

## ARTIKEL

# Contributie-analyse: achtergronden, methode, toepassingen en problemen/beperkingen

**Themareeks Methoden en benaderingen in de beleidsevaluatie, Dick Hanemaayer, Jos Mevissen, Valérie Pattyn (red.)**

*Frans L. Leeuw\**

### De methode in hoofdlijnen

Wie over contributie-analyse (hierna: CA) leest in het handboek *Beleidsevaluatie in theorie en praktijk* (Van der Knaap et al., 2023, p. 208) zal er vermoedelijk aan overhouden dat deze methode iets van de laatste jaren is, vooral het werk van (wijlen) John Mayne (2019) betreft en in de wereld van de ontwikkelings samenwerking wordt toegepast. Zoals verderop zal blijken, is dit beeld onvolledig. Ook het SEO-rapport *Durf te leren, ga door met meten* (2022) schetst een onvolledig en op punten zelfs twijfelachtige beschrijving van CA, bijvoorbeeld als gesteld wordt dat deze methode erop gericht is ‘causaliteit te bepalen, vertrekkend vanuit de beleidstheorie [en] indicaties verzamelen[d] voor optreden van een keten van uitkomsten’. Zo simpel is het niet.

Deze bijdrage poogt verdieping in wat CA omvat aan te brengen, oog te hebben voor kennis theoretische en methodische aspecten ervan en voorbeelden te presenteren. Aan de orde komt:

- wat de achtergronden van CA zijn;
- wat de kernelementen ervan zijn (kennistheoretisch, methodisch);
- in welke contexten en beleids- en professionele velden CA gebruikt wordt;
- hoe de methode te duiden is in termen van kwalitatief/kwantitatief of mengvorm;
- hoe CA zich verhoudt tot min of meer vergelijkbare methoden en bij welke evaluatiestroming deze methode aansluit;
- welke beperkingen, respectievelijk problemen er zijn bij het toepassen van CA.

### *De kern van CA in vijf punten*

- 1 *Kan onderbouwd worden dat, als er sprake van is dat een (beleids)interventie gevolgen (bedoelde en onbedoelde) heeft, deze gevolgen mede door de interventie tot stand zijn gebracht en niet door andere factoren?* Een van de doelen van CA is dat

\* Frans L. Leeuw is socioloog en emeritus hoogleraar Maastricht University. Eerder o.a. hoogleraar evaluatieonderzoek Universiteit Utrecht, directeur WODC, Hoofdinspecteur Onderwijsinspectie en Directeur Algemene Rekenkamer.

Frans L. Leeuw

- 'beleidsmakers of onderzoekers [daardoor] meer zekerheid krijgen over de bijdrage van die interventie en wat eventuele interne of externe factoren daarbij zijn' (Van der Knaap et al., 2023, p. 208).
- 2 *Hoe en waarom draagt de interventie daaraan bij?* Met andere woorden: welke factoren, processen en mechanismen maken dat er sprake is van een bijdrage van de interventie aan de gevolgen?
  - 3 *Hoe zijn deze inzichten weer te geven in een theory of change, veranderings- of beleidstheorie of narrative?* Delahais (2023) noemt dit de fase van 'theory-development'.
  - 4 *Hoe sterk staat deze theory of change (theory-testing fase)?* Daartoe wordt (empirische) evidentie verzameld over onderliggende mechanismen en contexten die relevant zijn voor de interventie(s) en over de rol van andere factoren (Dinh et al., 2019, p. 566; Mayne, 2012, p. 273).
  - 5 *Wat zijn de hoofdconclusies en welk antwoord kan gegeven worden op de vraag wat de bijdrage van de interventie(s) aan de geconstateerde gevolgen is (of zal/kan zijn)?* Delahais (2023, p. 3) refereren hierbij aan wat hij 'a final specificity of CA' noemt, namelijk het genereren van *contributie-verhalen*.

In de afgelopen jaren zijn verschillende studies verschenen waarin op de *methodische aspecten* van CA is ingegaan (Mayne, 2017; 2019; Delahais, 2023; Delahais & Toulemonde, 2012; Lemire et al., 2012; Choi et al., 2023). Een recent overzicht biedt het 'Special Issue' van het *Canadian Journal of Program Evaluation* (2023), waarin contributie-analyse en het werk van Mayne centraal staat. Ik kom daar later op terug.

