES ARE NOT NECESSARILY “CHANGES”**

02



### 28.2.3 Status quo as a divergent counterexample

As I understand it, even when a positive change is expected and reported by many respondents, if we insist on only asking for changes (as QuIP suggests), we suppresses the experience of others who did *not* experience a change. As they can only provide causal chains ending in a change, their experience is missed out. To be sure, as QuIP generally only includes respondents who have been exposed to the project, if a subset of respondents do not report a positive change, at least we can count them. But the respondents might have other information to give about other important non-project factors when the status quo was maintained which simply goes missing if we only ask about changes. What if they would have been able to report that they didn't get good crops because of the presence of a pest, but that information never gets written down?

Embracing the idea of “Differences” rather than “changes” would allow QuIP to evaluate projects with the aim of maintaining the status quo, although obviously the questions would have to be formulated a little differently.

### 28.3 Preceding items don't have to be changes either

As I understand it, there is less insistence in QuIP that the antecedent items which respondents report as *causes* of the final outcomes themselves have to be changes: they might for example be “states” or “events”.

### 28.4 How to focus on the relevant contrast

The trouble is that it is not always obvious *what* we are supposed to be contrasting with the present state of things – isn't this subjective?

Even insisting on “changes” – in the sense of differences from a baseline – does not remove this subjectivity, because what counts as the baseline is also up for grabs. If we ask a teacher how the new teaching method is affecting student performance, they are unlikely to think of a baseline derived from student scores at beginning of the year, as children will improve anyway. They are more likely to contrast performance with similar cohorts at the same time of year.

**SECTION 28. CLARIFYING HOW TO ASK QUIP QUESTIONS IN CONTEXTS WHEN THE FINAL OUTCOMES ARE NOT NECESSARILY “CHANGES”**

94

So in a specific research context it may be necessary to ask in an open way which suggests a variety of different axes of contrast: “how was your experience in the maternity ward, how did it compare with what you had expected, with what other mothers had told you, with your previous birth (if you had one)?”

In the variable-based approach, the global question is of the form “what causes what around here, in domain X?” which is less susceptible to this kind of problem; but even so, the questions actually asked in the field need to be tailored not only (of course) to the context but in particular to the intervention and the kind of Differences it is likely to make.

## 28.5 Drivers

### 28.6 Focus on initial items (“drivers”)

I think that classic QuIP gives a special role to the notion of drivers as the beginning of a causal chain; the project of interest is, of course, by definition a driver, but other “beginnings” get special status too? I’m thinking of Fig 5 in BSDR (2017), which shows a table of outcomes attributed to each driver.

The word “driver” seems to suggest that the beginning of a chain is the wicked initiator, and the intervening items are just innocent bystanders transmitting an impulse. But most real chains aren’t like that are they?



In this example, unemployment might be a relevant and actually present causal factor even without government ineptitude. I’d have thought you’d want to see “unemployment” in Fig 5 even though it might never appear at the beginning of a chain? In other words, does it matter to QuIP that an item (apart from the project itself) is at the beginning of a causal chain, does that give it a special status?

So perhaps we can say:

- The basic unit of “classic” QuIP coding is the causal chain.
- The final item in the chain is a change experienced by the respondent in a given domain during a given reference period.
- Going backwards from the final outcome, each link in the chain codes how the influencing item causally explains (or partly explains?) the consequence item.

*SECTION 28. CLARIFYING HOW TO ASK QUIP QUESTIONS IN CONTEXTS WHEN THE FINAL OUTCOMES ARE NOT NECESSARILY “CHANGES”*

95

- The logical nature of the preceding items is not specifically defined; anything is allowed which could count as a (partial) causal explanation of the item it influences. This can include events, changes during the reference period, and contrasts with hypothetical alternatives.
- Each item and each causal link has to be something which the respondent themselves has directly experienced/witnessed.
- Each chain is tagged with an assessment of the extent to which the final item (the change) can be attributed to the evaluated project.

## Section 29

# Coding individual causal fragments using propositions (revised)

Propositional maps differ from variable-based maps in many ways. Propositional maps encode answers to the question: what happened here and what caused it? Variable-based maps encode answers to the question: what causes what around here?

Here, I will discuss the most basic coding rule for bare-bones propositional causal maps, and discuss minimal **extensions** to it. My aim is to be able to code statements like this:

The small government subsidy helped, but it wasn't enough. Due to big price rises, my available income still went down a bit.

– using the propositional approach: propositions (and sets of propositions) causing other propositions. You are still dealing with historically true stories, not chatting about abstract models or what-might-happen-if. You still have something closer to an evaluation than to research. I'll try to show here how this is still possible as long as we assume that different contributions are combined in a linear, additive way (“linear superposition”) as often assumed in natural sciences.

[But: I argue that even the most basic causal coding anyway presupposes an understanding of an underlying causal claim behind the propositional claims, and that those causal claims always go beyond any specific piece of evidence. So it is not the case that “ooh, these extensions are risky, they go beyond the evidence to talk about abstract stuff”. Even the simplest propositional causal claim relies on background models of how things work.]

### 29.1 Basic coding rule

The basic coding unit is of the form “B causes E”.

B and E are both propositions. Both are true.

A proposition like B or E only has two possibilities; it is the case (which indeed it is), or it is not the case.

A proposition expresses a *Difference*. A good example of a Difference is a change, but other things can be Differences too (states, events, etc). It says that things are one way rather than some other, perhaps expected, contrast.

### Examples

- I bought a cow (as opposed to not buying one)
- Yields have got worse (as opposed to not getting worse)
- I am fairly happy (as opposed to not being fairly happy)
- There is unemployment in the country at the moment (as opposed to there not being unemployment)

We have no way to code a proposition on its own; only as part of a causal statement.



We can be vague about the timing of things. The consequence has to come after the influence, or be concurrent in such a way that effects are still happening after causes.



#### 29.1.1 Every causal propositional claim is actually three claims

The basic coding rule encodes three claims:

1. the influence item is true
2. the consequence item is true
3. *the first caused the second*

This last claim is really important. We do not encode simple observations that things merely happen together. We are in fact encoding, in a minimal way, the causal knowledge of the respondent, a minimal part of her implicit causal map of the world. There is no way that merely observing B and E on their own warrants the causal conclusion.

The way that people tell stories usually implies beliefs that there are causal connections between the links of the story. We can usually assume that causal claims are being made. They don't have to be made explicitly. "I got the cow. Now we have milk."

#### **29.1.2 The items have been witnessed**

In general, the propositions are / have been **personally true for, or at least personally experienced by the source** or sources. We treat the source as a **witness**.

#### **29.1.3 The propositions have contrastive implications**

If someone makes a claim of the form that "X caused Y", then this implies that if X had not happened in that way, Y wouldn't have happened, and that if Y hadn't happened, X can't have happened. This is a strong claim: within this narrow context, holding everything else constant, X is necessary and sufficient for Y. Whenever X is true, Y is true, and vice versa.

This is quite another thing from saying that the people making it have to be in possession of corresponding statistical evidence, and is compatible with a proposition-based as well as a variable-based approach. Mosquito-bite claims are based on a more fundamental experience of causation than comparing a state of affairs with a counterfactual. But they do have counterfactual implications.

#### **29.1.4 Can include claims about absence of expected states**

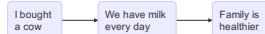
I was expecting support from a safety net program but they excluded me from being a beneficiary so I am in financial difficulties

There is no reason not to code this kind of claim; the influence item is still a Difference, a difference from an expected state. It is no more or less hypothetical than any other causal claim.



## 29.2 Mini Extension: causal chains

Often we have to code narratives which take the form of chains. In this case, as the intervening item is one and the same proposition, we can join the two mini-maps into a causal chain. In practice we can do this as a single piece of coding. This is what QuIP has always done.



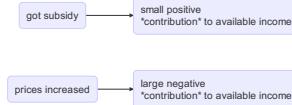
This involves what I have elsewhere called the “chaining rule”. This also implies that we can deduce that as B causes C and C causes E, also B causes E. Of course this might seem naïve, of course other things explain my family’s health, but then all causal claims like this are naïve. Indeed, in QuIP, C is offered as an explanation of E and B is offered as an explanation of C.

## 29.3 Extension: propositional claims about *additive contributions*

The small government subsidy helped, but it wasn’t enough. Due to big price rises, my available income still went down a bit.

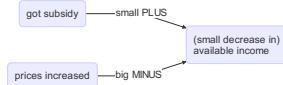
It isn’t obvious how to do this with the propositional approach, keeping the simplicity of “this causes that”.

We can write this:



The contributions, on the right, are expressed as simple propositions; our respondent can really provide evidence about each contribution, for example shows us the receipt from the subsidy and the record of money spent in shops. (This is only possible because “available income” is conceived of as an additive, linear combination.)

But there is another fact here, the resulting small decrease in available income, which we can also measure and report independently and express as a proposition. And as if by magic, this decrease is equal to the price increase minus the subsidy. I’m not sure if this is due to a causal law or a definition (what would we do if we found there was a discrepancy?).



In this diagram, we have combined the two contributions into one resultant proposition. I’ve put the information about the actual result in brackets. You can think of it as just a summary of the two separate original claims as in the previous diagram, or you can read it as something new: “the combination of getting the subsidy and the price increases caused a small overall increase in available income”. The second version involves multiple causation.

It seems that the causal model in the background involves continuous variables too, but the actual causal claims are still expressed as propositions.

We have chosen to make the distinction about the size and polarity of the influences on the arrows rather than on the items. We have discussed this kind of notation before, we can standardise it by perhaps using a notation like “strength=.3” instead of “small PLUS”.

We have our first case of what looks like multiple causation, with two arrows pointing to the same item. But there is no need for all the general mechanics of multiple causation; we can treat these like two separate causal claims, as we did in the previous diagram. This is because of the special magic of assuming that the different influences combine separately, additively, without interaction. One influence doesn't need to know about the other influence.

Here, we *have* explicitly coded the fact that the respondent said that the overall result was negative, by writing “small decrease....” on the consequence item, but we didn't need to because anyone could have worked that out, from the fact that roughly speaking big MINUS + small PLUS = small MINUS.

In the app, it doesn't matter if you code the two contributions at the same time or separately. In each case, the consequence item is named as the resultant (in this case, small decrease), as long as the arrow mentions the actual contribution, like this:

```
got subsidy--small PLUS-->small decrease in available income  
prices increased)--big MINUS-->small decrease in available  
income
```

... and the individual claims wouldn't be false; the first could be read “got the subsidy, this made a small positive contribution to available income (which experienced a small resultant decrease overall)”. **You can even do this in Excel**, you don't need to link the two rows, as long as there is a way to express the strengths of the contributions in each row. In the app it is also possible to code both arrows at once, as you would more generally in the variable-based approach.

Remember, this is all about combining separate statements from the same source which mention a single consequence. This is not the same as aggregating different statements from different sources which happen to mention the same item. We'll look at that later.

### 29.3.1 Encoding *incomplete* claims about additive contributions

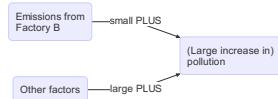
The river pollution levels have risen a lot; one part of the reason is that factory over there.

Factory B did not cause *much worse* pollution. There is an implicit claim that there are other causes, that the contributions add up in some way, and that the net effect is a large one. So we could code the above like this:



But it would be confusing because it suggests the small contribution causes the large change.

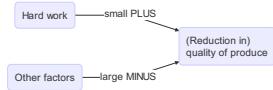
Perhaps we should code like this because there is really extra information which shouldn't go missing. I have put the information about the Difference on the consequence item in brackets.



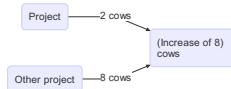
There are issues here about scaling – what counts as large, a lot, etc.

#### 29.4 Examples we can code

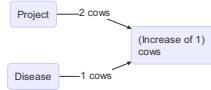
We work hard but the gain is very minimal and much less compared to what we invest. I don't know whether it is due to the climate change or the problem with the soil but the land isn't giving sufficient product.



The project gave us two cows, that was really useful, and the other project gave us eight.



The project gave us two cows, that was really useful, unfortunately one died of disease.



[More...]

## 29.5 Examples we cannot code

The project trained us about cows, but the other project never gave us any, so we got no benefit and no milk.

We can't code this with our simple additive model because the contributions are not independent. It is hard to say what the contribution of the project was to the outcome (which was zero).

## 29.6 That's it

That's it.

We can cover most claims, even most of the problematic ones, using these simple ideas. Anything which doesn't involve an interaction.

The result of several acts of causal coding is a heap of unconnected mini-maps and/or chains. We have no way so far of doing anything else with them, like aggregation. In fact, even classic QuIP goes beyond this, with (conservative) rules for joining individual maps. That's what we need to look at next.

## Section 30

# Combining causal fragments from different sources using propositions

### 30.1 Same item

There is an easy extension to this approach as already implemented in “classic QuIP”.

When two propositions from two different causal fragments can be considered more or less the same, create the combined map: join up the pieces on the common items. As discussed elsewhere there are four possibilities:

1. Chaining. (The common proposition is an influence in one fragment and a consequence in the other.)
2. Shared influence. (The proposition is an influence in both fragments; unproblematic.)
3. Shared consequence. (The proposition is a consequence in both fragments; what arithmetic do we use for this? The result doesn't have to be consistent: E was caused by B but no, it was caused by C? The composite map can't be considered as a larger claim that things are like this, only as an aggregate of different claims)
4. Shared arrow. (Combining fragments in which there is a shared influence and a shared confluence; here the issue is how we combine the arrows, e.g. making them fatter and saying “2 citations”.)

The first two are fairly unproblematic.

There is also the problem of the meaning of an individual item when it appears in, say, 20 different fragments from 20 different sources. Is it still a proposition, but a new one? Does it say “yields increased (for 20 farmers)”? Does it generalise to “yields in this area increased”? Does it merely say “yields increased (in 20 different cases)”?

[More to discuss here]

### 30.1.1 Extension 2: Subsuming items under more general items

This is also a part of classic QuIP. It makes it much easier to aggregate fragments, as above, by recoding *different but similar* items into broader categories.

There are some issues here too though. If we have

Heavier rains have led to worse crops

as well as

Hotter temperatures have led to drier ground

can we recode the left-hand side as “climate change” in both cases? It is not actually reflecting what the respondents said; they weren’t talking about climate change, just rains or temperatures.

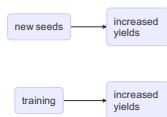
[More to discuss here]

### 30.1.2 Extension 3: Recoding items from different sources as gradations of more general items

(Respondent P:) here in village Q, the new seeds have doubled our yields

(Respondent R:) here in village S, the training has slightly increased our yields

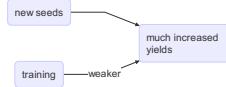
I don’t think classic QuIP does this. In classic QuIP, all we can do is this:



.... which can then be combined into one map.

This is a bit of a fudge, because the meaning of “increased yields” changes depending on how you look at it. It isn’t clear does it mean “at least some increase, no matter how small” or “some average-ish increase”?

The extension I suggest here allows us to explicitly encode the *degree* of a relationship:



We have discussed this before, perhaps using a notation like “strength=.3” instead of “weaker”.

Here we have two separate causal fragments from separate sources which we have combined afterwards using the “shared consequence” rule. The important (implied) step was first to recode this



as this



We still have propositions, and causal claims about simple links between them. But in the *aggregated* map we have combined two propositions into one. Creating something like a proto-variable. This trivial-seeming trick allows us to combine maps on propositions which are in a sense common, but differ by degree, without losing the information about the degree.

### 30.1.3 Extension 3a: allowing negative gradations

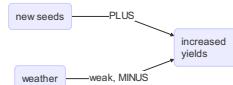
If we are going to go along with extension 3, there is no reason not to allow negative strengths too.

So we can encode this

(Respondent P:) here in village Q, the new seeds have doubled our yields

(Respondent R:) here in village S, the weather has made our yields slightly worse

as this:



This extension does however highlight the problem of the meaning of an aggregated item which represents “the same” item in several original sources; we don’t in general even necessarily know whether to phrase it as “increased yields” or “decreased yields”. We are really only justified in saying “changed yields”. We should not be trying to make an actual statement about any change in the total yields amongst many farmers. We don’t have enough of the right kind of data for that.

## 30.2 What these extensions still don’t do

This model allows for background causal maps in which all influences are of an additive (or subtractive) nature. Things make other things higher, or better, or worse, etc, perhaps more or less likely, and that’s it. There is no notion of the shape of the influence and (more importantly) there is no notion of any kind of interaction between multiple influences. So for example there is no room for the difference between AND and OR, see below. And there is no room for necessary or sufficient conditions.

Without wanting to, we have ended up with background theories like those common in natural science: separate, additive influences between numeric variables. This is because we wanted to keep things simple, and these kinds of models have emerged over millennia to fit that bill.

I will finish by pointing out a few of the things we can’t do.

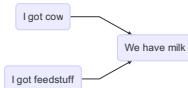
Other vegetables don't grow in our locality since the soil is swampy. The government agents particularly kebele principals gave support through distributing vegetables and fruits for planting. But we tried it and it couldn't grow in our locality but it was good for people in other kebeles.

This is an explicit AND claim (you have to have seeds AND good soil). Or it's an explicit claim that good soil is necessary but not sufficient.

### 30.2.1 Can't distinguish between AND and OR

I don't think this is a big deal. AND claims and OR claims do appear in QuIP transcripts. But not often enough to warrant a lot of attention to the difference between the two. For example, in the example of the cow and the feedstuff, it is probably enough to note that both the cow and the feedstuff have some kind of a PLUS influence on milk (though in fact the influence of one is not independent of the influence of the other).

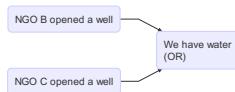
I got the cow, and I also got the feedstuff, which I couldn't have afforded. Thanks to both, my family now has milk.



In the example of the cow and the feedstuff, the underlying causal map is based on a conjunction; E is true if and only if B is true and C is true. There is no way we could deduce this underlying map from this single observation, and there is no way the respondent could either. The respondent is gifting us the implied underlying causal knowledge, just as they did in the simpler version of the coding rules, above. The *two propositions which they report to us* do not include anything about, for example, cases in which there was a cow but no feedstuff; but the implied causal map does.

This is the subtle difference to the case below: here too, all three propositions are true. As in the case above, the propositions only say that B and C are true, and E is true, and that if B and C hadn't been true, E wouldn't be true. But the underlying maps which we infer from the narrative and the context give us the additional information which distinguishes between AND and OR; what would happen if B was true and C false, or vice-versa.

We had no clean water, and then two different NGOs came and dug wells! We'd only really need one of them, but anyway we are happy now.



### 30.2.2 Necessary / sufficient conditions

Reminder:

- a report of a necessary condition being fulfilled says: B happened, E happened, and B is a necessary cause of E, which means that E couldn't have happened without B. Or: if B doesn't happen, E doesn't happen, but *I don't know what happens if B does happen*.
- a report of a sufficient condition being fulfilled says: B happened, E happened, and B is a sufficient cause of E, which means that NOT-E couldn't have happened without NOT-B. Or: if B does happen, E does happen, but *I don't know what happens if B doesn't happen*. So in a sense a sufficient claim, like a necessary claim, is weaker than an ordinary causal claim.

We could argue: we can't encode this kind of information, because a respondent can't deduce something as advanced as a necessary or sufficient condition on the basis of a single story. They have to have more information, e.g. more observations, to be sure of this claim.

But we can say the same about ordinary causal claims too: they can't be justified on the basis of a single observation or report either.

So we *could* code like this: B happened, and E happened, and B is a necessary cause of E. We encode the necessary/sufficient information not in the way we encode the propositions but in the way we encode the background causal information. As encoding a causal link always implicitly involves encoding the respondent's background causal knowledge, why shouldn't that knowledge include necessary or sufficient links too, as well as about ordinary causal links?

The biggest problem I have with actually coding these links is that whether you think of a condition as necessary or sufficient depends on the contrast you happen to be thinking of.

We really wanted milk. We have the feedstuff and an empty shed, we just didn't have the cow. The cow was necessary, the missing piece. No other kind of help (e.g. livelihood training) would have done this.

or

We really wanted milk. We have the feedstuff and an empty shed, we just didn't have the cow. The cow was sufficient. There are other kinds of help (e.g. deliveries of milk to the door) which would have done this; but the cow was enough.



## Section 31

# Coding causal maps with propositions

Propositional maps differ from variable-based maps in many ways. Propositional maps encode answers to the question: what happened here and what caused it? Variable-based maps encode answers to the question: what causes what around here?

Here, I will discuss the most basic coding rule for bare-bones propositional causal maps, and discuss some **extensions** to it.

Even with the extensions, you still keep the core advantages of the propositional approach: you have propositions (and sets of propositions) causing other propositions. You are still dealing with historically true stories, not chatting about abstract models or what-might-happen-if. You still have something closer to an evaluation than to research.

[But: I argue that even the most basic causal coding anyway presupposes an understanding of an underlying causal claim behind the propositional claims, and that those causal claims always go beyond any specific piece of evidence. So it is not the case that “ooh, these extensions are risky, they go beyond the evidence to talk about abstract stuff”. Because the simplest propositional causal claim relies on background models of how things work.]

### 31.1 Basic coding rule

The basic coding unit is of the form “B causes E”.

B and E are both propositions. Both are true.

A proposition like B or E only has two possibilities; it is the case (which indeed it is), or it is not the case.

A proposition expresses a *Difference*. A good example of a Difference is a change, but other things can be Differences too (states, events, etc). It says that things are one way rather than some other, perhaps expected, contrast.

Examples

- I bought a cow (as opposed to not buying one)

- Yields have got worse (as opposed to not getting worse)
- I am fairly happy (as opposed to not being fairly happy)
- There is unemployment in the country at the moment (as opposed to there not being unemployment)

This also means that, seen like this, the items have *no gradation*. They are always one thing or the other. There are no possibilities in between. So “I am fairly happy” is not specifically contrasted with, say, “I am very happy”, but with any circumstances in which “I am fairly happy” is not true.

We have no way to code a proposition on its own; only as part of a causal statement.



We can be vague about the timing of things. The consequence has to come after the influence, or be concurrent in such a way that effects are still happening after causes.



### 31.1.1 Three claims

The basic coding rule encodes three claims:

1. the influence item is true
2. the consequence item is true
3. *the first caused the second*

This last claim is really important. We do not encode simple observations that things merely happen together. We are in fact encoding, in a minimal way, the

causal knowledge of the respondent, a minimal part of her implicit causal map of the world. There is no way that merely observing B and E on their own warrants the causal conclusion.

The way that people tell stories usually implies beliefs that there are causal connections between the links of the story. We can usually assume that causal claims are being made. They don't have to be made explicitly. "I got the cow. Now we have milk."

### **31.1.2 The items have been witnessed**

In general, the propositions are / have been **personally true for, or at least personally experienced by the source** or sources. We treat the source as a **witness**.

### **31.1.3 The propositions have contrastive implications**

If someone makes a claim of the form that "X caused Y", then this implies that if X had not happened in that way, Y wouldn't have happened, and that if Y hadn't happened, X can't have happened. This is a strong claim: within this narrow context, holding everything else constant, X is necessary and sufficient for Y. Whenever X is true, Y is true, and vice versa.

This is quite another thing from saying that the people making it have to be in possession of corresponding statistical evidence, and is compatible with a proposition-based as well as a variable-based approach. Mosquito-bite claims are based on a more fundamental experience of causation than comparing a state of affairs with a counterfactual. But they do have counterfactual implications.

### **31.1.4 Can include claims about absence of expected states**

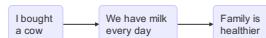
I was expecting support from a safety net program but they excluded me from being a beneficiary so I am in financial difficulties

There is no reason not to code this kind of claim; the influence item is still a Difference, a difference from an expected state. It is no more or less hypothetical than any other causal claim.



### 31.2 Mini Extension: causal chains

Often we have to code narratives which take the form of chains. In this case, as the intervening item is one and the same proposition, we can join the two mini-maps into a causal chain. In practice we can do this as a single piece of coding.



This involves what I have elsewhere called the “chaining rule”. This also implies that we can deduce that as B causes C and C causes E, also B causes E. Of course this might seem naïve, of course other things explain my family’s health, but then all causal claims like this are naïve.

### 31.3 Mini Extensions: certainty and trust

There is no reason why we shouldn’t also encode additional attributes like the respondent’s degree of certainty:

I think the new irrigation methods improved my yields (this is the kind of thing which ought to work, and the results were encouraging) but I’m not really sure. [Coding: low certainty]

Or we can code the coder’s trust in the statement:

I saw a documentary and the state of the crops is definitely due to contrails. [Coding: low trust]

But as we have no way to even aggregate individual causal fragments, – no arithmetic for them – encoding certainty or trust amounts to no more than making a note alongside the fragment. They are not very interesting.

### 31.4 How does classic QuIP extend this basic idea?

That’s it. The result of one act of causal coding is just one mini-map, or a chain of them. The result of several acts of causal coding is a heap of unconnected mini-maps and/or chains. We have no way so far of doing anything else with them, like aggregation. We have no way of encoding multiple causation.

In fact, even classic QuIP goes beyond this.

### 31.4.1 Extension 1: Combining causal fragments from different sources: same item

There is an easy extension to this approach as already implemented in “classic QuIP”.

When two propositions from two different causal fragments can be considered more or less the same, create the combined map: join up the pieces on the common items. As discussed elsewhere there are four possibilities:

1. Chaining. (The common proposition is an influence in one fragment and a consequence in the other.)
2. Shared influence. (The proposition is an influence in both fragments; unproblematic.)
3. Shared consequence. (The proposition is a consequence in both fragments; what arithmetic do we use for this? The result doesn't have to be consistent: E was caused by B but no, it was caused by C? The composite map can't be considered as a larger claim that things are like this, only as an aggregate of different claims)
4. Shared arrow. (Combining fragments in which there is a shared influence and a shared confluence; here the issue is how we combine the arrows, e.g. making them fatter and saying “2 citations”.)

The first two are fairly unproblematic.

There is also the problem of the meaning of an individual item when it appears in, say, 20 different fragments from 20 different sources. Is it still a proposition, but a new one? Does it say “yields increased (for 20 farmers)”? Does it generalise to “yields in this area increased”? Does it merely say “yields increased (in 20 different cases)”?

[More to discuss here]

### 31.4.2 Extension 2: Subsuming items under more general items

This is also a part of classic QuIP. It makes it much easier to aggregate fragments, as above, by recoding *different but similar* items into broader categories.

There are some issues here too though. If we have

Heavier rains have led to worse crops

as well as

Hotter temperatures have led to drier ground

can we recode the left-hand side as “climate change” in both cases? It is not actually reflecting what the respondents said; they weren't talking about climate change, just rains or temperatures.

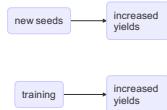
[More to discuss here]

### 31.4.3 Extension 3: Recoding items from different sources as *gradations* of more general items

(Respondent P:) here in village Q, the new seeds have doubled our yields

(Respondent R:) here in village S, the training has slightly increased our yields

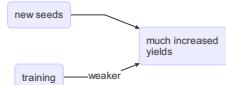
I don't think classic QuIP does this. In classic QuIP, all we can do is this:



... which can then be combined into one map.

This is a bit of a fudge, because the meaning of "increased yields" changes depending on how you look at it. It isn't clear does it mean "at least some increase, no matter how small" or "some average-ish increase"?

The extension I suggest here allows us to explicitly encode the *degree* of a relationship:



We have discussed this before, perhaps using a notation like "strength=.3" instead of "weaker".

Here we have two separate causal fragments from separate sources which we have combined afterwards using the "shared consequence" rule. The important (implied) step was first to recode this



as this



We still have propositions, and causal claims about simple links between them. But in the *aggregated* map we have combined two propositions into one. Creating something like a proto-variable. This trivial-seeming trick allows us to combine maps on propositions which are in a sense common, but differ by degree, without losing the information about the degree.

#### 31.4.4 Extension 3a: allowing negative gradations

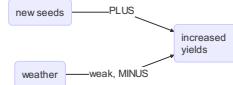
If we are going to go along with extension 3, there is no reason not to allow negative strengths too.

So we can encode this

(Respondent P:) here in village Q, the new seeds have doubled our yields

(Respondent R:) here in village S, the weather has made our yields slightly worse

as this:



This extension does however highlight the problem of the meaning of an aggregated item which represents “the same” item in several original sources; we don’t in general even necessarily know whether to phrase it as “increased yields” or “decreased yields”. We are really only justified in saying “changed yields”. We should not be trying to make an actual statement about any change in the total yields amongst many farmers. We don’t have enough of the right kind of data for that.

### 31.5 That’s still not enough

Above, we aggregated mono-causal links or chains into maps with multiple links to individual items. But that isn’t the same as encoding individual respondent statements which themselves make claims of multiple causation.

I’ve had a good hack through the Save the Cow transcripts and these cases come up just too often and there is no adequate way to code them at all.

#### 31.5.1 No way to encode explicit claims of multiple causation

The small government subsidy helped, but it wasn’t enough. Due to big price rises, my available income still went down a bit.

We can’t encode that as

```

subsidy --> available income increased
price rises --> available income decreased
  
```

because we’d be contradicting ourselves about available income.

We can encode it as

```
subsidy + price rises --> available increase decreased a bit
```

but then the left-hand side is just a monolithic block. We can’t get at the bits, for example we can’t combine it with another map which just mentions the subsidy.

### 31.5.2 No way to encode implicit claims of multiple causation

The river pollution levels have risen a lot; one part of the reason is that factory over there.



The respondent is a witness to both propositions, but the causal claim is of a *contribution to a whole*. Factory B did not cause much worse pollution. There is an implicit claim that there are other causes, that the contributions add up in some way, and that the net effect is a large one.

We could chicken out and just code like this:



in which the “pollution worse” is ambiguous about how much worse. The contribution is *small* while the net effect is *large*; which claim is being made? To understand the consequence item, we have to shift between the two interpretations.

### 31.6 Solution: propositional claims about contributions, and assume additive causation

The small government subsidy helped, but it wasn’t enough. Due to big price rises, my available income still went down a bit.

We can encode that like this:



Here we are encoding the influence of two propositions on a third. All three propositions are simple and true. But where necessary, we explicitly note that the consequence items **are *contributions* rather than end results**.

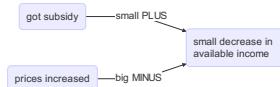
We haven't explicitly coded the fact that the respondent said that the overall result was negative, but we can see that, from the fact that big MINUS + small PLUS = small MINUS.

We have chosen to make the distinction about the size and polarity of the influences on the arrows rather than on the items.

Our respondent can really provide evidence about each contribution, for example shows us the receipt from the subsidy and the record of shop prices.

There is no need for all the general mechanics of multiple causation; we can treat these almost like two separate causal claims. Only the word "contribution" reminds us that this consequence item can have multiple causes. This is because of the special magic of assuming that the different influences combine separately, additively, without interaction. We don't even need to have two arrows pointing to one box.

But in practice we can treat the following as equivalent to the above:



Here, we *have* explicitly coded the fact that the respondent said that the overall result was negative, but we could have worked that out, from the fact that big MINUS + small PLUS = small MINUS.

You could say there are **five** propositions here:

1. got subsidy
2. small PLUS contribution of subsidy to available income
3. prices increased
4. big MINUS contribution of prices to available income
5. small resultant decrease in available income

1 causes 2, 3 causes 4, and 4 and 2 taken together cause 5 (or perhaps this is a definitional rather than causal link). In the background is an implicit causal theory which we can model in a number of equivalent ways. The simplest is probably to say that in the background there are continuous variables, and the resultant decrease is calculated according to a function which is always the same: just a generic kind of addition / subtraction.

In other words, we have used essentially the same approach as that used above to combining separate statements from separate sources which happened to mention “the same” consequence. These rules are confusingly similar, but really not the same!

I think this is as far as QuIP wants to go. It covers nearly all the tricky statements I've seen so far.

I think this is probably where QuIP should stop. Enough extensions already.

### 31.7 What these extensions still don't do

This model allows for background causal maps in which all influences are of an additive (or subtractive) nature. Things make other things higher, or better, or worse, etc, perhaps more or less likely, and that's it. There is no notion of the shape of the influence and (more importantly) there is no notion of any kind of interaction between multiple influences. So for example there is no room for the difference between AND and OR, see below. And there is no room for necessary or sufficient conditions.

Without wanting to, we have ended up with background theories like those common in natural science: separate, additive influences between numeric variables. This is because we wanted to keep things simple, and these kinds of models have emerged over millennia to fit that bill.

I will finish by pointing out a few of the things we can't do.

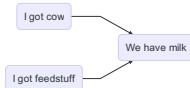
Other vegetables don't grow in our locality since the soil is swampy.  
 The government agents particularly kebele principals gave support through distributing vegetables and fruits for planting. But we tried it and it couldn't grow in our locality but it was good for people in other kebeles.

This is an explicit AND claim (you have to have seeds AND good soil). Or it's an explicit claim that good soil is necessary but not sufficient.

### 31.7.1 Can't distinguish between AND and OR

I don't think this is a big deal. AND claims and OR claims do appear in QuIP transcripts. But not often enough to warrant a lot of attention to the difference between the two. For example, in the example of the cow and the feedstuff, it is probably enough to note that both the cow and the feedstuff have some kind of a PLUS influence on milk (though in fact the influence of one is not independent of the influence of the other).

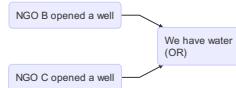
I got the cow, and I also got the feedstuff, which I couldn't have afforded. Thanks to both, my family now has milk.



In the example of the cow and the feedstuff, the underlying causal map is based on a conjunction; E is true if and only if B is true and C is true. There is no way we could deduce this underlying map from this single observation, and there is no way the respondent could either. The respondent is gifting us the implied underlying causal knowledge, just as they did in the simpler version of the coding rules, above. The *two propositions which they report to us* do not include anything about, for example, cases in which there was a cow but no feedstuff; but the implied causal map does.

This is the subtle difference to the case below: here too, all three propositions are true. As in the case above, the propositions only say that B and C are true, and E is true, and that if B and C hadn't been true, E wouldn't be true. But the underlying maps which we infer from the narrative and the context give us the additional information which distinguishes between AND and OR; what would happen if B was true and C false, or vice-versa.

We had no clean water, and then two different NGOs came and dug wells! We'd only really need one of them, but anyway we are happy now.



### 31.7.2 Necessary / sufficient conditions

Reminder:

- a report of a necessary condition being fulfilled says: B happened, E happened, and B is a necessary cause of E, which means that E couldn't have happened without B. Or: if B doesn't happen, E doesn't happen, but *I don't know what happens if B does happen.*
- a report of a sufficient condition being fulfilled says: B happened, E happened, and B is a sufficient cause of E, which means that NOT-E couldn't have happened without NOT-B. Or: if B does happen, E does happen, but *I don't know what happens if B doesn't happen.* So in a sense a sufficient claim, like a necessary claim, is weaker than an ordinary causal claim.

We could argue: we can't encode this kind of information, because a respondent can't deduce something as advanced as a necessary or sufficient condition on the basis of a single story. They have to have more information, e.g. more observations, to be sure of this claim.

But we can say the same about ordinary causal claims too: they can't be justified on the basis of a single observation or report either.

So we *could* code like this: B happened, and E happened, and B is a necessary cause of E. We encode the necessary/sufficient information not in the way we encode the propositions but in the way we encode the background causal information. As encoding a causal link always implicitly involves encoding the respondent's background causal knowledge, why shouldn't that knowledge include necessary or sufficient links too, as well as about ordinary causal links?

The biggest problem I have with actually coding these links is that whether you think of a condition as necessary or sufficient depends on the contrast you happen to be thinking of.

We really wanted milk. We have the feedstuff and an empty shed, we just didn't have the cow. The cow was necessary, the missing piece. No other kind of help (e.g. livelihood training) would have done this.

or

We really wanted milk. We have the feedstuff and an empty shed, we just didn't have the cow. The cow was sufficient. There are other kinds of help (e.g. deliveries of milk to the door) which would have done this; but the cow was enough.



## Section 32

### Types of variable

Above, I have briefly explained why I have chosen *variables* as the best candidate for the nodes in our causal maps.

So here are a few points about variables.

We will need to distinguish different types of variable. Here are a few key types:

- continuous, limited variables like percentage, usually specified as going from zero (“nothing”) to 1 (“everything”). I call them “[lo-hi variables][#lo-hi]” but I would love to hear of a better name. Can also be expressed as a %, because this is more familiar to people. Our interpretation of these numbers is very loose, they might be proportions or probabilities or numbers expressing membership of a fuzzy set Zadeh (1973), Ragin (2008).
- false/true variables like “the project is implemented (yes or no)”
- continuous, unlimited variables like height, income
- ..... and various others, see xx.

When actually coding causal maps, we will mostly use lo/hi variables, as false/true variables are a special case of them, with just the levels 0 (no) and 1 (yes).

## Section 33

# Causal thinking is essentially contrastive thinking

Variables contrast the actual (or imagined) state of things with possible alternatives.

All causal claims intrinsically have contrastive meaning. (I say “contrastive” rather than “counterfactual” as the latter would more strictly only refer to past things which can no longer be changed). If A claims that contrails causally influence the weather, and in particular that if there are a lot of contrails, the weather will be worse, but has no opinion at all about what the weather would be like (controlling for any other influences) if there were fewer contrails, in particular doesn’t claim it would be any better, then they haven’t understood what “causal influence” *means*. In other words a useful and comprehensive general framework for causal maps in social science can start by treating the elements as *variables* in a broad (and mathematical rather than statistical) sense, i.e. as things that can / could be / could have been different.

[xx add more]

## Section 34

# Lo/hi Variables, types of variable, and contrasts

My strong suggestion will be that we can make a lot of sense of causal maps if we think of the nodes/factors/vertexes/tags as variables *taking values between 0 and 1* (or maybe -1 and 1). I call variables of this kind “lo/hi” variables for the want of a better word.

Variables of this type are common in some parts of social science, e.g Kosko (1986).

Standardising our causal maps to use these kinds of variables in most cases will make most tasks a lot easier, in particular the hard task of specifying a suitable “soft arithmetic” for combining and simplifying hosts of causal fragments.

We can say, “confidence in the President is around 20% whereas for the former President it was around 50%” or “the project was performing at about half of its potential” even if we don’t make it clear what the numbers mean exactly. This method is pretty good for expressing a lot of the variables we encounter, so wellbeing of .9 would be pretty amazing, 0 is impossibly bad.

This value can have quite general meaning:

- the value within an empirical range e.g. “this summer is as hot as we have ever had” (this summer temperature .95)
- the value without reference to an empirical range e.g. “the teachers’ skills were appalling” (teachers’ skills .1)
- the proportion of a set of things which have a binary property e.g. “most of the children seem happy” (happiness of the children .75)
- the strength of the membership of something in a certain set e.g. “country X is only partly democratic” (country X democracy level .5)
- the probability of a binary variable e.g. “the chances of war are now very low” (chance of war .1)
- the strength of our information about a binary variable

The point is our respondents will anyway use such language which we need to reproduce, and also that when coding and then combining causal fragments, we

sometimes have to commit to expressing a non-numerical claim with a rough number.

These ranges might also be tied to some kind of empirical distribution, so “Income from farming = .95” would mean it is in the middle of the top 10%, etc.

The terrible danger of these kinds of numbers is that if we use them, someone will accuse us of trying to achieve an unjustified level of “accuracy”. Whereas we don’t in fact want accuracy. But we don’t want to lose the kind of broad-brush information about small, large, largest, negligible, which our respondents are actually telling us. My suggestion is to hardly ever show the numbers and use them only for our “internal” calculations. Occasionally it might be useful to use semi-standardised phrases like “very high” / “high” etc, and/or empirically-grounded phrases like “in the bottom 10%”.

Another advantage of lo/hi numbers is that not only the current *value* of variables but also in some sense the *strength* of arrows can be captured with them (though we might like to encode the strength of arrows in a broader range, from **-1**, rather than 0, to 1).

We can directly interpret lo/hi information about the effect of a causal package on a consequence variable as an analog to *effect size*: 1 means total control, 0 means none. When we “zoom out” to ask about direct effects of a variable or package on variables further downstream of it, this effect size information is particularly interesting. It is the key to answering a core demand on our app and on reasoning with causal maps; to be able to say that, for example, variable B is a much more important influence on outcome O than variable C.

We can already start to guess how this kind of soft arithmetic will work – for example, if E is already at 90% and C has a 100%, positive, influence on E, and then we set C to 90%, we are at least sure E isn’t going to decrease and we’d expect it increase a bit. Whereas if the influence of C on E is strongly negative, say -90%, then if we set C to be large rather than small, we’d probably expect E to drop.

The *details* of how this arithmetic might work are up for grabs. But I will try to sketch out some solutions in the next sections.

It is relatively simple to extend this idea to the range -1 to 1, i.e. including the idea of variables with negative levels.

We already encode *arrows* as having *negative* influences.

## **Section 35**

### **“For each .....” variables**

We also need to deal with repetition and multiple cases (e.g. when we know about the effects of X on *a* child's health, but we have many children in the programme.....). In particular: - when we have more than one level e.g. multiple children embedded in multiple classrooms - when we have higher-level effects e.g. when classroom effects are not just an aggregate of child effects - time / duration (for each day, for each semester .....)

[xx add more]

## Section 36

# Coding specific influences – individually and in combination

### 36.1 Summary of the argument so far

I argued<sup>1</sup> that as evaluators and social scientists we are faced with a bag full of different fragments of causal information which most often will mention several different factors or variables. We need to understand how to encode and synthesise such fragments. In particular, we can use causal maps to show how the fragments link together. Causal maps (actually, directed causal networks) are great because they show how our (putative) knowledge of the world is constructed from relatively stable, portable, fallible, snap-together pieces.

I argued that we need rules to translate backwards and forwards between fragments of (written) causal information and causal maps. In the one direction, the question is how to encode causal information in the form of maps; in the other direction, the question is how to interpret those maps.

Most generally, we need rules about how to *reason* with causal maps (and therefore with the sets of causal fragments which they encode) – how to see if one map, e.g. a summary map, (or a summary statement) can be deduced from another. I argued that showing how to reason with causal maps is the same thing as explaining the meaning of their constituent parts and conventions. The maps which result when we encode a causal fragment in a causal map, or join up two mini-maps to make a larger one, *mean* something; but what are the translation rules?

I have presented some of these rules, part of what I call “Soft Arithmetic” - e.g. the basic rule for coding a single causal fragment, the rules for joining smaller maps together, etc.

Quantitative social scientists have desks covered with numbers and they have arithmetic and statistics to help them make deductions with them.

We have desks covered with causal fragments. This is primary information and cannot be reduced to measurements of the values of variables and observed

---

<sup>1</sup>There aren't any references to literature in this summary, but there are some in the preceding sections

correlations between them. We start at (putative) causality, the point which pre-causal, quantitative social scientists set out for but never reached. We use an arithmetic for causal fragments, which I am calling Soft Arithmetic. We use it anyway when we reason with fragments of causal evidence, so we may as well spell it out.

I argued that looking at the *structure* of the maps alone (what links to what) is not enough. (I think this is Rick Davies' approach; or maybe he does indeed take note of the contents of the causal functions but makes the reasonable assumption that all the functions are simple “increasing” or “PLUS”.) Sometimes we also need to ask and answer questions about comparisons (“Is this outcome higher in this situation than in that situation, e.g. without the project?”), comparisons of comparisons (“Is the influence of this project as big as the influence of that project?”), and maybe absolute judgements of the form “This factor is high” or “This influence is large”. Such questions cannot in general be answered purely by looking at the structure (there are a few, relatively trivial exceptions. For example, we can merely count the number of mentions of a causal link, but by ignoring the contents of those causal links, e.g. “this is necessary for that”, or “this decreases that”, we cannot draw many very interesting conclusions. We only know that people said that this factor was somehow causally relevant to that factor, but we don't know anything about *how* it is causally relevant, which is usually what we want to know.

I argued that the requirement to be able to make some sorts of comparisons (and especially, comparisons of comparisons) is logically equivalent to the requirement to encode the values of the variables in our maps using numbers of some sort. (Though these need not be numbers from the usual space of interval numbers like -7 and 504.34; in particular, I am interested in variables whose values vary between 0 and 1.)