### *Achtergronden*

CA komt in belangrijke mate voort uit werk dat vanaf de jaren zeventig/tachtig onder de naam 'performance auditing' verricht werd bij verschillende nationale rekenkamers en onderzoeksinstituten. In Nederland gebruikte de Algemene Rekenkamer de term 'doelmatigheids-/doeltreffendheids-audits' (Leeuw, 1993). Mayne werkte in Ottawa bij het Office of the Comptroller General en produceerde in 1999 een discussiepaper voor de Canadese Rekenkamer, getiteld: 'Attribution Through Contribution Analysis: Using Performance Measures Sensibly'. Hij was ervan overtuigd dat een van de problemen waar performance auditors tegenaan liepen, de moeilijkheid was om onderbouwde uitspraken te doen over de doelmatigheid en doeltreffendheid van beleidsinterventies en over de mate waarin de ervaren (maatschappelijke) problemen daardoor meer of minder werden opgelost. Mayne introduceerde CA als een manier om een stap verder te komen dan tot dan toe gebruikelijk was. Tot aan zijn overlijden (2022) heeft hij aan de verfijning van CA gewerkt, iets dat geïllustreerd wordt door het bestaan van verschillende 'generaties' van CA die hij onderscheidde (Bohni et al., 2023; Leeuw, 2023).

Door de jaren heen is een steeds sterkere verbinding met *realist evaluations* tot stand gebracht, zoals verderop duidelijk wordt. CA positioneert zich daarom als een methode die onderdeel vormt van de bredere '*theory-driven traditie*' in het beleids- en evaluatieonderzoek.

Een van de andere achtergronden is de behoefte om daar waar (gerandomiseerde of andere [quasi-]) experimentele designs niet kunnen (of mogen), toch enig zicht te krijgen op wat doorgaans het attributievraagstuk wordt genoemd: het zoeken en vinden van oorzaken van een verschijnsel. *Contributie wordt dan wel gezien als een zwakkere vorm van attributie* (wat deels ook zo is).

#### *Kwalitatief, kwantitatief?*

CA is – tot nu toe – vaker kwalitatief van aard dan kwantitatief en dit mede omdat het gericht is op het vinden van een verdedigbare ‘theory of change’, die uitspraken bevat over de waarschijnlijkheid dat een interventie (van welke aard dan ook en inclusief combinaties van interventies) heeft bijgedragen of kan (en zal) bijdragen aan het realiseren van beleidsdoelen.<sup>1</sup> Met de komst van op *Large Language Models* gebaseerde ChatGPT-apps (zoals Gemini, Venice.ai en Co-Pilot) en NLP<sup>2</sup> heeft het onderscheid tussen ‘kwali’ en ‘kwanti’ echter aan betekenis ingeboet. Immers, de omvangrijke tekstbestanden (ook bestaande uit ‘gewone mensentaal’) die met moderne NLP-technieken geanalyseerd worden, waarbij teksten data zijn, maken zowel kwalitatieve als kwantitatieve analyses mogelijk.<sup>3</sup>

#### *In welke evaluatiestroming hoort deze methode thuis?*

CA vertoont aanzienlijke overeenkomsten met de *realist traditie* (Leeuw, 2023), zeker daar waar het gaat over (gedrags)mechanismen en het CMO-model (‘context-mechanism-outcome’), het ontwikkelen van een beleids-/interventietheorie/ToC, het opsporen van (‘hidden) mechanisms’ en het analyseren van bestaand onderzoek op dit gebied. Er zijn echter ook andere relevante ‘families’ die kunnen helpen de identiteit van CA goed te begrijpen. Een daarvan is de familie van *causaliteitsbenaderingen* (Palenberg, 2023, p. 390 ff; Stern et al., 2012). Zij maken een onderscheid tussen counterfactuele causaliteit (bij gerandomiseerde experimenten), generatieve causaliteit (waar realist evaluatoren zich wel van bedienen), INUS-causaliteit, geïntroduceerd door Mackie (1965), en probabilistische causali-

- 1 In de praktijk wordt ook aandacht besteed aan gewenste/ongewenste neveneffecten. Daar is bijvoorbeeld sprake van als de beleidstheorie niet over neveneffecten rept, maar er wel evidentie is dat deze er zijn.
- 2 Natural language processing, of in het Nederlands natuurlijke taalverwerking, combineert statistische technieken met machine learning-technieken. Met inzet van NLP kan automatisch informatie uit omvangrijke en ongestructureerde tekstuele data gehaald worden, gecategoriseerd worden, geanalyseerd worden, enzovoort.
- 3 Werken met NLP maakt het doorakkeren en inhoudsanalyseren van miljoenen woorden in gewone mensentaal, bijvoorbeeld uit de implementatiepraktijk of afkomstig van meldingen van burgers en ‘beneficiaries’ van beleidsprogramma’s, veel beter mogelijk, waarbij ook voorspellingen van wat de nabije toekomst aan ‘ervaringen/problemen’ kan brengen, opgesteld worden.

Frans L. Leeuw

teit.<sup>4</sup> Elementen van deze vier zijn bij CA terug te vinden. Dat geldt ook voor het concept ‘attributie’, zoals eerder genoemd. Dit wordt wel omschreven als ‘the proportion of observed y prevented by the hypothetical absence of x’ of, in iets gemakkelijker termen: ‘as whether the event y would still have occurred in the hypothetical absence of x’ (Yamamoto, 2012, p. 1, 3; Pearl & Mackenzie, 2018). Attributie ligt met andere woorden dicht aan tegen wat counterfactuele causaliteit genoemd wordt. White (2010, p. 159) schetst in een zelfde lijn de verhouding tussen contributie en attributie en noemde de discussie daarover in 2010 een van de belangrijke debatten in evaluatieland.

*Process tracing* is – volgens sommigen – een andere relevante ‘familie’ (Befani & Mayne, 2014). ‘Contribution analysis comes out of the field of evaluation, while process tracing (PT) has emerged from the analysis of historical events. Process tracing is a method that was originally formalised by Bennett (2010) and others. In PT, a causal mechanism that is believed to explain the outcome, is theorised in the form of a number of interlocked components that are all necessary for the causal mechanism to exist (Beach and Pedersen 2011).’ Befani en Mayne (2014) schetsen overeenkomsten en verschillen tussen beide en werken uit hoe de ene methode de andere kan bevruchten.

#### *Toepassingen op beleids-/professionele velden*

CA wordt op tal van beleidsterreinen toegepast: het varieert van evaluaties van ontwikkelingssamenwerking, de werking van ‘nudging’, policy advocacy<sup>5</sup> en waterbeleid tot studies waarin nagegaan wordt of communicatie- en ‘awareness beleid’ van een land bijdraagt aan ‘national character building’ van jongere generaties in dat land, of (en hoe) onderwijsprogramma’s en curricula bijdragen aan ‘health graduate capabilities’, wat de invloed van een stichting is op onderzoek in de levenswetenschappen en wat de invloed van een ‘students’ unions internal communication response to the Covid-19 pandemic is on staff motivation’. Een lijst met

- 4 Counterfactuele causaliteit begrijpt oorzaak en gevolg door middel van het vergelijken van de echte wereld met een fictieve wereld waarin de interventie niet bestaat. Generatieve causaliteit begrijpt causaliteit door middel van causale mechanismen: door die te benoemen wordt (generatieve) causaliteit blootgelegd. Mackie (1965, p. 245) introduceerde INUS-causaliteit: Insufficient but Non-redundant part of an Unnecessary but Sufficient condition. ‘In this case, the so-called cause is, and is known to be, an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result. (...) In view of the importance of conditions of this sort in our knowledge of and talk about causation, it will be convenient to have a short name for them: let us call such a condition (from the initial letters of the words italicized above), an inus condition’. Ten slotte de probabilistische causaliteit; die kadert causaliteit weer anders in en doet dat in termen van waarschijnlijkheden. Bij deze vorm van causaliteit vergroten oorzaken de kans op effecten, in plaats van ze deterministisch te veroorzaken. Palenberg (2023, table 1, p. 392) geeft een kernachtig overzicht inclusief voorbeelden van vragen die erbij passen.
- 5 Advocacy en preciezer: beleids-advocacy is het beïnvloeden van beleidsbesluiten in een bepaalde richting met behulp van ‘social marketing’, waaronder overheidsvoorlichting of via het wijzen op de rechten van de mens in ‘nieuwe’ situaties of wanneer ‘nieuwe’ problemen zich aandienen. Zie <https://advocacyguide.icpolicyadvocacy.org/21-defining-policy-advocacy>. De vaak niet uitgesproken assumptie is dat langs die weg vraagstukken kunnen worden opgelost (of in ieder geval verminderd) én dat dit ‘for the social good’ is.

verwijzingen naar relevante studies op diverse beleidsterrein is hieronder toegevoegd.