You are now getting justifiably uneasy at this talk about applying numbers to causal maps and the conclusions we draw from them. But:

- we are fully aware of the fragility of most such conclusions
- we will rarely mention actual numbers, we only want to make judgements (like “B is probably larger than C”) on the basis of them
- we will be boringly insistent in frequently saying “the evidence isn't strong enough, I'd rather not make any comparison here”
- but crucially, we are aware of our responsibility not to *always* refrain from judgement and make *too* many type-II errors; sometimes the evidence, encased as it is in many layers of fuzzy caveats, is so strong that we have to make a judgement: yes or no, high or low, and to help our clients and readers to do so too. And from a Bayesian perspective, the right question is: what's your best bet, experts in QuIP or Outcome Harvesting or whatever, does the program work or not? Hiding behind the null hypothesis is so last-century; and it is irresponsible too.

So we also need rules about how to encode information about quantity and extent, and direction / shape of causal influence.

The rules of Soft Arithmetic are actually embedded in the app. So when you ask the app to show, for example, the effects on E of C and B, these are the rules which are being followed.

In the first part of this guide, I talked generically about the “content” of a causal link and about the function which it embodies. I then went on to look at those causal contents. I suggested that they can be understood as very general functions from an influence variable (or a package of influence variables) to a single consequence variable. This function is most often vague, ill-defined, and not obviously numerical. So a dance student who copies a teacher’s moves is embodying the function “do the same as” which maps the space of teacher-moves onto the space of student-moves. It’s quite easy to see when the student is keeping or breaking the rule, even though we’d never be able to formulate it numerically. Yet, it is still a function.

Ideally, Soft Arithmetic would help us deal with any kind of function including those which are completely non-numerical; but for the purposes of this guide (and the app) I am henceforth going to stick to “lo/hi” variables and the functions aka causal links between them. “Lo/hi” variables vary between 0 (low) and 1 (high), similar to those found for example in Fuzzy Cognitive Maps and related approaches. A large proportion of the functions which we are likely to want to code can in fact be coded using quite a small set of functions (four so far, including “necessary” and “sufficient”), leading from (packages of) lo/hi influence variables to lo/hi consequence variables.

Contrast here with Goertz and Mahoney (2006)

That’s the end of this summary of the argument so far; now I will proceed with developing the next part of Soft Arithmetic.

## 36.2 Encoding causal influences between lo/hi variables

How can we possibly have enough symbols to write down the different ways one or more variables can influence another? Even with variables which have rational numbers as their values and even with the strongest linear assumptions there are uncountably many such functions and combinations: exponential, interacting, chaotic, and so on.

In the most basic situation, we have information about some unspecified influence of one “influence variable” on one “consequence variable”.



The influence is hardly ever completely unspecified. We might hear things like this:

- You can't have E without B
- Mostly, Bs are Bs
- More of B means more of E
- B happened and then E happened
- E was poor because of B
- B is good for E
- B reduces E
- B is quite good for E
- B doesn't have much effect on E

More generally there is more than one influencing variable

- B and C combine synergistically to influence E
- B and C have contrary influences on E
- You can't have high performance without both high effort and high ability;

and so on. Yet, we are unlikely to ever hear a neatly-specified function. How to encode this information?

Also, (a related problem): how to make sense of any kind of *weight* information like strength, trust, confidence, reliability, probability, class membership etc. Obviously these qualifiers do not all mean the same thing, but we will make a start by looking at how to implement even some generic version. I will show a way to *reduce the influence* of information which is deemed to be, for example, less reliable or trustworthy or simply expresses a weaker connection.

### 36.3 Breaking down the problem

What we will do is restrict and reduce the problem as follows:

- All variables are lo/hi variables, taking values between 0 and 1. This also includes false/true variables as a special case, with values just 0 or 1.
- For the moment, we will deal only with symmetrical or commutative functions in the sense that if  $E = f(B, C)$  then also  $E = f(C, B)$ . For example, this is true for addition but not for division. [am I sure about this?? we surely want to be able to code that one variable is a suppressor in a package]
- We will specify a limited (as small as possible) bunch of functions to encode causal information which are designed for lo/hi variables, i.e. they are all defined on the same domain and range, so they all tell you how to go from values between 0 and 1 to produce another value between 0 and 1. These are functions such as AND, MULTIPLY, NECESSARY etc, chosen so that in combination with the attributes “strength” and

“contour”, we can code a good proportion of the functional relationships we are likely to encounter.

This also means that rather than defining a function like NECESSARY specially for the case of just binary 0/1 variables, we will straight away provide a generalisation which can also deal with intermediate values, and therefore deal with binary variables in particular, as a special case.

We will also, to keep our zoo of functions as lean and powerful as possible, we will try to provide functions (like MULTIPLY) which can be defined all at once for any number of influence variables. (Just as with ordinary arithmetic we don’t need to provide separate definitions for the “SUM” in  $SUM(b, c)$  and in  $SUM(b, c, d)$  just because there are different numbers of arguments.)

Thus armed, we will be able to tackle the thorny problem of how to combine *different* causal packages influencing the *same* variable (see earlier), making allowance for any weightings; and for the resulting causal maps, we will be able to “soft-calculate” the downstream effects of any given setting of the exogenous (“no-parent”) variables. More generally, we will be able to say things like, for example, “the effect of B on W, way downstream of it, is likely larger than the effect of C on W, especially when Z is hi”.

### 36.4 Some “mono functions”: functions with a single influence variable

Here are four fairly obvious “mono functions”: PLUS, MINUS, NECC and SUFF.

They are a formalisation of some really familiar ideas, which will probably seem standard and familiar in some ways but not in others (in particular, I am not aware of other similar attempts to bring necessary and sufficient conditions together with more ordinary “increasing” and “decreasing” functions together under one roof). I will explain them by means of graphs.

The first set of graphs are derived from the same simple situation in which one variable influences one other variable, as in the diagram below. I’ve called this function “PLUS” because it tries to force the consequence variable to take the very same value as the influence variable. It could also be called IDENTITY or perhaps POSITIVE or INCREASE.

This is the first such diagram in which the specification of the function is not preceded by three dots, because we are actually using a function (“PLUS”) defined within Soft Arithmetic. When coding, this corresponds to the difference between merely adding some notes about the function (information which the app cannot make use of) and using a drop-down input widget so that the app “knows” which function is meant and can use this information in subsequent calculations.

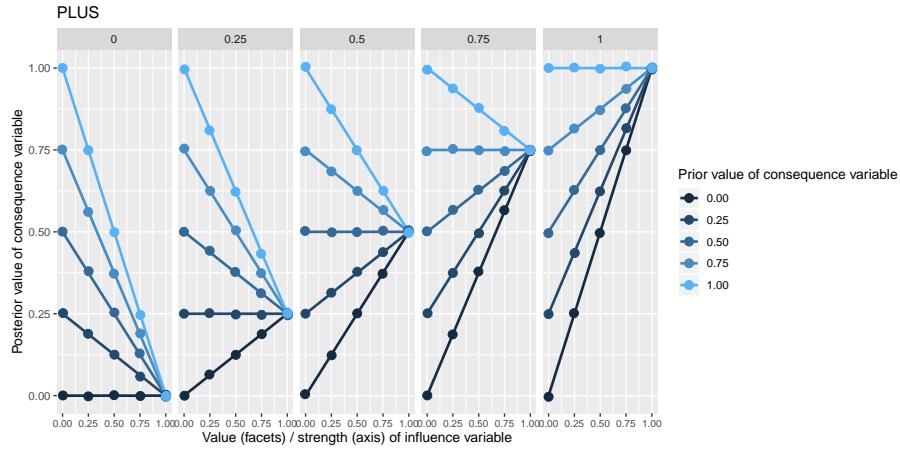


Below there is a set of graphs which show the value of G (on the y axis) given the value of the single influence variable (the five different facets or sub-graphs) and the strength from 0 to 1 of that influence, shown on the x-axis of the individual sub-graphs.

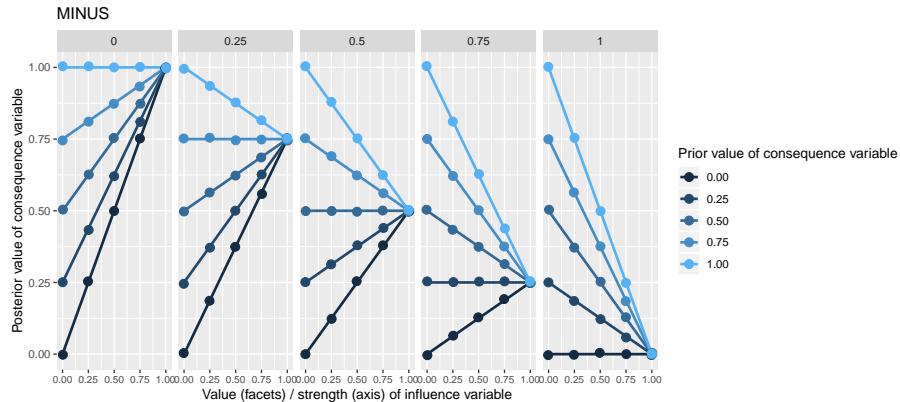
The different coloured lines are generated by different possible *prior* values of Y, i.e. values which it would have taken if X hadn't been influencing it.

Only five different values are used here for the influence variable and the prior value of the consequence variable, and for the strength of the influence, but any arbitrary number of values between 0 and 1 is valid.

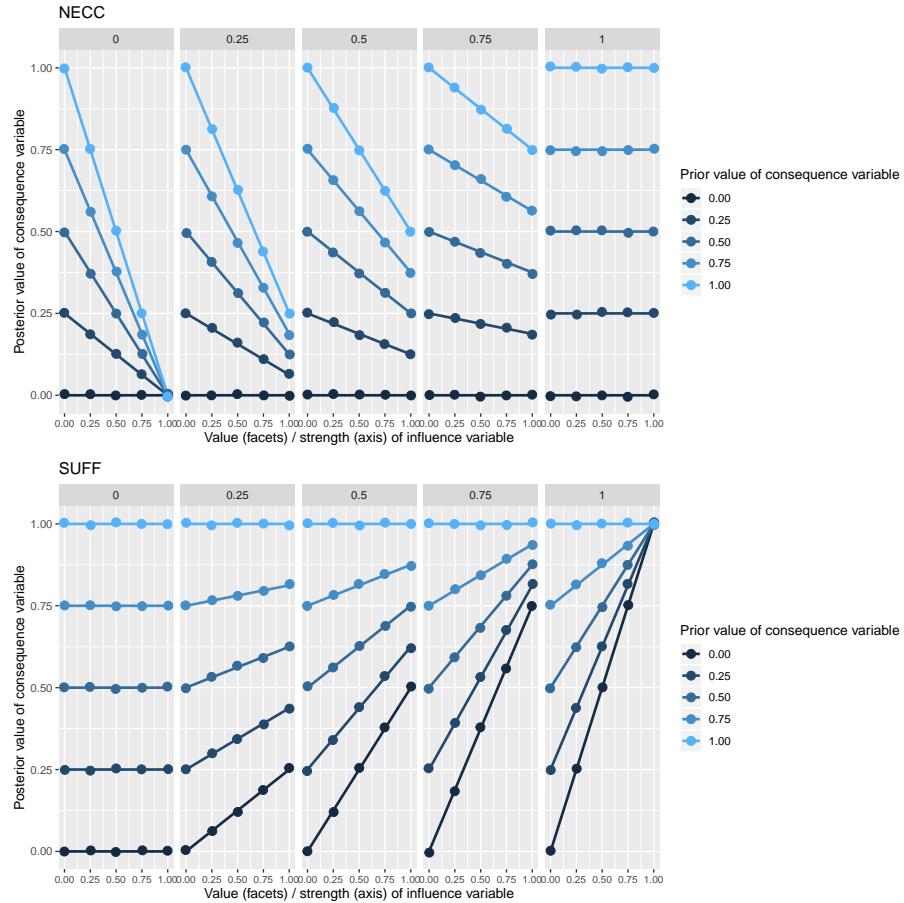
So in the right-hand column of each facet, where strength=1, we claim that the training completely determines the quality of the farming – a very unlikely situation; i.e. the “strength” of the function is set to 1. The function PLUS says that in these positions, the influence variable completely determines the consequence variable in the simplest sense that it is forced to have the same value as the influence variable. All the coloured points are overlaid on top of one another (they are slightly “jittered” randomly to make this overlaying visible) because the prior value of the consequence variable has no influence on its posterior value, and there are no other influences. So in the right-and facet, when strength is 1, the posterior value of the consequence variable is 1. In the left-hand facet, when the strength is 1, the posterior value of the consequence variable is 0 because the influence variable is 0. On the other hand, when the strength of the influence is zero, in all five facets the value of the consequence variable is not affected.



The next example is the same idea, but the influence variable forces the consequence variable to take the opposite value.



The next two examples are much more interesting: continuous versions of necessary and sufficient conditions. The characteristic upper and lower triangles familiar from empirical presentations of these conditions Goertz(xx) are visible - for NECC, where the influence variable is 0 and for SUFF where it is 1. When the strength is less than 1, the influence variable does not completely determine the consequence variable. When it does not, the information about what value the consequence variable should nevertheless take does not come from arbitrary noise but *from the prior value of the consequence variable*. The influence variable serves to restrict the range of operation of this prior information.



Where does this prior value come from? Hopefully we do have some prior information, or at least a probability distribution. In the simplest case, nowhere. If we have no information, we can simply provide a “flat prior”, i.e. a distribution in which each level has the same probability. The app can cope with a discontinuous probability distribution with up to five levels, but doing Soft Arithmetic with many exogenous variables with unknown levels is computationally expensive. If there are three of them, there are  $5^3$  sets of calculations to do.

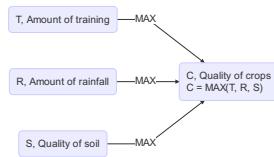
“Wait” you say: “what is the difference here between the influence variable having a value of .5 and the influence itself having a strength of .5”? Imagine we knew that a substantial income is necessary for admission to a particular profession. You just can’t enter it without that income – but in particular cases, someone’s income is not so high. So the value of the variable is middling, although the strength of the influence is 1. Alternatively, we might suspect that a substantial income is kinda necessary for admission, but the extent to which it is really necessary is in doubt. We could say the strength of the relationship is only middling, even for people with top incomes. (FCMs have a lot of cases like this.)

We could also include functions which change the contour of the incoming variables, such as squaring. But not now.

### 36.5 Some “multi functions”: functions for packages with more than one influencing variable:

These kinds of ideas can be generalised to packages of multiple influence variables. The most obvious functions are these:

- MIN
- MAX
- MULTIPLY



MIN and MULTIPLY are continuous analogues of Boolean AND. The diagram above suggests that crop quality is only as good as the weakest of the three influence variables. So if rainfall is poor, there is no point trying to improve soil or increase training. MULTIPLY embodies the same idea but is more aggressive.

It could also be useful to have measures of central tendency and dispersion, like

- MEAN
- STANDARD\_DEVIATION

.. which can be defined identically to their interval-scale counterparts.

It is important to note that all these functions are applied to a package (even if the package contains only one variable) at once; they are a property of the whole package, not the individual variables. For this reason it is a bit misleading to display the information on the arrows.

#### 36.5.1 SOFTADD

One additional multi-function is called SOFTADD. I think this may be the most frequent case. It fills a gap for an incremental, addition-like function for lo/hi variables. Every additional positive parameter to the function potentially adds at least something to the result. So if someone says just “the training helps increase crop yield, but so does the pre-existing skill level and of course the weather”, SOFTADD is the best option. If any of these influence variables have values equal to 1, the consequence variable will also be 1, but if they are

all below 1, the consequence variable will never reach 1, but will get ever closer to it.

Suppose:

- T, Amount of training = 0.3
- R, Amount of rainfall = 0.4
- S, Quality of soil = .5
- C, Quality of crops = SOFTADD(T, R, S)

SOFTADD is calculated by taking one of the parameters, then adding to it the second parameter \* the gap between the current total and 1, and so on until all the parameters are exhausted. Perhaps surprisingly it is commutative, the order doesn't matter. So in the example,

$$C = .3 +$$

$$(1-.3)*.4 +$$

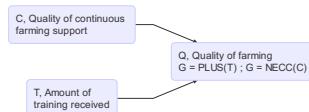
$$(1-(.3 + (1-.3)*.4))*.5$$

[Also, we could add asymmetric functions.]

### 36.6 “Package-free” combinations

Now life starts to get interesting. What happens if we combine two separate “package-free” pieces of information about the same consequence variable which do *not* come from the same source and where we have no information about how to combine them? This is a really common situation which happens hundreds of times say in a QuIP study.

These could involve the same or different functions.



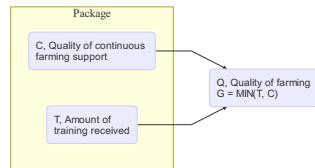
I already set out why I think this is an extremely tricky problem.

Let's go back to contrast it with this subtly but importantly different case, when we get the information about the influence of C and T together, as part of the same explicit causal package, presumably from the same source at the same time. Someone tells us “Ah, the quality of the farming depends on both the training and the support: it is only as good as the weaker of the two”. Here the source has explicitly combined the information into a package. To make

**SECTION 36. CODING SPECIFIC INFLUENCES – INDIVIDUALLY AND IN COMBINATION**

141

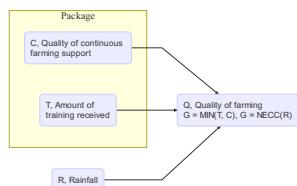
that clearer, I've added a box below, though I won't do this every time we have a causal package.



Except with the special strong assumptions of interval-valued variables under linear combinations, there is probably no default way to just jam together several different such causal fragments which are not part of the same package. For example because a function like “NECC” arguably has an opinion about how it should be combined with other functions, which is maybe different from the opinion which a function like PLUS has about how *it* should be combined.

If we do have information, from the context, about how different influences are to be combined, with some explicit function, we should explicitly join together all the influences on each variable as an explicit package with that explicit function. the problem I am focusing on here is “package-free” combinations for which we do not have any such information.

Finally, it is perfectly possible to have a causal *package* jammed together with another package which consists of a single influence. Here we have information about how to combine C and T but not about how to combine the influence of that package with the influence of R.



In the rest of this section, below, I make a stab at a generic way to combine any two or more such (packages of) influence variables when we have no information about how to combine them.

The jamming together of these different causal influences happens via an R procedure which basically takes the weighted average of the different influences on the consequence variable. (It takes the value which the consequence

variable would have under the influence each causal package, then calculates the difference between the prior value and this posterior value. This set of differences, one for each causal package, is then multiplied by the strength of each influence; so if the strength is 0, the resulting difference is zero. Finally, their arithmetical mean is taken and this is added to the original prior value of the consequence variable.) This is really easy maths, but the ramifications can be quite confusing.

It's also worth noting that this weighted average also works well with any kind of numerical variables, not just lo/hi variables.

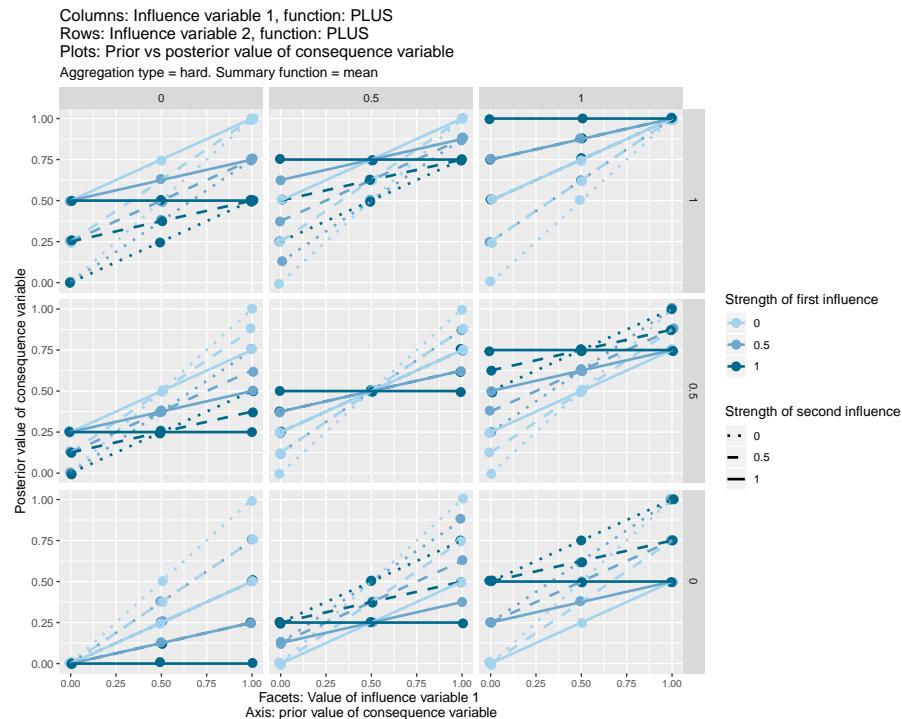
This “package-free” aggregation function is shown in the set of graphs below.

Each of the nine individual plots show how the posterior, i.e. final, value of the consequence variable on the y axis varies according its prior value, i.e. the value which it would have taken before these two influence variables touched it.

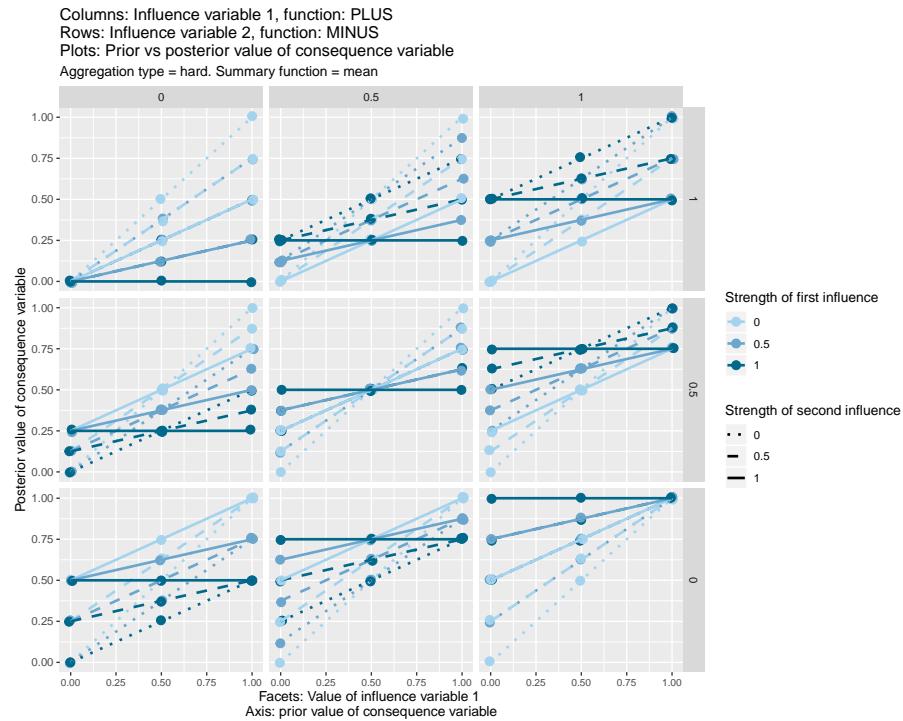
The three different vertical sets of facets or sub-graphs show the three values of the *first* influence variable and the three different horizontal sets of facets or sub-graphs show the value of a *second* influence variable.

The strength from 0 to 1 of the first influence variable is shown by the colour of the lines. The strength from 0 to 1 of the second influence variable is shown by the line type.

The graphs also show some additional options (`type`and `summary` function) which I have not yet discussed.

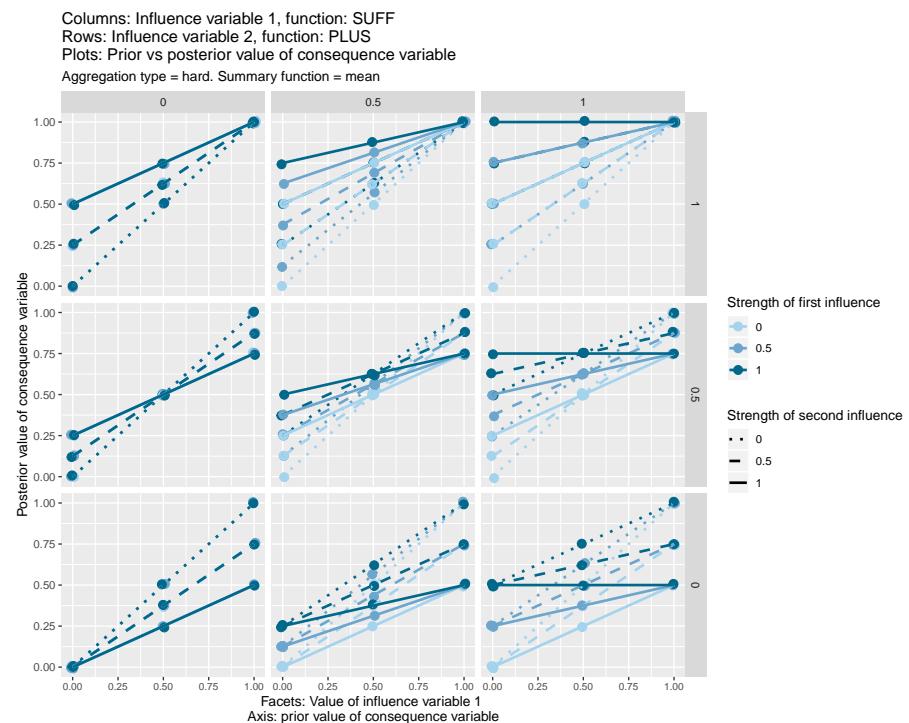
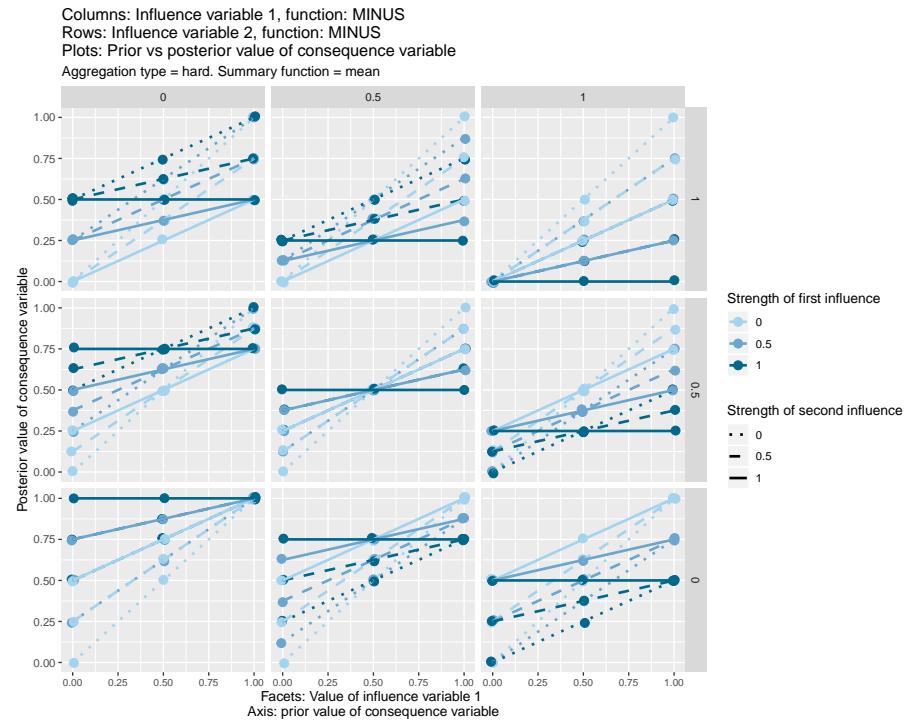


In these graphs it is helpful to look straight for the dark-blue, unbroken line; this is the line where both influences have a strength of 1. So in this first diagram, this line is always horizontal, which means the consequence variable is fixed to a specific value depending on the facet, i.e. depending on (the mean of) the values of the influence variables. This first diagram is in a way the most problematic, because it highlights the weakness of the “weighted mean” approach: if two forces are acting in the same direction on the consequence variable, we would probably expect it to be affected by their sum, not their mean. See xx.



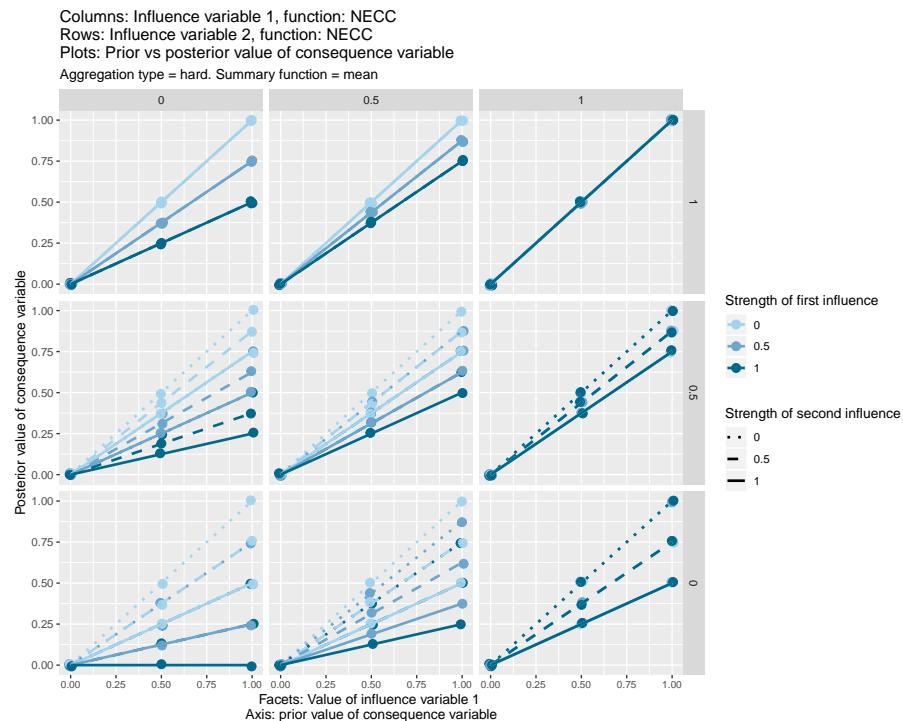
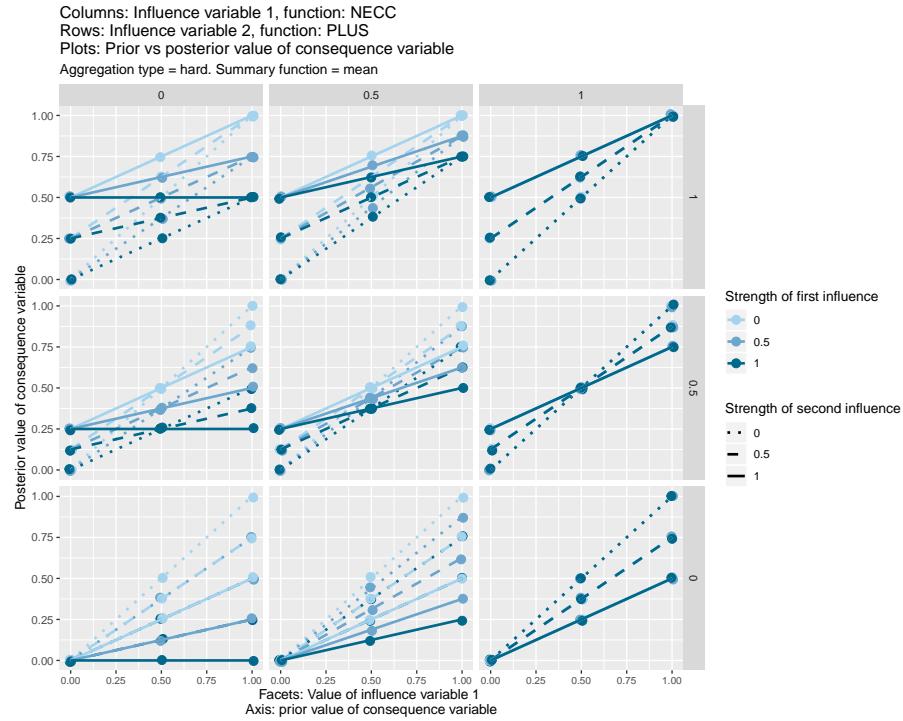
**SECTION 36. CODING SPECIFIC INFLUENCES – INDIVIDUALLY AND IN COMBINATION**

144



**SECTION 36. CODING SPECIFIC INFLUENCES – INDIVIDUALLY AND IN COMBINATION**

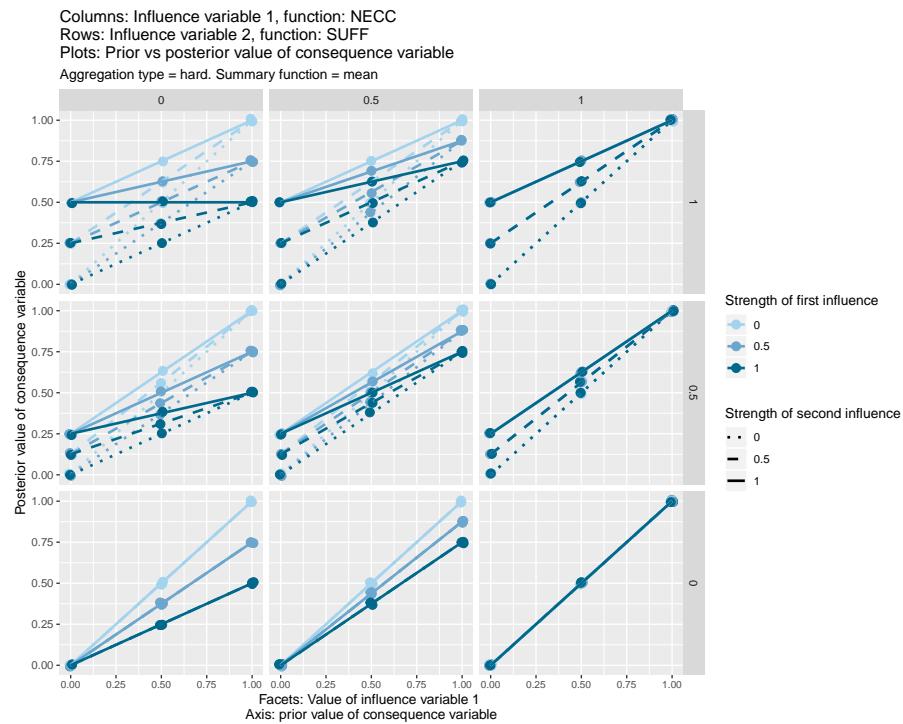
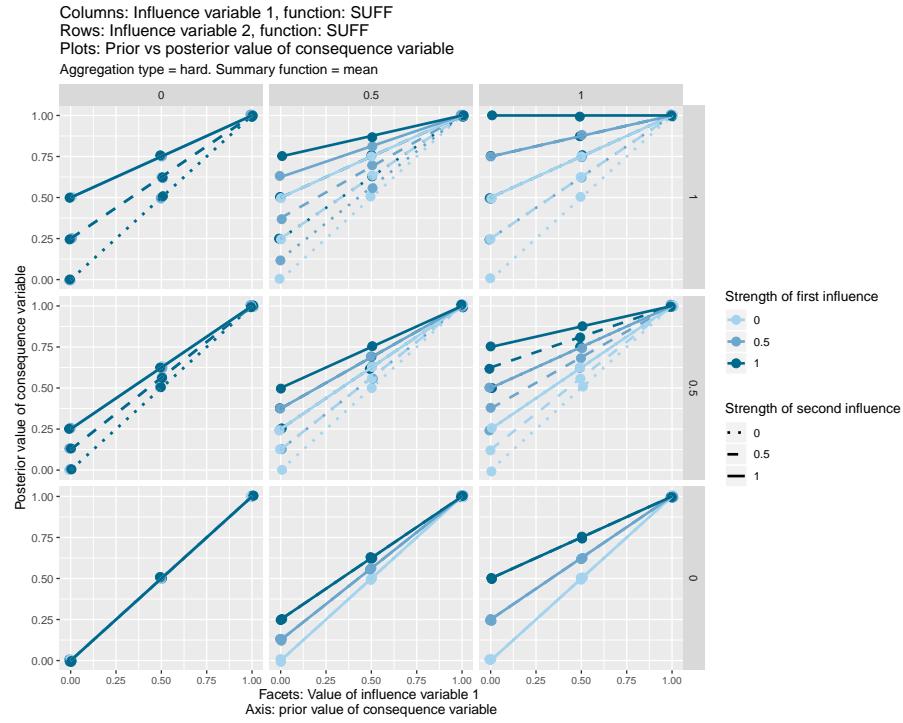
145



When both conditions are necessary, and both take a value of 1 (top-right facet), they have strictly no influence on the consequence variable, and all the lines are  $y=x$  lines, i.e. the consequence variable just retains its previous value.

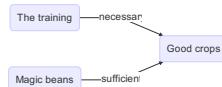
**SECTION 36. CODING SPECIFIC INFLUENCES – INDIVIDUALLY AND IN COMBINATION**

146



### 36.7 Contradictions?

In general, it is not coherent for *the same source* to claim different functional influences from the same or different sets of influence variables *onto the same consequence variable*. This might be surprising. But for example let's think what it would mean if someone claimed simultaneously a necessary and a sufficient condition on the same thing.



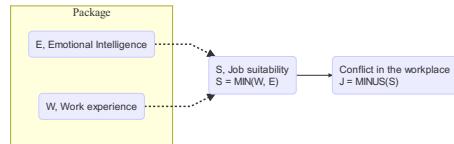
..... they couldn't, could they? If the beans are sufficient, you don't need to know anything about the training, so it can't be necessary. If the training is necessary, the beans can't be sufficient (because if there was no training, the beans wouldn't be enough).

This case is shown in one of the graphs above; the result is that the two influence variables transmit no effect to the consequence variable, which remains at its prior value.

So we expect sources to organise their information into packages, and if they don't, we'll do it for them.

### 36.8 Building up more functions from these building blocks

As we piece together fragments of causal information into a larger network, following these causal rules (or getting the app to do that for us), we will see how downstream variables are in principle influenced by the values of the upstream variables. But we can also “virtually” piece together more complicated functions, if we need them, by using conceptual links rather than causal links. For example, here we define a conceptual or latent variable “Job suitability” – note the dotted lines – as the minimum of emotional intelligence and work experience; and we show how this inversely affects the level of conflict in the workplace. By “zooming out” we can see that  $J = \text{MINUS}(\text{MIN}(E,W))$ ; we have constructed a more sophisticated function by piecing together two from our handful of basic functions.



### 36.9 Coding influences within in the app

The app already realises the coded networks as actual R objects, within which the variables or nodes have an attribute called `level` e.g. 0 or 1, but they can take other values too. Also there is an attribute called `fun` which stores the function which the variable uses to calculate its own level on the basis of the levels of the variables directly upstream of it. (As there might be more than one package of influence variables, the `fun` attribute is stored with each such package.)

Our job when coding the causal link is to specify this function using the drop-down widgets etc in the app's UI (and if there is more than one combination of influencing variables, to code one function for each set). We should also code the level of the variable, if we know it.

In general, the variables are typeless so you can give any variable a level of 1, or -2090, or “fish” or anything you want, and the function attached to the variable(s) directly downstream of it can maybe make sense of that, or maybe not. The function can be the name of any R function, including those mentioned here. But the philosophy of the app is focused on lo/hi variables.

## Section 37

# One function to rule them all?

The preceding sections have been quite involved. Can't we keep it simple? Can't we just code every causal claim in the same way, e.g. as somehow some kind of positive influence, and leave it at that? That might save us a lot of time when coding especially when we don't really have the necessary information, and it would save us thinking about all the philosophy?

By all means.

### 37.1 One universal type of function for coding causal claims with exclusively lo/hi variables.

There are some obvious candidate functions. Coding a connection between (packet of) influence variables and consequence variables could be consistently understood as any one of these:

- Total control
- Bare control
- Necessary
- Sufficient
- Somehow-positive control

Probably only the last makes sense in our context. It's amazing though that there are whole schools of social science built on the exclusive use of sufficient and even necessary (Dul, xx) conditions. However if we really want to use one universal function for everything, we have to remind ourselves that there is no going back – we couldn't decide to use just, say, sufficient conditions and then change our minds and ask to code one specific causal link as, say, expressing a necessary connection.

Indeed, there are such strategies, here are two:

The purest and least sophisticated strategy to encode causal information, “**content-neutral coding**”, does no coding of the causal *content* of the link at all. We interpret an arrow as meaning only

C has *some* kind of causal influence on E, but we aren't saying, or don't know, what

The next step would be a strategy like “**purely-positive coding**” in which we do provide some minimal encoding of the content, but always as some kind of “positive” influence, like this:

C has some kind of, in some sense, positive or increasing causal influence on E, but we aren't saying, or don't know, more than that.

But we should remember that, in contrast, the (raw) information provided by our respondents can may in fact have arbitrary causal content. So they might actually say things like this:

C is almost irrelevant to E

or

C drastically reduces E

or

C is necessary for E

or

C definitely has no causal influence on E

as well as

C has a strong and positive influence on E

etc.

We would struggle to meaningfully encode most of these kinds of information using either “purely-positive” or “content-neutral” coding.

By all means we could use “content-neutral coding” to encode willy-nilly the example statements just given as arrows from C to E (or whatever). But we wouldn't be able to do much with the resulting network. We would hardly be justified even in *aggregating arrows* (combining several cases in which several sources mentioned the same arrow, for example, (in the resulting network) using the width of the combined arrows to show the number of mentions). That would be to mix apples and pears.

We will have to try be at least a bit more sophisticated in how we encode causal information as arrows in a causal map.

## **Section 38**

### **Coding a claim about the absence of a causal link**

Is it really the same as strength=0. I think not.

## Section 39

# Strength / importance

### 39.1 Including the idea of strength in our functional equation

Usually we think of functions as having the form

$$E = f(V)$$

where V is some set of influencing variables e.g. B, C etc, possibly just a set of one variable like B. Lately we have been using an extended form:

$$E_{posterior} = f(E_{prior}, V)$$

In order to include the important idea that influences rarely have the power to completely shift the consequence variable from 0 to 1, we have already allowed for functions of the form

$$E_{posterior} = f(E_{prior}, V, s)$$

where  $E_{prior}$  is the prior level of E, the level without the influence of V, and where **s is the strength of the influence**, between 0 and 1.

There are different possible versions of the rules which govern the way strength information is processed, (and which therefore determine exactly what we mean by it). For example how much the prior value of the consequence variable is taken into consideration.

So really this means revisiting and expanding our basic coding rule to include coding for strength.

“Strength” uses a generic way of *tempering* or *weakening* causal information, in particular we use it when the respondent themselves says that the connection is weak.

We can imagine applying this kind of “tempering” transformation for a variety of other purposes, for example because the causal link has only a small amount of evidence, or we don’t trust the respondent. I have added another attribute which tempers the effect of a variable in the same way. It is actually called

“trust” but it works at the moment as a placeholder for any of a range of different qualifiers — trust/trustworthiness, confidence, explicitness (Gary/James version 6) etc. These issues are of course not the same and should (probably) not be encoded with a single attribute.

If “trust” is 100%, we treat the information as gospel. If it is zero, it has no influence on our causal map at all.

- From: “source P believes / asserts that ‘A causes B’ and we trust source P 100%” we *can* (by definition) conclude that A causes B
- and, from “source P believes / asserts that ‘A causes B’ and we trust source P 0%” we have no information about whether or how much A causes B

But with values in between 0 and 1, the equation is a bit more difficult to formulate.

Basically, from a Bayesian perspective,

The strength of the upgrade to our information due to a report with a standard tempering of .3 would be correspondingly reduced, i.e. multiplied by a factor of .3.

### **39.2 Direct ways of eliciting how important is one variable’s contribution to another. (Fiona)**

Where the coding of the strength of an influence is particularly important, we might want to specifically ask our respondents questions about it. The coder should consider this information when coding the “strength” attribute of the package of influencing variables.

## Section 40

# Causal Inference

What kinds of questions can we answer with causal inference? These:

- can diagram X be deduced / synthesised from (a more complicated) diagram Y?
- show how having even a simple feedback loop makes causal inference often unpredictable
- what is the overall causal power or “contribution” or effect size which one variable has on any descendent?
- which information is unimportant?