#### *Waarvoor en wanneer is deze methode zinvol inzetbaar*

CA kan ingezet worden in *ex ante* onderzoek om te helpen bij het schatten van de (vermoedelijke) gevolgen van beleid. *Ex durante* onderzoek kan ook van deze methode gebruik maken; daarbij zal de aandacht doorgaans vooral naar implementatie uitgaan. Leeuw (2012) heeft voorbeelden uitgewerkt waarbij kennis over gedragsmechanismen die opgedaan is bij onderzoek van generieke beleidsinstrumenten (zoals 'sticks, carrots, sermons & pillories'<sup>6</sup>) gebruikt kan worden bij het inzichtelijk maken van de kansen op succes of falen van er min of meer op gelijkende interventies in meer of mindere gelijkende contexten. Deze aanpak wordt wel aangeduid met de term 'subsumering' of 'subsummatie' (zie verderop).

Het gebruiken van CA in *ex post* evaluaties komt ook voor. Zie daarvoor de lijst met verwijzingen naar voorbeelden van CA-studies, maar vooral de twee uitgewerkte voorbeelden (een over ontwikkelingshulp, de ander over voedselinterventieprogramma's).

#### **Aandachtspunten bij het toepassen, inclusief beperkingen waar CA onderhevig aan is**

De methode is niet zonder beperkingen en problemen. Aandachtspunten zijn:

- de causaliteitsproblematiek; de onderzoeker moet bedacht zijn op het binnensluipen van (impliciete) claims dat CA (toch) 'causaliteit aantoot'. Daar past de oproep bij om 'confirmation bias' te voorkomen, bijvoorbeeld bij het zoeken en gebruiken van onderzoeksliteratuur. In meer algemene zin is het aan te bevelen bij het gebruik van CA een fallibilistisch perspectief<sup>7</sup> te hanteren: geen enkele claim op contributie, laat staan attributie is a priori evident, maar zal uiteindelijk steeds getoetst moeten worden;
- hoe mechanismen en de assumpties erover te articuleren zijn en op een repleceerbare manier. Juist omdat mechanismen vaak niet voor het oprapen liggen, is explicatie van de gebruikte manieren om ze te vinden belangrijk. Dat geldt ook voor het omgaan met alternatieve verklaringen die wijzen op andere factoren dan de veronderstelde contributieve;
- hoe de 'theory of change' op te stellen: gebeurt dat via deductie, inductie of abductie;
- hoe de mate wordt bepaald van de invloed van de verschillende contributiefactoren wanneer er sprake is verschillende (typen) interventies die min of meer gelijktijdig ingevoerd zijn of worden (de complexiteitsthematiek);
- hoe de verhouding tussen bevindingen uit de literatuur/onderzoek die als 'ondersteunend, onderbouwend' voor de narrative gezien worden en die juist als

6 Oftewel: wetgeving/regelgeving/handhaving; wortels = financiële prikkels; preken = voorlichting, communicatie/propaganda en schandpaal = reputatie-vernietiging via naming en shaming.

7 Zie verderop.

Frans L. Leeuw

weerleggingen of (empirische) kritiek op deze narrative te zien zijn, te wegen: wat wordt hoe zwaar gewogen en waarom?

Deels betreffen deze aspecten de validiteit en betrouwbaarheid van de methode en daarmee samenhangend de repliceerbaarheid.

Wat betreft de *kosten* van een CA is weinig meer te zeggen dan dat die ten eerste afhankelijk zijn van hoe empirische gegevens ter toetsing van het opgestelde ‘contributieverhaal’ (Van der Knaap et al., 2023, p. 308) verzameld worden. Is nieuw (primair) onderzoek nodig, dan zal een CA doorgaans aanzienlijk duurder zijn dan wanneer volstaan kan worden met het gebruik van systematische reviews en syntheses van door anderen, aan de hand van expliciete criteria, beoordeelde onderzoeken. Ten tweede: op grond van een analyse van vijf jaar contributie-onderzoeken door Delahais & Toulemonde (2012, p. 286) is vastgesteld dat het uiteenrafelen van de elementen van de ‘theory of change’ en het toetsen van deze theorie aan de hand van literatuur (vooral als van (bron)documenten en interviews met beleidsrepresentanten gebruik moet worden gemaakt) onverwacht ‘costly’ is.