... and more.

### 40.1 Specifying the map

#### 40.1.1 Exogenous variables and the Grid

Ideally the coder would also specify the levels of all the endogenous or “no-parent” variables, the ones at the beginning of the causal chains. In practice, much of this information isn’t known in any detail.

If one or more of the influencing variables do not have a level specified for them, the app opens up a “Bayesian grid of scenarios” and calculates the consequences for several different levels, usually 3, of the exogenous variable in question.

So if there are 5 exogenous variables, this means there will be  $3*3*3*3*3 = 35$  combinations to work out. Non-parametric reasoning is computationally expensive.

Anyway, we can’t display all of those possibilities to the user, so we have to ultimately resort to statistics like the mean of the values for the variables across all the possibilities.

#### 40.1.2 Inference in action

When you’ve specified your diagram, the app can calculate the consequences on the downstream variables.

By default, the app then calculates the mean of all the scenarios in the Grid.

At the moment, the app can cope with a network which included feedback loops, but it does not actually follow the causal path right around the loop but stops before repeating itself.

## Section 41

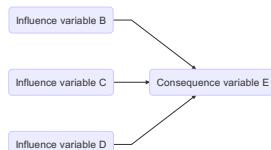
# Upstream (Bayesian) inference

Is it enough to process information *down* the network? No. True, cause can by definition only flow down it. But we might have information which implies something about antecedents, which means we have to reason *backwards*. For example, we might know that the student failed to fulfill one of two conditions for entering a college, and we know they entered it, so we know they must have fulfilled the other condition.

When we have coded our causal arrows following the rules set out earlier, we can use this information to calculate specificity and sensitivity, so we can use Bayes' rule to modify our prior information about the upstream variable(s).

Not implemented, but soon.

This way it will be possible to do Process Tracing with causal maps Tansey (2009),Befani and Stedman-Bryce (2017),Law (2005)



## Section 42

# Contour

Influences don't have to be straight. It is possible to select different contours.

- quick-start: (`contour=quick-start`)
- slow-start: (`contour=slow-start`)
- S-shaped: (`contour=S-shaped`)
- U-shaped: (`contour=U-shaped`)
- threshold: (`contour=threshold`)
- straight: (`contour=straight`) (the default. I don't say "linear" as that can mean other things too)

### 42.1 Flipping

Also, each consequence variable can be "flipped" in relation to the combination of its incoming variables. All of these combinations work with any number of incoming variables greater than 1. Flipping is the same as adding a "MINUS" variable in between.

## Section 43

### Valence

Hearts and wedges: Marking which variables are deemed to be *important* / *valuable* (“heart” symbol?); and which can be *intervened* on (“wedge” symbol?) In theory, this information can be a way to model or understand or predict someone’s behaviour (“they’re going to want to intervene on X in order to get Y”). For this to work we need to also encode the fact that they believe that X leads to Y: a causal map about somebody’ causal map. We can also generalise this to deal with *multiple* actors each of whom can value and intervene upon different (sets of) variables.

I’ve been having a battle with myself about metalanguage here. The problem is with the word “value” which is the best word to talk about the state or level of a variable but also the best word to talk about the valence which something has for us, how good it is.

## **Section 44**

### **Valence and direction**

Need to discuss the way ordinary language confuses the valence of something (I like it) with the polarity of the influence(s) affecting it.

“Plus” vs “minus” aka “positive” vs “negative” influences

Also need to discuss the fact that the psychological implications of “flipping” a variable are not equivalent to the logical implications.

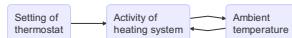
E.g. coding a variable as “High levels of heat” vs “Low levels of cold” etc.

See also Mahoney and Goertz (2004)

## Section 45

### Effect: two kinds

In a stable universe we can calculate the effect of a variable by tweaking it and describing its projection on some variable of interest. So we can see how twiddling the thermostat setting changes the behaviour of the system, dependent also on ambient temperature.



But we can also describe the effect of *removing the variable from the system*. For example, what happens to the rest of the household when the thermostat is removed?

## Section 46

# Interventions and differences

You can specify “base” and “intervention” values for (some of the) exogenous variables, in which case the app can calculate the “effect” of an intervention on a downstream variable, defined as the “difference” between the predicted value under intervention and the predicted base value, i.e. the value without intervention.

### 46.1 Shortcuts

Type (.9--.3) after a variable name to designate a variable as `intervention=.9; base=.3`.

Type (--.3) to designate a variable as `base=.3`.

So this:

```
Variable 1 (1--0)
Variable 2 ( --0)
```

..... means that we intervene to set Variable 1 to 100% rather than 0%, and that variable 2 just has a base level of 0, we don't intervene on it.

## Section 47

# Contribution

Can we tell from our causal map what the contribution is of a certain intervention on a variable downstream of it?

Conversely, can we take a specific Difference on a downstream variable of interest and ask, what is the “contribution” of an upstream variable to that Difference?

It looks so easy, especially when we have a causal map to hand. In fact, as Pearl insists, we *have* to have a causal map to hand when we try to soft-calculate the contribution which C makes to E.

We have soft-specified the strength of the links and the values of the other exogenous variables, what could be easier than pressing a button and soft-calculating the contribution of the intervention? i.e. the value of outcome O is much higher in the scenario where the project is run than in the scenario where it is not run.

Information about contribution is probably number 1 on our client’s list when we are helping them construct a causal map.

But ...

There are a lot of extreme difficulties outside the linear, quantitative comfort zone...

### 47.1 Why reporting correlation is not helpful

When calculating the downstream effect of an upstream variable in terms of the Difference made by it, we should be careful to avoid shortcuts like looking at a correlation coefficient. Correlation is a poor way to spot non-linear effects and in particular is bad at recognising necessary and sufficient conditions.

It’s worth noting that the formulation of the downstream effect of an upstream variable in terms of the Difference made by it is immune to this problem.

When examining the overall effect of a continuous variable on another, Differences are not enough. Need to have some kind of display.

[add]

## Section 48

# Information about the *source* of our causal information

Above, we looked at how we can encode ‘C causes E’. Not at how we can encode “P believes that C causes E”.

### 48.1 Applying the “trust” attribute source-by-source

In section xx we also looked at qualifying causal information according to its quality, which might be less than 100%.

We might consider extending this idea so that individual sources can be given individual and usually different trust attribute, so that an expert source might get a value of say .7 and a community rumour a value of say .1.

The strength of the upgrade to our information due to a report from a source with a standard quality qualifier of .3 would be correspondingly reduced, i.e. multiplied by a factor of .3.

### 48.2 Including information about sources

A more fundamental move is to include information about sources as a meaningful component in our causal maps with a role in our Soft Arithmetic for causal maps.

see later.

## Section 49

### Clusters of similar maps

Markiczy and Goldberg (1995) suggest a method for summarising a set of causal maps, one each from a set of respondents, which does not rely on piecing them together but on finding typical clusters of maps, and then presenting the central map from each cluster along with the typical respondent characteristics.

The way they suggest of clustering graphs depends on the concept of distance between graphs. There are various methods Gao et al. (2010) to measure the distance between (pairs of) graphs, some making use of the idea of the maximum common subgraph. The most common of these methods (like Levenshtein distance) do not make use of node or edge attributes – information like the direction, polarity or strength of the influence between nodes.

Markiczy and Goldberg (1995) also point out that also here we need to ask ourselves if the absence of a variable or an arrow is relevant or not.

For example, is the graph

$B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \rightarrow G \rightarrow H \rightarrow I$

more similar to

$B \rightarrow C$

or to

$B \rightarrow C \rightarrow D \rightarrow F \rightarrow H \rightarrow I$

On the one hand, the second is a perfect subgraph of the first – it has, in a sense, no mistakes; on the other hand, the third has three mini-maps in common with the first (but also some incompatible parts).

It probably makes sense to penalise the second more than the third, because if we were to say that 1 and 2 were close, we would produce results with very small graphs, and we can already see which are the most popular mini-maps just by looking at the frequency of arrows.

I haven't yet implemented this and am looking for a suitable R package.

## **Section 50**

### **Maps of maps**

Very often some actor's own internal causal map is a relevant factor in our causal map of the whole field. For example, in behaviour change interventions.

We need to be able to encode not only the fact that

C influences E

but also

stakeholder Q believes that C leads to E

and maybe

Q values E

(and/or Q values C, see Powell(xx)).

## Section 51

# Probability density functions

I just realised that it's easy to solve the problem of paradoxical amounts of certainty about the strength of causal links.

The problem was this:

To calculate the strength of a link when combining information from different sources about its strength, we'd usually take some kind of mean or average. Here are two cases when people tell us, with an ordinary amount of certainty, about the strength of a causal link between B and E:

- ten people say it is 0, and another ten tell us it is 1, the average would be .5.
- twenty people tell us it is .5, so the average is also .5

It would be pretty bad if one can't distinguish between these two cases. In the first case, we'd want there to be a lot of uncertainty (almost no useful information) when modelling the influence of B on E.

The standard way to deal with this is to store information about link strengths as probability density functions rather than single numbers, and it turns out this is really easy to implement.

So what difference does this make to the rules? When doing things like calculating the aggregated strength of the link and/or calculating how influences flow down the causal chain, the app doesn't use individual numbers like .5 but actually probability density functions (PDFs): uncertainty clouds. So the app can still report the averages (which would of course be the same as before) but can also display information about the cloud of certainty / uncertainty if asked. More importantly, if there is a lot of uncertainty about a link, there is a lot of uncertainty about the way causal impact flows down it.

Goertz & Copestake (xx) suggest including “certainty” of a link in the coding form itself, ... but the “paradox” above is more about the certainty we have in aggregated links rather than about coding uncertainty directly.

But sure, if the strength of aggregated links can have uncertainty, directly coded links can have them too. Two ways this could be implemented:

- Just include a slider or text box for reported “certainty”, so the coder could just code say .5 for the strength, as well as a certainty of say .8, which would usually create a small cloud of other possibilities on each side of .5 (the coder wouldn’t need to see anything happening, but the information would be recorded)
- Or, (I think this is better for most cases) there doesn’t need to be *any* change to the coding interface; everything just gets coded with a medium amount of uncertainty.

Actually, allowing PDFs for strength necessitates allowing PDFs for the values of variables too. This doesn’t necessarily imply any further difference to the coding form / inference rules / app, but it is available if required. I’ve already allowed for the possibility of coding the actual values of variables, so that when coding something like “The quality of the training course is important for the level of farmer skills, but the quality of the course was quite low”, it is possible to code the information about the quality as well as the causal link. Now with PDFs it would be possible to also code the certainty of that information if required (though I doubt this is necessary), and more interestingly it would be possible to report the uncertainty of that information, as well as the mean, when aggregated across sources.

## **Part III**

# **Simplification**

## Section 52

# Simplifying causal maps: aggregation and filtering

The best answer to most causal questions about a causal map is another causal map: a simpler one which can be derived from the first.

So you have coded multiple fragments of causal information. The app *can* put them together into a combined network with no aggregation at all – one line for each causal link etc – but usually this will need simplifying.

Only a very few stakeholders will actually want or be able to explore all of this information: a great central product would be a one-page graphic which can be understood with a couple of minute's viewing, perhaps with appendices showing additional information. And even those who have the time and interest will want to start at a certain level of aggregation.

In the QuIP protocol, only one code is made per respondent per domain or question, even if a link is mentioned several times. This can already be considered a kind of summary.

### 52.1 Why simplify a causal map?

To count as knowledge, it has to fit in your head.

A presentation of a causal map should fit on a page of A4.

It should be simple enough that the user, with a bit of work, can get a *causal gestalt* – this might even be likened to muscle memory, a physical feeling for how things fit together, what you have to wiggle to make something else jiggle.

Data-led simplification makes use of different metrics which we can define for the variables, arrows etc in the map. The following sections look at different available metrics.

---

How are we going to aggregate a network?

## 52.2 The tables available

The app starts its aggregation looking at these three tables:

### 52.2.1 Table of Statements

This has not only the texts of the actual statements but also optionally user-defined attributes like source ID, gender, district, question number, whatever columns were in the data which was imported at the beginning of the project. Also, if a column has a special name like `Age.District.Sex` and contains values like `22.North.M` it will be split into separate attributes, say, `Age = 22`, `District = North`, `Sex = M`.

### 52.2.2 Table of Variables

This table has label and ID, level, cluster, value (in the sense of valence/importance/valued-ness, not level), fun (see below) and any user-defined attributes.

### 52.2.3 Table of Arrows

This table has label and ID, the full text of the statement upon which it is based, any more specific quote, level, cluster, value (in the sense of valence/importance/valued-ness, not level), fun (see below) and any user-defined attributes.

At the moment, the app now carries out the following steps, in this order. I'd thought that this process would have to be completely user-defined, but in fact at least part of this ordering is inevitable. So the order of the steps is hard-coded but the user can switch the steps on or off, and tweak them as described by fiddling with the settings in the `Display` tab in the app.

## 52.3 The steps taken by the app when simplifying a network according to user commands and/or automatically if requested.

### 52.3.1 Inference

(Not strictly part of the simplification process but mentioned here to keep its place in the chain of calculations.)

If `infer` is set to `T` or `TRUE`: Calculate the levels of downstream variables given the levels of upstream variables. This means that for each variable, the app has to know how to combine the upstream information. That is what is stored in the `fun` (i.e., "function")attribute of each variable, e.g. `AND` and `OR`. The `fun` attribute can be set manually in the Arrows table or in some cases by using the "Interaction" widget in the Coding tab.

### 52.3.2 Calculation of source-by-source data

(Not strictly part of the simplification process but mentioned here to keep its place in the chain of calculations.)

Here we can calculate metrics like “conspicuous absence”, see xx.

### 52.3.3 Merging together variables which are in the same cluster

I think clustering is the most important way in which the researcher simplifies a causal map, and the app needs to help them with it.

*If variablemerge is set to TRUE:* Any variables which have the same letter or string (“A”, “4”, “Cluster1” or whatever) in the Cluster column are merged into one. Any arrows to the component variables are re-routed to this new cluster variable, TODO: how are variable attributes like valence aggregated?.

By default, the variables in the cluster are still stored separately and are combined “live” each time the visualisation is refreshed. But there is also a button to combine them permanently, i.e. the information about the separate variables is lost for ever (though the step taken should be stored in the project log TODO).

(TODO: take into account when links between variables have been specified as “definitional” or “conceptual”; these are good candidates for clustering)

(TODO: finalise algorithm which auto-suggests sensible clustering)

(TODO: ensure that clusters appear in the automatic filters, see below)

### 52.3.4 Merge information about Statements into the arrows which are based on them

This happens automatically. So an arrow “knows” all the attribute information from the Statements tab which belongs to this statement, e.g. Source ID.

### 52.3.5 Filter sources

If there are any such attributes in the Statements tab e.g. Source ID, gender, then filter buttons appear in the Display tab which allow the user to exclude specific categories e.g. women, people from District X, etc. The app then excludes all arrows which are encoded from any sources with that category. This functionality is intended for live exploration of the diagram and the settings of the filter buttons is not stored when the project is saved.

### 52.3.6 Hard-coded calculations

Hard-coded calculations for the arrows combining the attributes. At the moment, only `wtrust = strength weighted by trust`. (TODO: allow user-defined calculations)

### 52.3.7 Merge arrows

If `arrowmerge` is set to `TRUE`: All arrows connecting the same pair of variables are merged into one arrow. The attribute `frequency` i.e. count of the number of arrows is added automatically to the new combined arrow. Other attributes (strength, trust, sensitivity, source, label, etc) are aggregated too (at the moment only sum and mean are calculated, but TODO the app should calculate live any aggregations required by the user for the visualisation e.g. max, min, user-defined functions etc).

### 52.3.8 Calculate QuIP-style metrics, e.g. summaries per domain.

### 52.3.9 Filter out arrows below a minimum frequency

If `arrowminimumfrequency` is larger than 0, any arrows with frequency below this threshold will be excluded from now on.

### 52.3.10 Add arrow attributes to variables

If `variablejoinededges` is `TRUE`: The attributes of arrows to and from a variable are added to it. So for example the variable now “knows” how many women mentioned it as the target of a causal claim. This is an important feature of the app and really useful for the final visualisation. This involves two stages of aggregation: first functions like e.g. mean and sum are used to aggregate the different arrows coming into (`to`) or going out from (`from`) a given variable, and secondly a function is used to aggregate the `from` and `to` information. This means that at this stage, a variable can potentially have information like `mean_age_mean_mean`, i.e., working backwards, take the mean of the arrows merged in step 5 above, then the mean of arrows from several variables linking from or to this one, then the mean of this `from` and `to` information. Also available are a host of attributes like `from_age_mean_sum` and so on. It’s quite bewildering!

### 52.3.11 Filter out variables below a minimum frequency

If `variableminimumfrequency` is larger than 0, any variables with frequency below this threshold will be excluded.

Now the aggregation has been completed; the next step is how to use it in the visualisation.

## Section 53

# Aggregating and filtering beliefs

- How to combine multiple such fragments? (this is QuIP bread-and-butter)
  - e.g. if source P believes A (completely) causes B and that B (completely) causes C, do we know that they believe that also A (completely) causes C? It's a logical consequence, but how rational can we assume our sources to be?
  - (Fiona's problem) e.g. if source **P** believes that A (completely) causes B and source **Q** believes that B (completely) causes C, can we conclude anything about A and C? In other words, if one source tells me about one arrow in a chain, and another tells me about another which links with it, who told us about the whole chain?
  - (Generalisation problem) e.g. if source P believes A (completely) causes B, can we conclude anything about what some larger population believes?
- How to *summarise* multiple such fragments?
  - In particular:
    - \* How do we combine different overlapping statements in which we have different degrees of trust? See xx

## **Section 54**

### **Aggregation and filtering based on face value**

You can remove items by hand because they are not relevant to your research interest. For example remove items which are not upstream of important outcomes, or downstream of important inputs like interventions.

Or ...

## **Section 55**

### **Aggregation and filtering based on particular research questions**

Often you might want to prepare a different view of your causal map in order to answer a specific research question.

## **Section 56**

# **Aggregation and filtering based on metrics**

### **56.1 Network metrics**

[add more]

## **Section 57**

### **Citation intensity**

In QuIP, Citation Intensity is the citation count divided by the maximum possible citation count (no. Respondents x no. Domains).

## Section 58

# Conspicuous absence

This metric can be calculated on the basis only of structural information, with no knowledge of the contents or nature of the causal links.

The basic idea is to look at the *lack* of edges: node pairs (aka variable pairs) which could have an edge (aka arrow) but don't, or have only a thin edge, i.e. with very few mentions. The made-up example below is supposed to be built up on the basis of edge information derived from say 15 separate respondents. Is the low number of mentions of the arrow between "better crops including feedstuffs" and "healthier pigs" (below) understandable because they were anyway rarely mentioned together by the different respondents: maybe respondents tended to talk about the pigs, or the crops, but not both? Or is it *conspicuous by its absence*? We can look at the 15 individual "mini-maps" produced by the 15 respondents. Did those two nodes in fact appear together in the many of the respondents' mini-maps but without an edge between them? This is something that we can *not* tell just by looking at the map, but a confusion-matrix-type structural analysis *can* tell us.



One could build a global score for each edge, to complement the score for the frequency of mentions: so, 50% = this edge is present in 50% of the mini-maps in which both nodes were mentioned in some way or other. So if it turned out that "better crops" and "healthier pigs" appeared together in only two mini-maps, it would get a score of 100% and we would take it more seriously than if it had a score of say 10%. (We could also go on to look at all the arrows

which are not in the map and talk about *how much* they are lacking - a kind of negative causal link.)

The algorithm calculates this score by calculating the individual respondent-level adjacency matrices for the edges mentioned by each respondent, but scoring for *absence* rather than *presence* of edges. So each adjacency matrix only has a 1 for pairs of nodes which are mentioned by the respondent but for which no edge was mentioned. Then combine them into one large adjacency matrix of “anti-causality”. The trouble is that I am not sure about the details — not sure if the score should really be punishing situations when people in group 1 say that X leads to Z, and group 2 say that it leads to Y which is *between* X and Z — that needs some more thought. In this case, you couldn’t really say that arrows from X to Z are conspicuously absent, more that these people are thinking in a more detailed way.)



In the same way, you can look at the total lack of an arrow between Better Communication and Better Wellbeing and ask aha, is this because they were never mentioned together? Or is its absence more conspicuous?

## Section 59

### Not forwards

This metric merely shows whether an arrow in the causal map does not point forwards, i.e. it points backwards or to another variable in the same “column” as it.

Which specific arrow is identified depends on the specifics of the algorithm which calculates the layout (Sugiyama). But the great thing about the metric is it reveals where there are cycles in the network. If you were to delete these arrows, there would be no cycles.

Often, when laid out sensibly, there are fewer loops than people think.

## Section 60

### Homogeneity of paths

(Fiona's suggestion for a metric)

Scenario 1) 20 people might have said that A links to B and another different 20 might have said that B leads to C with no overlap between the groups of people.

Scenario 2) the same 20 people say that A links to B *and* that B leads to C.

Conventional ways to combine the information from the different sources would produce the same diagram in both scenarios. But we want the user to see that there is some kind of weakness in scenario 1. Can't we just show this on the arrows somehow?

In the case of isolated paths with no forks in them, which of course are very rare, this wouldn't present a big problem. Assume that we are showing, as we do, now the total number of mentions for each section represented by for example the width of the arrow.

We can construct a measure which we could call "homogeneity" for each section. So if in a long path, each of the sections were mentioned by more or less the same people the stretch would have high homogeneity and you could for example show it unbroken. In contrast, for an arrow in which separate sections were mentioned by different groups of sources, it would have low homogeneity and you could show that for example with a very broken arrow with lots of gaps in it. Or by making it almost transparent.

However I don't think this works in the much more common case when paths are of course constantly diverging and rejoining.

For example, suppose you have an arrow from B to C and then an arrow from C to X and another from C to Y; suppose there was high homogeneity from B to C to X in the sense that both of these sections were mentioned by many sources but the section from C to Y was mentioned by a different bunch of sources. How would you mark this?

**The problem is that homogeneity as I've described it as a metric of entire paths** and so you can't really show it in individual sections when there are forks, i.e. when one section can have different homogeneities because it is part of different paths.

### 60.1 (non-) solution 1)

What one could do is see whether there is any clustering within sources rather than within variables. So you might find there is a bunch of people who tend to mention many arrows the same and another bunch of people who mention a different set of arrows. It would certainly be possible to automatically or manually create subgroups of respondents. Then there are ways to show how different sections of paths were mentioned by these different subgroups, but it doesn't really answer the question.

### 60.2 Solution 2)

You can certainly report the homogeneity of each individual path in the Report tab of the app, but that would be quite long-winded.

### 60.3 Solution 3)

You could show the same information interactively. For example when you click on B in the example above, you would see the width of the section from C to Y shrink but the section from B to Y would stay the same. However I am not a big fan of information which you can only discover by twiddling.

### 60.4 Solution 4)

You might want to summarise the most important results in Solution 3 e.g. in a legend on the diagram, where you could (automatically) mention individual paths with particularly high or low homogeneity. You could probably develop a metric for the overall homogeneity of a whole diagram, and you might want to mention if a whole diagram had unusually low homogeneity.

### 60.5 Solution 5)

Allow the viewer to “play” the different sources one by one, if there aren’t too many of them.

## **Part IV**

# **Visualising**

## Section 61

# Visualising and formatting causal maps

Now the network has been aggregated, there are decisions to make about:

- encoding important facts about the network e.g. in terms of colour, width, visibility, etc
- decoration - other things which make the network easier to read, such as orientation, gaps between variables, etc.

Here we will deal only with the first.

### 61.1 Conditional formatting

In the **Display** tab, the user can set almost any visual feature of the network to reflect almost any attribute of the variables, arrows etc.

For example:

- Hide all variables with some attribute equal to, or less than a certain number, or containing a certain string ....
- Make the width or transparency of the arrows reflect the number of mentions
- Colour arrows according to the average age of people mentioning them

etc.

This can seem a bit overpowering so the trick will be to have plenty of sensible pre-sets and templates.

### 61.2 Hard-coded formatting

(Very debatable:) e.g.

- make an arrow dotted if it is coded as a definitional/conceptual rather than causal link
- put a symbol on an arrow if its strength is less than 0

### 61.3 Constructing labels and tooltips

Arrow and variable labels and tooltips can be constructed by the user to show e.g.

- the actual label originally provided by the user e.g. “Income due to pig farming”
- any attributes available to the app, e.g. list of sources, average age, etc.

### 61.4 Focus

*If `diagramfocus` is TRUE:* This is mainly for interactive viewing, not for printing / exporting. As the user clicks through the different statements in the corpus, the arrows coded on the basis of this statement are highlighted and the others fade. (There is also a “play” button which plays through all the statements like a video.)

## Section 62

### Coding using the UI (outdated)

- 62.1 Show that the influence of one variable on another has a particular contour.**

Click a button to put a symbol on the middle of the arrow.

- 62.2 Show that variable X is part of the definition of variable Y.**

Select a choice from a drop-down. An arrow is drawn from X to Y but it is dashed.

People have improved health ..... Older people have improved health

- 62.3 Show that variables X and Y are linked by definition**

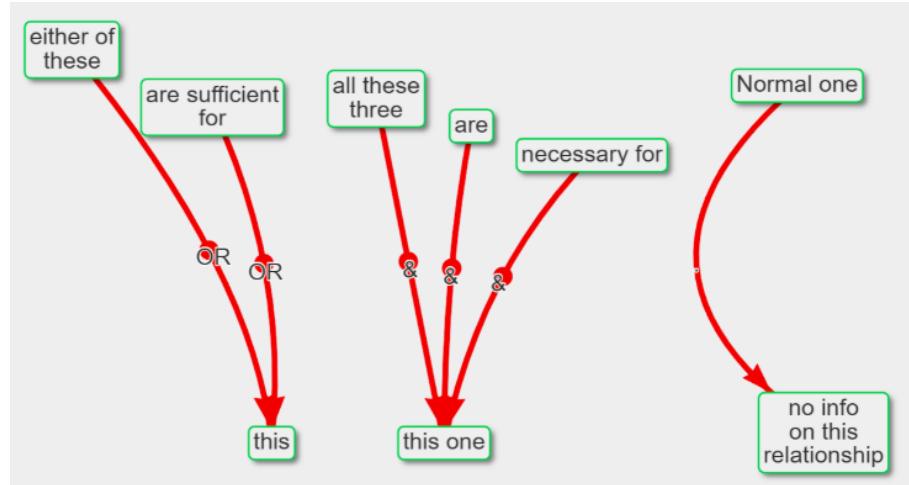
They overlap in meaning.

Families spend more time together ..... Fathers interacting more with their children

Select a choice from a drop-down. An arrow is drawn from X to Y but it is dashed and there is no arrowhead to show direction.

#### 62.4 Show that a set of influence variables influence the consequence variable *collectively*.

Click a button to put them all in the same group. They all get a circle on the middle of the arrow and the desired symbol e.g. & on top of it.



## **Part V**

## **Summary**

## Section 63

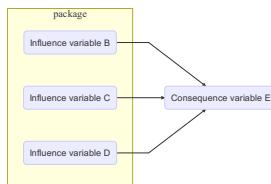
# Summary of the rules for inference in causal maps, aka “Soft Arithmetic”

### 63.1 The inference rules for causal maps

These are rules for reasoning, drawing inferences, with causal maps and/or ordinary, causal, narrative sentences. These rules tell us how to go from one set of maps and/or narratives to another.

### 63.2 The mini-map coding rule

Information like “the influence variables B, C and D all have some kind of causal influence on the consequence variable E” can be coded with a **mini-map** in which one or more variables (the “influence variables”) are shown with arrows leading to another (“the consequence variable”). The information and the map are equivalent.



I call this a “coding rule” but more generally it is the first rule of inference for causal maps. This first rule tells us how to go from a fragment of causal narrative to a diagram (coding) and back again (interpretation).

Mini-maps are the atoms of causal maps. One action of coding something produces one mini-map. You can build up any causal network from them. In 90% of practical applications, a mini-map will just contain a **single** influence variable. But we don’t want to get stuck when we need a **package** of two or more influence variables, e.g. when someone says “Both B and C affect E” or even “B and C interact to affect E”.

It’s crucial that a mini-map codes information about causality, not co-incidence. So the causal map “ $C \rightarrow E$ ” should **not** be interpreted along the pattern of “if you observe (a high level of) C you are more likely to observe (a high level of) E”, though that may or may not be a corollary of the causal information. The strongest and most correct interpretation is “if you intervene in the system and manipulate C, which may involve breaking any causal links from other factors to C itself, then this manipulation will produce a corresponding effect in E”.

The mini-map is also equivalent to a functional expression:

$$E_{posterior} = f(B, C, D, E_{prior})$$

This just says that the value of E is influenced via B, C, and D according to a function  $f$ , which shifts the value of E away from our previous or *prior* best guess about E to a value determined not only by the influence variables but also that prior value.

### 63.3 Focus on “lo/hi” variables and functions between them

The function  $f$  above could be anything. The app itself is not fussy and should be able to cope with any function for which a definition exists or can be written in R. BUT in Gary / James’s article, and I think in QuIP, we focus on what I call “lo/hi” variables and functions between them. lo/hi variables vary between 0 “as low as you can get” and 1 “as high as you can get”.

So in the world of lo/hi variables, we will focus on a limited set of functions.

### 63.4 Recording the actual values of variables

It is also possible to record the reported actual value of a variable. So this involves coding information about the variable, not about the causal link, i.e. the coder is using a different widget in the app / a different part of the coding sheet.

In the case of lo/hi variables, this value can have quite general meaning:

- the value within an empirical range e.g. “this summer is as hot as we have ever had” (this summer temperature .95)
- the value without reference to an empirical range e.g. “the teachers’ skills were appalling” (teachers’ skills .1)
- the proportion of a set of things which have a binary property e.g. “most of the children seem happy” (happiness of the children .75)
- the strength of the membership of something in a certain set e.g. “country X is only partly democratic” (country X democracy level .5)

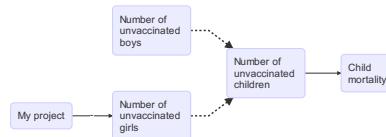
- the probability of a binary variable e.g. “the chances of war are now very low” (chance of war .1)
- the strength of our information about a binary variable

### 63.5 The rule for conceptual links

Conceptual links like “B is part of the definition of E” can be coded analogously to causal links but should be clearly distinguished from them.

The number of unvaccinated children is a conceptual combination of the number of unvaccinated girls and the number of vaccinated boys, but it isn’t caused by them. This is NOT a causal link!

The rules for causal maps which include conceptual connections too is really neat. All you have to do is code (in the app / the coding sheet) the fact that the link from a variable / a package of variables is conceptual. So for example if my project increases the number of girls who are vaccinated, and we also know that total child mortality is partially caused by the number of unvaccinated children (showing conceptual connections as dotted arrows):



..... then the causal influence travels along the lines in the direction of the arrows, oblivious to the fact that some of them are dotted, and we can infer that my project likely helped to suppress child mortality. Nevertheless the dotted lines are not causal connections.

There are two options for coding a conceptual link: directed, as above, and undirected (when the concepts are not the same but overlap in meaning).

### 63.6 The rules for coding different types of influence; single influence variable

These are the most important functions when (as usual) we just have a single influence variable:

- PLUS:

$$E_{posterior} = B$$

- MINUS:

$$E_{posterior} = 1 - B$$

- NECC:

$$E_{posterior} = B * E_{prior}$$

- SUFF:

$$E_{posterior} = 1 - (1 - B) * (1 - E_{prior})$$

... it is easy to add more.

“PLUS” is just your normal “positive” influence; if B is high, E will be high; if B is low, E will be low, and so on.

“MINUS” is the opposite. High B causes low E.

“NECC” is just a necessary condition. For example: “you can’t have E without B”. Or “if you have E, you must have / have had B”. But NECC is generalised to cope with when B (and E) are not exactly 0 or 1. So you might know that a democratic government is necessary for (say) racial tolerance, but the country in focus is only partially democratic.

“SUFF” is just a sufficient condition. For example: “if you have B, you will get E too”. SUFF is also generalised.

### 63.7 The rules for coding different types of influence; packages of multiple influence variables

The most obvious thing to do is just generalise the above functions, so for example:

- NECC:

$$E_{posterior} = B * C * D * E_{prior}$$

this is equivalent to the frequently mentioned “necessary AND”. When any of B, C, D ... are zero, the value is zero. When all of them are 1, the value is just the prior value of the consequence variable, i.e. the influence variables have no effect.

Similarly for SUFF: - SUFF:

$$E_{posterior} = 1 - (1 - B) * (1 - C) * (1 - D) * (1 - E_{prior})$$

We can extend PLUS and MINUS in a similar way, but obviously the value can too easily hit 1 for PLUS and 0 for MINUS. There are different possible ways to deal with this, for example HARDPLUS aka HARDADD just truncates the value at 1. Another possibility is SOFTADD and I am looking at Bayesian ways to combine this kind of information. There are other, simpler functions which do not take into account the prior value of the consequence variable:

There are other, simpler functions for more than one influence variable which do not take into account the prior value of the consequence variable:

- MIN:

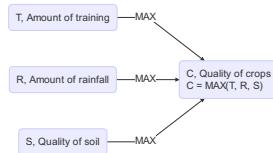
$$E_{posterior} = \min(B, C)$$

- MAX:

$$E_{posterior} = \max(B, C)$$

- MULTIPLY:

$$E_{posterior} = B * C$$

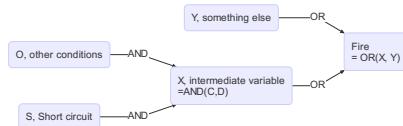


MIN and MULTIPLY are continuous analogues of Boolean AND. So we don't need AND, because for binary variables you can use either, but you should think about the difference too. The diagram above suggests that crop quality is only as good as the weakest of the three influence variables. So if rainfall is poor, there is no point trying to improve soil or increase training. MULTIPLY embodies the same idea but is more aggressive.

Any other function which makes sense over lo/hi variables, i.e. between 0 and 1, can be used, for example measures of central tendency and dispersion, like the *average* of the variables in a package. These can be entered directly into the app / coding sheet.

It is important to note that all these functions are applied to a package (even if the package contains only one variable) at once; they are a property of the whole package, not the individual variables. For this reason it is a bit misleading to display the information on the arrows.

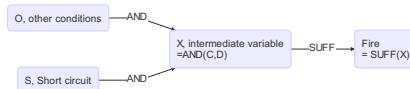
### 63.8 INUS



Personally I find INUS a big yawn, and there are different ways to render it. Its main interest is in trying to emulate the ordinary-language expression “B is a cause of E”. But as we think harder about causation, this kind of expression loses its interest. The best answer to the question “what causes what here” is just a causal map, or a simplified causal map (derived from the original using the rules of inference for causal maps), perhaps with any pedagogical / visual aids we can give the reader to understand connections and phenomena which are not obvious.

The INUS-like configuration below is somewhat equivalent to Mackie’s original. One can argue about whether the ANDs should be NECC or whether the ORs should be SUFFs. Insofar as there are other real conditions Y mentioned, there is no need to provide an option to code INUS directly because the Y (or its constituent parts) need to be coded anyway, and INUS is by its nature a complex beast with multiple parts.

I guess this is a possibility too:



It would be possible to offer a specific option to directly code this kind of combination if that was deemed important.

### 63.8.1 SOFTADD

One additional multi-function is called SOFTADD. I think this may be the most frequent case. It fills a gap for an incremental, addition-like function for lo/hi variables. Every additional positive parameter to the function potentially adds at least something to the result. So if someone says just “the training helps increase crop yield, but so does the pre-existing skill level and of course the weather”, SOFTADD is the best option. If any of these influence variables have values equal to 1, the consequence variable will also be 1, but if they are all below 1, the consequence variable will never reach 1, but will get ever closer to it.

## 63.9 Coding the strength of the influence of a package

We can also code the strength of the influence of a package, also between 0 and 1<sup>1</sup>

---

<sup>1</sup>It is also possible to think in terms of a *negative* strength, but as we already have MINUS influences, we don’t strictly need this.

### 63.10 Causal claim coding form

Pick / select / name the “**from**” variable(s) and the “**to**” variable.

- Code the type of the influence(s). Default is PLUS, but also possible are NECC, SUFF, MINUS, maybe others.
- Strength of link (according to the source) between 0 and 1
  - If (rarely) there is one or more arrows with a **different influence** or **different strength** within this package, need to click / select / specify this differing influence and/or for this variable(s)
- Add specific / “broken” aka cited text
- Source and context (can be automatically coded by the app. Can then also be linked back to the source and to characteristics of the source)
- Add the “Complete text” (also added automatically)
- Is this a conceptual link? Default is NO, but possible also are “directed” or “undirected”.
- Certainty (default is .5). Can this also optionally differ between links within the same package?
- Possibly other kinds of qualifier e.g. trust, explicitness, but only really useful if we have inference rules to tell us what difference they make

There is no reason to specifically code a causal chain as this can be constructed from several atomic pieces of coding. Trying to code multiple links at the same time would just mean repeating the interface n times. An app should give feedback and show where each new link fits in the grander scheme, and anyway records the fact that the pieces all come from the same quote.

There is no reason to specifically code cycles or loops because this can be done just by coding  $A = f(A)$  or  $A = f(B)$  and then  $B = g(A)$  etc.

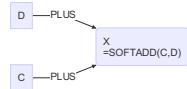
### 63.11 Problems

Some of these are really thorny.

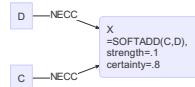
- what happens when coding creates a cycle. The only situation where one might want to code such a cycle directly is in the case of so-called “self-loops” i.e. “A influences A” (or “A and B influence A”). At the moment this is allowed in the app, and the coder can select the requisite function (e.g. MINUS) as required, as usual.
  - a coder might code “C influences D” and, immediately after or later, “B influences A”. This is also fine and also does not need a special interface. However in both these cases, causal inference becomes potentially very tricky!
- Variables vs. events
- what to do when variables have a specific time attached to them
- how to combine “package-free” combinations of influences mentioned separately by different sources. With OR? With SOFTADD?

- how to code information about a zero influence
- what role does the number of mentions of a causal link play when a causal map is aggregated / simplified?
- Can't we do it simpler? e.g. just restricting ourselves to PLUS and MINUS influences? Do we really need NECC and SUFF?
- It's much easier if the strength and type of each influence remains the same for each arrow within a single causal package.
  - (otherwise we need a separate set of tables)

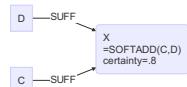
Yes: (the default)



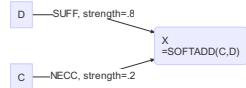
Yes:



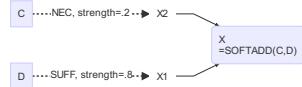
Yes:



Harder to code directly, because it is necessary to store separate information for each arrow, not just about the package. How frequent?



There is a workaround:



## **Part VI**

### **The why question**

## Section 64

# The “*why* question” as a generic method in social research

Up to now we have looked at how to process fragments of causal information. But overlapping that is a prior questions of how to get those fragments in the first place.

Here I'll briefly highlight what seems to me the minimum research strategy for eliciting causal maps, stripping off other “nice-to-have” elements. It seems to me this minimum research strategy (“the why question”) is interesting in its own right.

The “why question” method: formulating a single question about what influences what, and posing it to many sources or respondents to elicit many fragments of causal maps.

There are various different, related, methodologies in the literature. Márkiczy and Goldberg (1995), Trochim (2017), Copestake, Morsink, and Remnant (2019).

### 64.1 Steps to asking “the why question”

#### 64.1.1 Formulate ...

... a single question about what causally influences what. There are various different scenarios.

- QuIP-like: “What factors influence this important thing, and what influences those factors?”
- Appreciative: “Tell us a story about the most important positive effect the project had (what were the enabling factors, what were the consequences)”
- Impact: “Tell us a story about most important positive or negative effects the project had (what were the enabling factors, what were the consequences)”
- Research: “What do you think are the most important causes and consequences e.g. of this thing [e.g. climate change]”

- Goalfree (Sensemaker-like) “What is the most important issue right now, and what are its causes and consequences”

In each case, you can tweak in various ways

- emphasise causes and/or consequences,
- restrict respondents more or less to predefined items (etic versus emic)
- specifically ask about causes of causes and/or consequences of consequences (i.e. more than one link away from the main focus – though people often mention these things anyway).
- specifically ask people for how beneficial (or detrimental) the variables are. For example asking at the end, “... and which of these things are most important to you?”
- ask people for another example when they have finished the first (and even then ask for a third or fourth example)
- code only for structure (“X influences Y”) and parameters (“X is a strong but negative influence on Y”) of the causal network, or you can also code information about actual levels of variables (“right now, X is very high”) and even contrasting or counterfactual levels of variables (“X would be low if Z was happening”).

#### 64.1.2 Pose the question

... to several sources or respondents. You can always add, for example, official documents or even your own considered opinion to the list of evidence and perhaps assign them a high level of trust. You’ll usually ask a couple of additional questions about each source like gender, age, status.

#### 64.1.3 Get the answers into a spreadsheet

... with a column with the replies and a columns for the additional questions. (If you have allowed people to go back and provide another example, it would be good to have an additional column to record the ID of the source)

The subsequent steps (Code, Simplify / synthesise / Analyse) have already been covered in earlier sections.

### 64.2 When would you *not* want to use this approach?

- If you can guess the main answers - which / how many sources mention which things, you might be better off just formulating those issues as closed questions
- If this is a theory or theory of change which has already been discussed a lot, so there is already pretty much a consensus

So this works best with causal structures which we don’t yet know too much about. This probably comes under the buzzword “complexity-aware”.

---

*SECTION 64. THE “WHY QUESTION” AS A GENERIC METHOD IN  
SOCIAL RESEARCH*

201

“To reduce the redundant data, semi-structured CCM (Laukkanen, 1994, 1998) uses a different format. The interviewed participants answer a series of questions about what factors they perceive to influence the focal phenomenon (anchor theme) and what consequences follow from it.”

## **Part VII**

### **The why question**

## Section 65

# The “*why* question” as a generic method in social research

A *method*: formulating a single question about what influences what, and posing it to many sources or respondents to elicit many fragments of causal maps.

### 65.1 Steps to asking “the why question”

#### 65.1.1 Formulate ...

... a single question about what causally influences what. There are various different scenarios.

- QuIP-like: “What factors influence this important thing, and what influences those factors?”
- Appreciative: “Tell us a story about the most important positive effect the project had (what were the enabling factors, what were the consequences)”
- Impact: “Tell us a story about most important positive or negative effects the project had (what were the enabling factors, what were the consequences)”
- Research: “What do you think are the most important causes and consequences e.g. of this thing [e.g. climate change]”
- Goalfree (Sensemaker-like) “What is the most important issue right now, and what are its causes and consequences”

In each case, you can tweak in various ways

- emphasise causes and/or consequences,
- specifically ask about causes of causes and/or consequences of consequences (i.e. more than one link away from the main focus – though people often mention these things anyway).
- specifically ask people for how beneficial (or detrimental) the variables are. For example asking at the end, “... and which of these things are most important to you?”

- ask people for another example when they have finished the first (and even then ask for a third or fourth example)
- code only for structure (“X influences Y”) and parameters (“X is a strong but negative influence on Y”) of the causal network, or you can also code information about actual levels of variables (“right now, X is very high”) and even contrasting or counterfactual levels of variables (“X would be low if Z was happening”).

#### 65.1.2 Pose the question

... to several sources or respondents. You can always add, for example, official documents or even your own considered opinion to the list of evidence and perhaps assign them a high level of trust. You’ll usually ask a couple of additional questions about each source like gender, age, status.

#### 65.1.3 Get the answers into a spreadsheet

... with a column with the replies and a columns for the additional questions. (If you have allowed people to go back and provide another example, it would be good to have an additional column to record the ID of the source)

#### 65.1.4 Start analysing

e.g. copy and paste these answers into the first tab of the app. The app imports them and breaks each answer into individual sentences.

### 65.2 When would you *not* want to use this approach?

- If you can guess the main answers - which / how many sources mention which things, you might be better off just formulating those issues as closed questions
- If this is a theory or theory of change which has already been discussed a lot, so there is already pretty much a consensus

So this works best with causal structures which we don’t yet know too much about. This probably comes under the buzzword “complexity-aware”.

---

“To reduce the redundant data, semi-structured CCM (Laukkanen, 1994, 1998) uses a different format. The interviewed participants answer a series of questions about what factors they perceive to influence the focal phenomenon (anchor theme) and what consequences follow from it. ... A typical session can cover two to four anchor themes (and the second data layers) and takes one to two hours.” Laukkanen and Eriksson (2013)

## Section 66

# The “*why* question” – asking about changes

### 66.1 “Changes”

I don't think the right way to elicit information about someone's causal map of a particular domain (e.g. in order to find out how a project influenced that domain) is to ask about changes since a certain time point. This is probably my biggest disagreement with the QuIP way of working.

One obvious counterexample:

the government stopped fixing the riverbank, but our NGO intervened to keep fixing it regularly. If someone doesn't fix the riverbank, there are more floods.

If you ask about changes, in the literal sense, there won't be any, because the intervention served to maintain the status quo. It might be there are contrast situations like other regions where the government withdrew and no-one stepped in, or a contrast with some previous situation, but there might not be. Either way, the question about changes will only work, if it works at all, because people implicitly understand that we really want to know about the whole causal map, not just a change over time. So why not ask that directly? Why not ask, “how do things work around here vis-a-vis floods? Why are there sometimes more or fewer floods?” This is what I like to call “the why question”.

BSDR has a whole wealth of experience and caveats re how to pose these questions, what works and what doesn't as a question. So I'm hesitant to question this obviously effective way of working. But, as an evaluator I've often posed the alternative question, “how does X work around here?” and found it fine.

## **Part VIII**

### **Using the causal map app**

## Section 67

# Using the causal map app – already partly outdated!

At some point this section will be integrated into the body of the rest of the guide. But there is no point doing that now as everything is in flux.

This app is in an alpha state. That means you can use it and it should work, you can even do real coding in it, but it may crash at any point and the app itself will change in the future so you might not be able to continue any projects which you started.

Pressing the “Interrupt” button will crash the app!

### 67.1 About the tables

E.g. the table in the `Import` tab.

These are featureful.

You can drag the columns of the arrow and variable tables around, and sort the columns.

To sort, just click on the name.

Sometimes a table doesn’t load properly; just click in it. They extend themselves when you paste in a lot of information, but they can seem to freeze.

### 67.2 Left-hand tabs for producing and coding your diagram

We’ll look at each tab in the interface, starting with those on the left-hand side.

### 67.2.1 Import

Theorymaker

Interrupt Update

Import Code Aggregate Merge Display

Edit column names

Paste your sources into the space below. The table will expand if you paste more rows.

The first column is the text. There are seven other columns for attributes like age or gender.

You can change the column names.

text	female	key2	key3	key
The populations of Clevedon, Yatton, Wrington, Winscombe etc are all growing steadily. Many more people are using bicycles to commute and with no clear plans for road expansions it's obvious that the Strawberry Line has huge	0			

Here you paste in statements from several sources or respondents from which you want to extract causal claims. You can paste straight in from Excel, one statement or source per line (more rows will magically appear).

To fix: if you paste in more than about 20 rows, the interface seems to freeze. So use “upload” option.

There additional columns for data like gender, age or role of the source. These extra categories make your diagrams much more interesting.

Change the names of these columns if required by clicking on “edit column names”

To decide:

- at the moment, if you delete or paste over the data from this tab, it is lost. Perhaps it is better to have a more conventional “Import” function in which new data is appended to existing data.
- whether to have an option to break up larger passages into smaller ones.

You *could* make diagrams without looking at this tab, going straight to the **code** tab and treating everything as coming from one source. It would be a very clunky way to construct a small network, maybe an interesting way of constructing a larger one.

But you'd be missing the point of the app, which is to code and combine multiple sources.

### 67.2.2 Code

The screenshot shows the Theorymaker application interface. At the top, there is a navigation bar with buttons for 'Interrupt' (highlighted in pink), 'Update', 'Import', 'Code' (highlighted in blue), 'Aggregate', 'Merge', and 'Display'. Below the navigation bar is a pager with buttons for 'Prev', 'Next', and numbers 1, 2, 3, ..., 20, followed by 'First uncoded'. The main content area contains a statement: "...The populations of Clevedon, Yatton, Wrington, Winscombe etc are all growing steadily. Many more people are using bicycles to commute and with no clear plans for road expansions it's obvious that the Strawberry Line has huge potential to connect neighbouring communities with transport links. I'd also love to see it extended out to Wrington along the old disused line.". Below the statement are several buttons: 'Add arrows' (with input fields for start and end variables), 'View and edit arrows', 'View and edit variables', 'Combine variables', and 'Advanced options'.

Here, you view each statement one by one, looking for causal links e.g. “X leads to Y”. In some statements you might find more than one such link; many will have none.

#### 67.2.2.1 Pager

The bar at the top shows a pager with buttons from 1 to the total number of statements. If you have pasted some statements in on the previous tab, you can now click through and view the statements. The button “first uncoded” is one you will be using all the time as it takes you to the first statement which you have not yet coded; you will often use it as another “Next” button.

As you switch between statements, the arrows you coded for that statement are highlighted in the diagram.

If you have set the “overview\_column” setting to be the name of one of the attribute columns in your table of statements, then a button will appear which

will show all the statements from the current source.

#### 67.2.2.2 Widget for adding arrows

- with smart search for existing variable names or ability to add new variables
- ability to add several variables at the “from” or “two” end of the arrow (but not both) - just type a comma or press tab to enter the next one
- option to add a quote for the arrow; just highlight any text in the statement to have it added automatically (FIX)
- keyboard navigation for rapid data entry (press tab to move between options)
- option to continue a chain, i.e. when you’ve coded an arrow from A to B, B is entered as the beginning of the next arrow (FIX)
- option to change the strength and direction (plus/minus) for the arrow and trust we place in the source
- option to mark this combination of incoming variables as interacting, which means a symbol is placed on the corresponding in the diagram. To do this, when coding the fact that more than one variable controls another, i.e. you have more than one variable name in the top box, check the “Arrows interact?” checkbox. There are also be some pre-set options for the interaction like &, OR etc.

This widget is the heart of the app. It is designed to make coding quick and easy.

In the upper box you type the names of one or more variables at the beginning of the arrow(s) and in the lower box you type the names of one or more variables at the end of the arrow(s). Separate the names with a comma or by pressing your **tab** key. What’s cool is that the names of existing variables are suggested as you type. You can accept the suggestion which has a grey highlight just by pressing **Enter**.

When you have at least one variable name in both boxes, add the corresponding arrow(s) by clicking “Add this arrow”, or just by pressing **Tab** until this button is highlighted and pressing **Enter**.

Tip: if you have a statement which implies a chain, e.g. A leads to B which leads to C, you can speed up coding by pressing “Add & continue the chain”; the arrow(s) are added but also the variables from the bottom box appear in the top box, to make life easier.

To decide: at the moment, you can’t put *several* variables in the upper box and several in the lower box.

#### 67.2.2.3 Widget to View and edit arrows

This panel lets you look at the arrows you have created and edit them. Remember to press “Update” when you’ve finished. This table, like most others, is sortable (click on column headers) and you can paste into it, right-click for additional options, etc.

Here you can edit:

- the quote which goes with the statement, by default it is the whole statement or what you highlighted during coding.
- the *trust* which you as a coder have in the arrow (you can give different levels of trust to different arrows mentioned the same source)
- the *strength* from -1 to +1 which this statement claims the arrow has
- the *label* which appears by default on the arrow

#### 67.2.2.4 Widget to View and edit variables

Here you can:

- rename variables by changing the “label”
- don’t touch the `id` column unless you know what you’re doing
- you can add a `group` which is useful for drill-down scenarios

You can add any columns you want. If you add one which means something to visNetwork, like `color.background`, this will be used by the app. You can add arbitrary attributes like `mynumber` and summaries for it will be available in the visualisation if you put things like `sum_mynumber_sum_sum` in the settings.

#### 67.2.3 Merge

The point of this tab is to simplify the diagram - to reduce the number of arrows and/or variables. There are quite a few different ways to do this.

The features in this panel leave the original variables and arrows intact, so you can filter and recombine them in a different way later. What is saved is the original data plus the settings, clustering etc you want to use to merge the data the way you want it.

##### 67.2.3.1 Clusters

A better way to simplify the diagram and reduce the number of variables, probably because we want to subsume several under a new or existing variable which is more general and expresses all of them. When you activate this option, only the simplified variable is displayed instead of its constituent parts. - e.g. at some point while coding you decide you want to combine two variables such as “Safer cycling” and “Evening cycling is safer” by merging the latter into the former, “Safer cycling”. - or you want to combine “Better over long-distance” and “Better over short-distance” into one variable called perhaps “Better running”.

- Three ways to specify clusters:
  - manually in the previous tab (`code`) by editing the variables table
  - press the button to auto-generate suggestions (uses the ICLUST procedure under the hood) - you can then check and edit these manually in the `code` tab.
  - or by using the dedicated `cluster variables` widget:
    - \* auto-suggests names of existing variables

- \* the name of the combined variable is a concatenation of the variables it includes (FIX)
- \* the attributes of the combined variable (value, frequency etc) is calculated as a summary/average of the variables it includes (FIX)
- Hide arrows with fewer than a specified number of mentions
- There are two more widgets for viewing but not editing the resulting sets of arrows and variables. These are automatically now combined with attributes of the sources of the statements, e.g. to calculate % of women who mentioned the arrow or variable, **frequency** (number of arrows which have been combined) and **title** which is the tooltip shown above the arrows in the diagram.

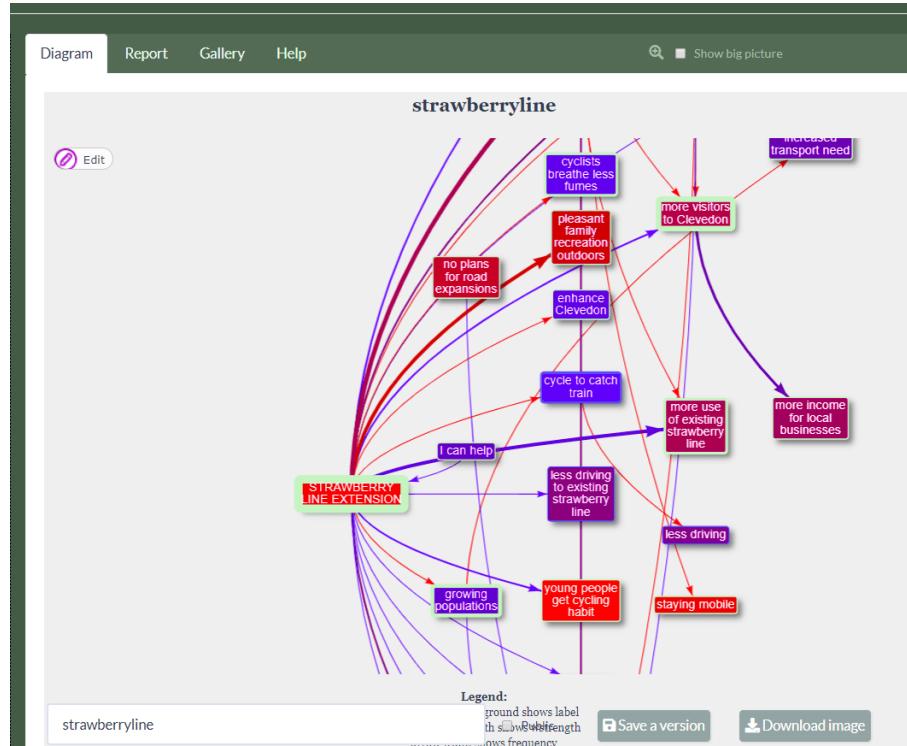
#### 67.2.4 Display

- Each display attribute like variable colour can be either:
  - set globally e.g. change the border colour of the variables (click the coloured area next to **color.border**)
  - set if... something equals, (or is greater than, or contains, etc) ... something else.
    - \* e.g. you can set all variables to be **hidden** if the % of women mentioning it is less than 50%
    - \* e.g. hide all variables containing the word “village” (under **variable**, click the control next to **hidden** and change it to **if...**, and then type **village**) (FIX: think of a way to have multiple such settings for one attribute, e.g. red for “village” and green for “town”)
  - set conditional on something e.g. set the colour or size of something to vary according to the number of mentions.
    - \* e.g. set the arrow colour to reflect the average strength (from +1 to -1), weighted by frequency of mentions and average trust (FIX: add some sensible presets)
    - \* making the width of an arrow dependent on the number of mentions (under **arrow**, click the control next to **width** and change it to **conditional on...**, and then select **frequency**)
    - \* making the colour of a variable depend on the percentage of sources who were women (assuming you have say “gender” or “female” as one of your columns in the **Import** tab, under **variable**, click the control next to **colour background** and change it to **conditional on...**, and then select **gender**)
- there are lots of other possibilities!
- **Diagram** settings (bottom section)
  - node spacing and level separation work only with the “hierarchical” option. They are important to make the text visible without crowding the diagram.

(FIX and DECIDE) getting the diagram to look “right” is quite a big deal as it is about opinion, taste & user-group as well as technology.

### 67.3 Right-hand tabs for outputs: viewing diagram etc

#### 67.3.1 Diagram



On the right is the diagram. You may need to press Update in the top-left corner to get the latest version.

##### 67.3.1.1 Ways to interact directly with the diagram

- You can drag and drop. However **at the moment, your edits will not be saved.**
- Zoom
  - zoom in and out by doing the two-finger gesture or scrolling with the mouse
- Pan
  - (move the diagram around by dragging an empty area)
- View descendants and ancestors when clicking on arrows or variables
- Add and edit variables and arrows
- View tooltips when hovering over variables and arrows.
- If you have used conditional formatting, a primitive legend is printed on the diagram which sort-of explains the meaning of the colours etc.

#### **67.3.1.2 Save**

- You can save a permanent link to your diagram. Type a name for it in the **Title** box, and press save. A link appears which you can click on, and/or copy and send to someone or bookmark for yourself. If you make further changes to your diagram, remember to save them again. If you want you can save them under the same title, or you can use a new title like “my-diagram-version-2”. The title cannot have spaces or unusual characters. Saves a version of the diagram and (most of - FIX) your present settings.
- Right-click to copy the diagram
- Save as png / html button

#### **67.3.1.3 Download your data**

e.g. statements, coding

#### **67.3.1.4 Report**

- Here we'd have some MicroStrategy-like things, either static reports and/or interactive widgets

#### **67.3.1.5 Gallery / library**

- Gallery of examples, user's recent diagrams, etc with optional short phrase describing on each one

## Section 68

# Using the causal map app for real-time, collaborative theory construction

Imagine you are at a public meeting about, say, the consequences of policy X. The moderator says, go to this link on your phone / tablet. The participants do that, and see a screen with just our two text boxes, where they can enter details of causal links relevant to policy X. They can see the whole crowd-sourced causal map as it emerges, perhaps on a screen at the front of the room. You could get a whole interactively sourced map in a few minutes. I don't think there is any other platform which will let you do this at the moment.

Details / things to tweak:

- If the people didn't already have a pretty clear idea of what a causal link is, you'd have to prime them with some examples.
- A weakness is that you might get a lot of different names for essentially the same thing like health / illness / family health / family wellness etc. So it would be cool if, when someone types something new, a pop-up appears where they are invited to say which existing factor this thing is most similar to. (I might build this functionality into the main app)

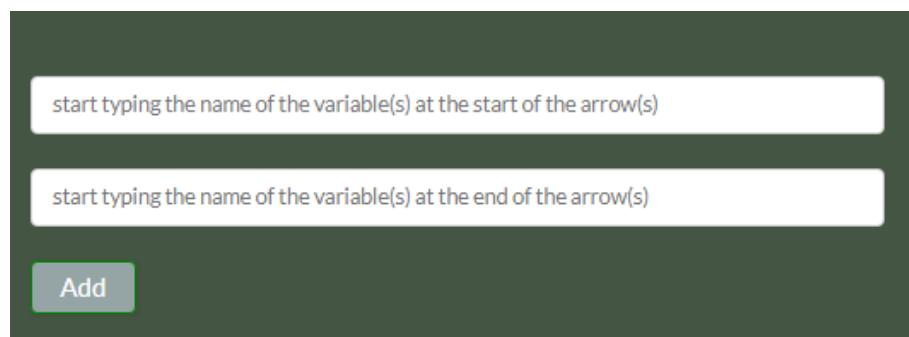


Figure 68.1: 1558002973590

anyway.) Then the map could be optionally automatically simplified by coalescing factors which are judged to be similar.

- It might be especially interesting with a room full of stakeholders who are actually part of the map, e.g. if the subject was well-being of elderly people and you had a room full of nurses, community health workers, relatives etc.
- When, as usual, the participants type the names of the factors beginning and end of each causal link, they can either type what they want or select from existing possibilities which pop up as they type. These possibilities can either be based only on the participant's own existing entries or on the growing list of everybody's entries.
- On their phones, the participants could either be allowed to see nothing, or just their own map, or the collective map.
- On their phones, the participants could be allowed or not allowed various extra things from the full version of the app, like the ability to add a note or evidence to a link, or to specify the strength or direction.
- This could either be done live and/or staggered over time like a web questionnaire.

Advantages / disadvantages:

- nice sexy wow factor
- perhaps less interesting for getting deep, quality evidence
- it's pretty simple for me to provide this simplified version of the application - I mean, it works now already but I'd need to work on hiding stuff in the interface which the participants wouldn't want or need to see.

## **Part IX**

### **Case studies**

## Section 69

# Case Study: the Strawberry Line

### 69.0.1 Theories gathered from activists

#### 69.0.1.1 The question I sent

*At last night's meeting, we talked about an easy exercise to help us build a "theory of change" for our group. Our "theory of change" will be a little story and/or diagram which will say, in a nutshell, what we are trying to do, and why. Which levers are we trying to press by doing the things we do, by e.g. hosting a petition, giving out stickers, and what is supposed to happen as a consequence?*

*One way to do this would be by all sitting round a flipchart, but another way is to ask each of you what you think, and I'll combine the ideas. We can ask the Council officers too and we can compare their ideas to ours.*

*So all you have to do, if you'd like to participate, is to send me your short answer to the two questions below. I've also provided a jokey "model answer",*

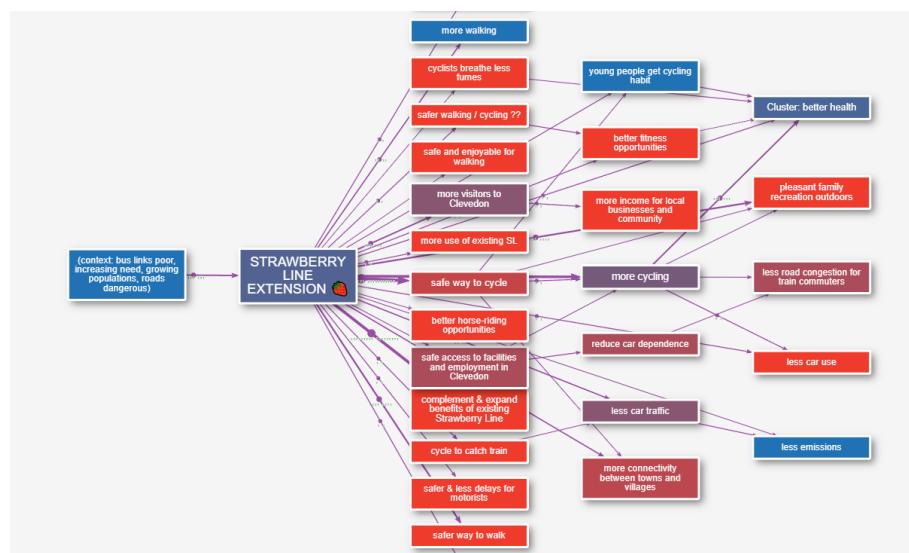


Figure 69.1: 1561808777750

*below, but you can write what you want, just 1-8 sentences altogether. There is no right or wrong way to answer and don't worry about spelling!*

*Here are the two questions: 1) What factors could lead to the successful extension of the SL to Clevedon? What do you think are the factors which will help make it happen (or make parts or sections of it happen)? Are there any factors that make it less likely? Do these helping (and hindering) factors influence one another? 2) What will happen because of the extension of the SL to Clevedon, once we have it (or parts of it)? What things will get better, will any things get worse? Do these things influence or reinforce one another?*

*Please send your answers to steve@pogol.net (having to post them here might put some people off). It isn't strictly anonymous though - I won't write your name when I combine the information into a bigger picture, but others might be able to guess who said what.*

*Jokey model answer:*

*I think the main point is to stick up so many posters that they can be seen from space. When we have enough, the aliens are sure to see them. As long as we make our idea clear enough, the aliens will obviously agree with us and either exert mind control over the council or just build the extension from space with lasers. Then everyone will see how easy cycling is and will plead with the aliens to build more and more cycle tracks, and everyone will rush to Clevedon to witness the new cycling paradise, as long as the aliens do not scare them off.*

#### **69.0.2 Stories gathered from the public**

(For another purpose)

[add more!]

## Section 70

# Case study: Global Young Academy. Tracing the paths of GYA's impact

The Global Young Academy supports young scientists around the world to connect with other young scientists, develop their careers and work towards solving global problems with science. We asked over 100 people, who had been supported by GYA or were otherwise involved in the programme, to tell us stories about positive changes which had happened because of GYA activities.<sup>1</sup> We used an online survey.

This research project was completed before work started on the Causal Map App, but using similar ideas. The causal map below was made by importing the old data into the new App.

We analysed the stories looking for examples of where people had said that *B* leads to *C* - for example, where someone said “I loved the regional meetings because they helped me widen my professional network”. Then all the *Bs* (like, *Global and regional interaction*) and all the *Cs* (like *networks, friendships, support*) were grouped into themes, which are the boxes in the diagram below: a Theory of Change for GYA. All the individual stories linking *Bs* and *Cs* are synthesised into one story using a pre-defined set of analysis steps.

Figure 1: causal map for the Global Young Academy. The numbers in the boxes (and the size of the boxes) reflect the number of people who mentioned a specific factor, and the width of the arrows reflects the number of people who mentioned that specific link. Arrows and boxes which were mentioned only by *women* are coloured *red*; those mentioned only by *men* are coloured *blue* (and shades in between red and blue were mentioned by both). Boxes with *white* borders were mentioned only by *younger* people; boxes with black borders were mentioned only by *older* people (aged over 40, including alumni, professors etc.). Grey shades in between were mentioned by both. In the online version of the diagram, if the

---

<sup>1</sup>We also asked about negative changes but this was in a different question.

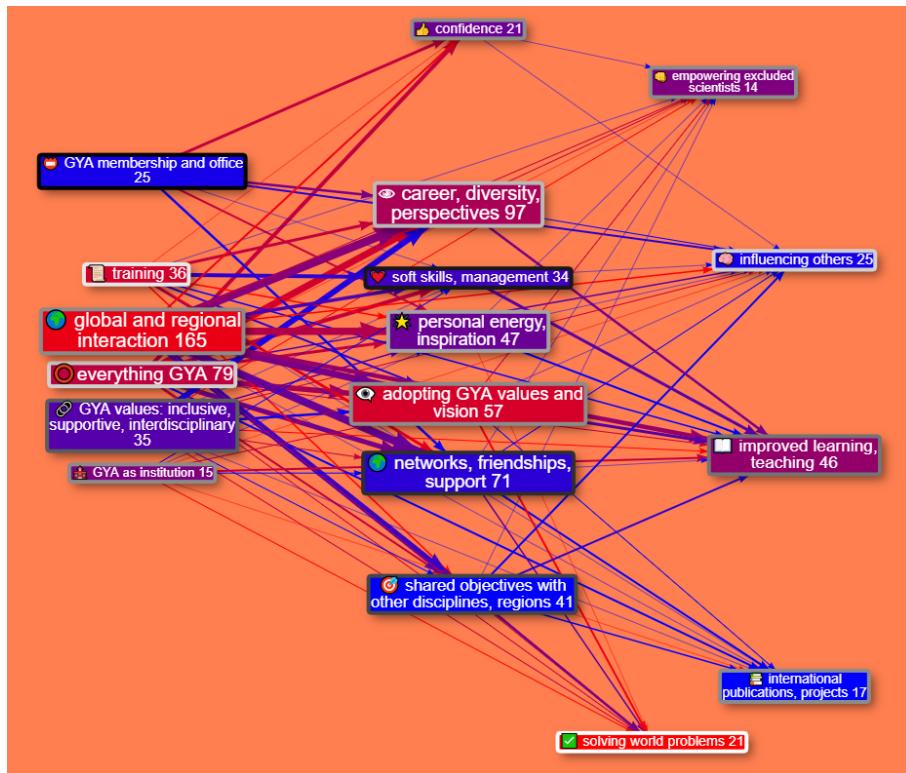


Figure 70.1: 1559329785486

user hovers over boxes and arrows they are given more information including individual quotes.

The factors which people mentioned sort themselves into three layers, reading from left-to-right: GYA inputs, individual impacts and broader impacts. So taken together, people told us about a range of GYA inputs which led mainly to a range of impacts on individuals. These impacts on individuals also led to some broader impacts like “improved learning & teaching” on the right-hand-side. People also quite frequently mentioned direct links from GYA inputs to these broader impacts. The diagram reveals a lot of insights, like these:

- People, especially women, reported that GYA activities, especially global and regional interaction like attending conferences, had many positive personal influences like establishing support networks with other young scientists.
- At the heart of the theory of change is that young scientists themselves adopt GYA’s values and vision, which enables them to go on to have broader impacts like improving teaching.
- Women, especially younger women, most often mentioned “solving world problems” as a broader impact, while men often mentioned interdisciplinary interaction leading to international projects and publications.

- Older people mentioned GYA membership and office-holding as important inputs, and most frequently mentioned how they had learned soft skills and in particular management skills through GYA participation, whereas younger people, especially women, mentioned training inputs and how they led to career opportunities.

The research report was very well-received, and the causal map was seen as central - even though we had not originally planned to analyse the stories in this way. It produced more discussion than the whole of the rest of the survey report which included dozens of conventional closed and open questions analysed with bar-charts etc. It provides not only an overall causal story but also number-based comparisons, e.g. between women and men; and the numbers are based on themes which emerged from the participants' own stories rather than being guessed-at in advance.

## 70.1 More details

Here are a few more details about this research project and the findings.

GYA's 10-year anniversary in 2018 was an opportunity to reflect on what has been accomplished so far and to inform the next 2020-2025 strategic plan to increase impact for the next 10 years. An impact evaluation was commissioned by the GYA to explore this impact further. The overall goal of this impact assessment was to understand personal narratives of impact and to generate data to inform development of the next 5-year strategic plan.

### 70.1.1 Aims

The aims of the impact assessment were to better understand:

- how members, alumni, and other young scientists experience GYA activities
- the impact of GYA activities on them (on the individual, their institutions and countries, and globally)
- to explore the contribution of the GYA to this impact

### 70.1.2 Methods

To collect these data, we sent a web survey to all GYA members, alumni, members of National Young Academies, and other young scientists (who may have had contact with GYA) inviting them to share their stories of the GYA. The survey was open from 22 Oct – 13 Nov 2018. Of the 683 people reached, 103 completed the survey.

A full report giving details of methodology and the complete analysis are available separately.

In the main part of the survey, respondents were asked for personal stories about "*how the GYA has influenced our members, other scientist/researchers,*

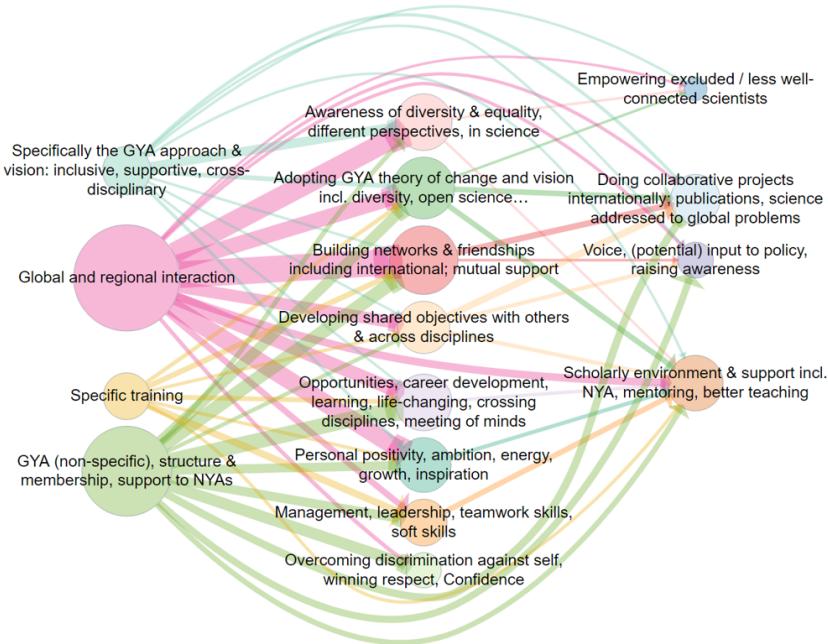


Figure 70.2: 1559329387947

*their institutions, countries, science and the world. These may include (but are not limited to) skills-building, personal development, friends, networks and connections, mentorship, primary research collaboration, publications, qualifications, policy papers published, policies enacted or influenced etc.”* Respondents could enter up to three stories as well as answering some closed questions like gender, age etc. Alongside some quantitative analysis of the responses, in the main part of the study the stories were qualitatively analysed.

### 70.1.3 Findings

The answers to the web survey were coded as described above. Initially we identified a larger group of different themes. We combined them step by step into a smaller set of over-arching themes.

Figure 2. The original causal map produced for the research report.  
There is no coding according to gender or age.

Figure 3: An intermediate version of the causal map, reanalysing the same data subsequent to the release of the report, produced with the new Causal Map App.

#### 70.1.3.1 GYA INPUTS

By far the most common theme in this group was **global and regional interaction**, in particular, meetings.

**SECTION 70. CASE STUDY: GLOBAL YOUNG ACADEMY. TRACING THE PATHS OF GYA'S IMPACT**

224

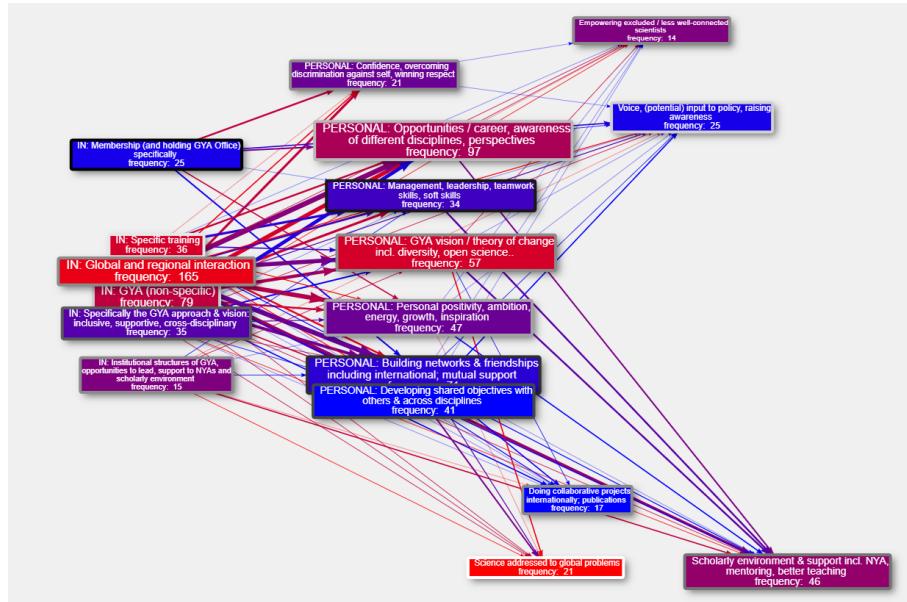


Figure 70.3: 1559226447376

I meet colleagues from other countries so I learned how they are organized in their National Young Academy (NYA). When I went back I decided to create a NYA [National Young Academy].

*Note, paragraphs formatted like the one above are quotes from actual respondent stories.*

Also, respondents often mentioned membership of GYA as relevant to their stories:

When I talk about my own Young Academy, which I often do, I always talk about how our Young Academy is also part of GYA and that gives our Young Academy more credibility and clout.

Being accepted as a member of the GYA gave me immense confidence boost...

Many respondents highlighted *GYA's inclusive, supportive, interdisciplinary values* as an important input in its own right.

I always felt like a bit of an oddball scientist because I do not conform to the “traditional” scientist model...

#### 70.1.3.2 INDIVIDUAL IMPACT

There was a wide variety of themes concerning impact on individuals. The most common was **Building networks & friendships, including mutual support.**

If I think about the greatest lasting effect of the GYA on me as a person, excluding benefits only to my career, the most significant impact the GYA has had on me is the network of friends I have built who are passionate about improving society through the work they do and the activities they volunteer on and they way they have changed my thinking.

We have described the second most common theme as *Adopting GYA's values and vision*. Here, a key personal impact is described as understanding and taking on board a specific GYA way of thinking which involves diversity, cross-disciplinary work, and taking new perspectives on familiar issues.

Realising that science truly is global, and excellent science comes from every country on earth

Some respondents describe it as quite a profound change in their understanding of and approach to science and even issues beyond science.

To give one specific story, will be the ECs response to the request from our funders to have a GYA initiative in response to the EU immigrations 'crisis'. In 2016 there was so much media attention on immigration of refugees of war into Europe, and how Europe was going to deal with this 'crisis'. Whilst the fleeing of thousands of innocent people from war was truly a humanitarian crisis, the perspective in the media came more from this being a social 'crisis' for the countries to which the immigrant were feeling, that they now had to 'deal with'. Thus, when the GYA held a meeting regarding Europe's response to this immigrate crisis, EC members from Africa countries brought to the table their objection to just an initiative, given the millions of refugees of war that Africa countries have been supporting for years, with no media attention. Their perspective was that in their countries, refugees were brought in and supported as brothers and sisters in need, not as invaders. This discussion will remain with me forever and continues to highlight the advantage to understanding a different perspective to global issues. but more importantly, the need to ensure that Europe and America do not dominate the conversation and perspective. That we continue to draw on the perspectives from a diverse experience.

A related theme was general *awareness of diversity*.

The GYA has profoundly affected how I think about cultural diversity.

This theme was later combined with another common theme: *new opportunities and career development.*

These opportunities brought me to many places, and allowed me to experience diverse cultures and connect with varied individuals, but more importantly, gave me an opportunity to learn new things, enrich my skills, discover more of my talents, and think beyond my own research field.

Another important strand, mentioned by women and men, young and old, was *confidence*. Some people said that holding GYA office helped to improve their confidence, and some mentioned that improved confidence was a factor in helping to empower excluded scientists.

The GYA has helped me to connect to a deep commitment I have to engage, to serve and to help build a better world.

*Management, leadership, teamwork skills and soft skills* were frequently mentioned, for example learning about participatory methodologies.

I started thinking about nurturing leadership in a strategical way and [thanks] to GYA for this piece of inspiration.

#### 70.1.3.3 BROADER IMPACT

Many stories mentioned broader impacts beyond those on individuals. The most frequently mentioned was *improved learning and teaching / support to the scholarly environment.*

After some planning and with the strong support of GYA members in the ASEAN region, ASEAN has established the ASEAN Young Scientists Network and the Responsible Conduct of Research Programme will be one of the key programmes of this ASEAN network.

GYA motivated me to set up the Benin Young Academy of Science, with their advices.

I feel that the existence of GYA is very important because it acts as a cohesive network for all other Young Academies

There was also frequent mention of concrete impacts such as *Doing collaborative projects internationally, publications and science addressed to global problems.*

Another GYA member and I have a joint research grant together ... In addition, we also had a joint research publication in a very high impact journal

#### 70.1.3.4 LINKS BETWEEN THEMES

The story fragments quoted above were given as examples of particular themes, but all the sentences in the stories which we categorised in fact contain a link from one theme to another.

The most frequently mentioned link was *Global and regional interaction* → *Building networks* ...:

I met incredibly inspiring colleagues from around the world that think like me, making me feel like I'm not alone which is often difficult when you are one of the few STEM PhDs in a tiny Caribbean island.

Another common link was *GYA (non-specific) ... membership* → *Overcoming discrimination ... Confidence*

Winning the membership itself has given me a strong motivation and confidence to speak up my thoughts and vision regarding young scientist's role in advancing science, health, and research in all regions of the world.

As well as people telling us about links from GYA inputs to personal factors and then on to impacts, sometimes they told us about links directly from GYA inputs (on the left of the diagram) to broader impacts (on the right of the diagram), such as the link *GYA (non-specific)* → *Scholarly Environment*...:

I feel that the existence of GYA is very important because it acts as a cohesive network for all other Young Academies

One particularly interesting set of links was from the theme “*Adopting GYA Theory of Change*”. This was often mentioned as a driver of broader impact.

GYA gatherings were always opportunities to strengthen my ideas and have insights into new ideas which were always brought back home to drive one initiative or the other in my national young academy.

This helps create a network and foundation for later long-distance work, but, moreover, helps to find common values and to identify shared ideas that can guide member activities throughout the year. Attendees think and talk about familiar and new ideas, about how to apply familiar ideas to new areas, and explore the scope for new cooperation across disciplines, countries and cultures.

## Section 71

# Appendix: previous work

Here are some things I've already written as steps on the road to understanding causal mapping. If you've read any of these, you might be interested in how "theories" relate to "causal maps".

### 71.1 From "Theorymaker" to "Causal Mapping" and from "Theories of Change" to "Causal Maps"

Previously I was mostly interested in "Theories of Change" from the planning approach of "how should we draw a diagram for this organisation's theory of change". The semi-formal logic and associated app is called "Theorymaker". That question overlaps with the question which most interests me now: "how should we encode this causal statement", and I believe the first is anyway just a special case of the second. So I've previously written about "Theories" as models of causal reality, now I'm referring to them as "Causal Maps" – it comes to pretty much the same thing.

### 71.2 Articles

- The Book of Why: The New Science of Cause and Effect. Pearl, Judea, and Dana Mackenzie. 2018 – Book Review This is now out at JMDE.
- Visualizing Theories of Change – how not to confuse causes and definitions". In: eVAL, Online Journal of the Evaluation Society in Bosnia and Herzegovina.
- Value in Theories of Change. JMDE.
- Additional notes on the definition of "Theory". Journal of MultiDisciplinary Evaluation, 15(32), 37-52

### 71.3 Presentations

- European Evaluation Society 2018 - presentation on Value in Theories of Change
- UK Evaluation Society 2018 - presentation on Judea Pearl and Theories of Change

- American Evaluation Association 2017 - A graphical language for evaluation

#### 71.4 Poster

Theorymaker poster – UKES, 2018

#### 71.5 Longer blog posts

- When an interview beats an RCT (with cartoon!)
- Scandal or yawn? Simpson's Paradox hits the IFRC
- Value for Money My first take on the VfM approach by Julian King and OPM, from a Theorymaker perspective.
- Evaluation Paradoxes – blog post at AEA365
- Various posts on LinkedIn

#### 71.6 Resources and apps

- The new app (online version, somewhat behind the development version)
- The app at GitHub
- Slides
- The original Theorymaker app
- Help for the original Theorymaker
- Theorymaker 3, the experimental version of Theorymaker – all the causal maps can be specified with typed text
- Causal Explorer

## Section 72

### References

- Befani, Barbara, and Gavin Stedman-Bryce. 2017. “Process Tracing and Bayesian Updating for impact evaluation.” *Evaluation* 23 (1): 42–60. <https://doi.org/10.1177/1356389016654584>.
- Bryson, John M, Fran Ackermann, Colin Eden, and Charles B Finn. 2004. *Visible thinking: Unlocking causal mapping for practical business results*. John Wiley & Sons.
- BSDR. 2017. “QuIP and the Yin/Yang of Quant and Qual: How to navigate QuIP visualisations,” 1–10.
- Copestake, James, M Morsink, and F Remnant. 2019. “Attributing Development Impact: the Qualitative Impact Protocol case book.” March 21, Online.
- Dul, Jan. 2016. “Necessary Condition Analysis ( NCA ): Logic and Methodology of “ Necessary but Not Sufficient ” Causality” 19 (1): 10–52. <https://doi.org/10.1177/1094428115584005>.
- Gao, Xinbo, Bing Xiao, Dacheng Tao, and Xuelong Li. 2010. “A survey of graph edit distance.” *Pattern Analysis and Applications* 13 (1): 113–29. <https://doi.org/10.1007/s10044-008-0141-y>.
- Goertz, Gary, and James Mahoney. 2006. *A tale of two cultures: Qualitative and quantitative research in the social sciences*. Princeton University Press.
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York: Farrar, Straus; Giroux.
- Kosko, Bart. 1986. “Fuzzy cognitive maps.” *International Journal of Man-Machine Studies* 24 (1): 65–75.
- Koulouriotis, D E, I E Diakoulakis, and D M Emiris. 2001. “Learning fuzzy cognitive maps using evolution strategies: a novel schema for modeling and simulating high-level behavior.” In *Proceedings of the 2001 Congress on Evolutionary Computation (Ieee Cat. No. 01TH8546)*, 1:364–71. IEEE.
- Laukkanen, Mauri, and Päivi Eriksson. 2013. “New designs and software for cognitive causal mapping.” *Qualitative Research in Organizations and Management: An International Journal* 8 (2): 122–47. <https://doi.org/10.1108/QROM-08-2011-1003>.

- Laukkanen, Mauri, and Mingde Wang. 2016. *Comparative causal mapping: The CMAP3 method*. Routledge.
- Law, International. 2005. *Process Tracing*. 1997.
- Mahoney, James, and Gary Goertz. 2004. “The possibility principle: Choosing negative cases in comparative research.” *American Political Science Review* 98 (4): 653–69. <https://doi.org/10.1017/S0003055404041401>.
- Markiczy, Livia, and Jeff Goldberg. 1995. “A method for eliciting and comparing causal maps.” *Journal of Management* 21 (2): 305–33.
- Moon, Katie, Angela M. Guerrero, Vanessa M. Adams, Duan Biggs, Deborah A. Blackman, Luke Craven, Helen Dickinson, and Helen Ross. 2019. “Mental models for conservation research and practice.” *Conservation Letters*, no. February: 1–11. <https://doi.org/10.1111/conl.12642>.
- Nadkarni, Sucheta, and Prakash P Shenoy. 2004. “Nadkarni and Shenoy 2004 -A causal mapping approach.pdf” 38 (2).
- Pearl, Judea. 1986. “Fusion, propagation, and structuring in belief networks.” *Artificial Intelligence* 29 (3): 241–88.
- . 2000. *Causality: Models, reasoning and inference*. Cambridge Univ Press.
- Ragin, Charles C. 2008. “Measurement Versus Calibration: A Set-Theoretic Approach.” *The Oxford Handbook of Political Methodology*, 1–31. <https://doi.org/10.1093/oxfordhb/9780199286546.003.0008>.
- Ragin, Charles C, and Paul Pennings. 2005. “Fuzzy Sets and Social Research.” *Sociological Methods & Research*. <https://doi.org/10.1177/0049124105274499>.
- Scavarda, A J, T Bouzdine-Chameeva, S M Goldstein, J M Hays, and A V Hill. 2006. “A methodology for constructing collective causal maps.” *Decision Sciences* 37 (2): 263–83. <https://doi.org/10.1111/j.1540-5915.2006.00124.x>.
- Scriven, Michael. 2012. “The logic of valuing.” *New Directions for Evaluation*, no. 133: 17–28. <https://doi.org/10.1002/ev>.
- Tansey, O. 2009. “Process tracing and elite interviewing: a case for non-probability sampling,” 481–96.
- Trochim, William M. 2017. “Hindsight is 20/20: Reflections on the evolution of concept mapping.” *Evaluation and Program Planning* 60: 176–85.
- Zadeh, Lotfi. 1973. “Outline of a New Approach to the Analysis of Complex Systems and Decision Processes.” *Systems, Man and Cybernetics, IEEE Transactions on SMC-3* (1): 28–44. <https://doi.org/10.1109/TSMC.1973.5408575>.