### Stappen in contributie-analyse<sup>8</sup>

*Stap 1: Beschrijf het (oorzaak-gevolg) probleem dat moet worden aangepakt, waaronder de reikwijdte van het probleem en de context waarbinnen het speelt. Schets ook hoeveel ‘zekerheid’ (‘assurance’, ‘certainty’) er moet zijn bij het adresseren van het probleem.*

Toelichting:

- CA wordt weliswaar regelmatig toegepast in beleidsonderzoek, maar kan ook op typen interventies (bijv. in de gezondheidszorg) worden toegepast. Contexten zijn belangrijk. Delahais (2023, p. 1) geven aan dat CA ‘starts from the change [= bijvoorbeeld de gevolgen die waargenomen zijn] to look for contributions, rather than from the intervention being evaluated’.
- Informatie over ‘assurance’ houdt in de praktijk bijvoorbeeld in aan te geven of volstaan kan worden met een relatief globaal antwoord, gebaseerd op bijvoorbeeld een snelle scanning van (beleids)documenten of dat aanzienlijk meer (tijd) nodig is. Verken daarbij de aard en omvang van de verwachte bijdrage en van andere belangrijke beïnvloedende factoren.

*Stap 2: Ontwikkel de theory of change (ToC) en schrijf de theorie uit.*

Toelichting:

- Bij het *ontwikkelen* van de ToC is het belangrijk te weten op welke manier/manieren dit gebeurt. In de literatuur zijn daarvoor verschillende methoderegels beschreven (Pawson & Sridharan, 2010; Mayne, 2019; Leeuw, 2003; Leeuw & Waltermann, 2022).
- Denk bij het *presenteren* ervan aan het maken van een deductief geordend overzicht van afzonderlijke hypothesen én hun samenhang, of aan een meer

<sup>8</sup> Bronnen die gebruikt zijn, zijn o.a. Mayne (2001, 2012, 2017, 2019), Delahais et al. (2012, 2023), Palenberg (2023), Leeuw (2012; 2023), Ton et al. (2019).

inductief tot stand gebracht overzicht of aan een ‘verhaal’ zonder een specifieke structuur. Ook wordt het CMO (Context-Mechanism-Outcome) model, dat bekend is uit de realist traditie, gebruikt om te helpen bij het uitschrijven van de theorie. Mechanismen zijn de (cognitieve, sociale, institutionele, biologische) ‘drivers’ achter gedragskeuzen waarvan gehoopt of verondersteld wordt dat zij de interventies laten ‘werken’. De rol van andere beïnvloedende factoren en concurrerende verklaringen dient onderkend te worden.

*Stap 3: Bepaal hoe omstreden de ToC is en welke ‘risico’s’ de theorie loopt.*

Toelichting:

- Omstreden betekent: bekritiseerd door andere onderzoekers op criteria als onderbouwing, reikwijdte, empirisch gehalte; bekritiseerd door beleidsmakers/stakeholders op onder andere bruikbaarheid en nut.
- Risico’s die een theorie loopt, zijn ten eerste dat de theorie geen of nauwelijks informatiegehalte heeft (en daardoor immuun voor kritiek is). Dergelijke theorieën sluiten met andere woorden niets uit, kunnen niet weerlegd worden en dragen evenmin bij aan kennis(groei). Een tweede risico is dat er een alternatieve verklaring/theorie bestaat over de bijdrage van de (of een sterk gelijkende) beleidsinterventie aan gedragsverandering, die minder risico’s heeft, een hoger ‘assurance-gehalte’ heeft en vaker met succes getoetst is. Zie Carvalho en White (2004) voor een fascinerend voorbeeld waarin theorie en anti-theorie, onderliggend aan een interventie van de Wereldbank, gearticuleerd en onderzocht werden.
- Waar mogelijk is het dienstig ‘*nested theories of change*’ te onderscheiden van meer generieke theorieën. Een voorbeeld: er bestaat een generieke theorie over de werking van financiële incentives om bijvoorbeeld gedragsverandering te helpen realiseren. Echter, wanneer het gaat over het stimuleren van het aanschaffen en handelen in specifieke emissierechten, zijn er aanpassingen, respectievelijk preciseringen nodig naar deze specifieke context bij het ontwikkelen van de ToC. Dergelijke sub-theorieën heten wel ‘nested’, omdat zij genesteld zijn in een bovenliggende algemenere (gedrags)theorie (Richards, 2019).
- Mayne (2017, p. 162) heeft criteria die zijns inziens gebruikt moeten worden bij het beoordelen van de ‘theory of change’ in onderstaande tabel samengevat.

**Table 1.** Criteria for Theory of Change Analysis

Overall Criteria	
<b>Understandable</b>	<b>Is the logic and structure of the ToC clear?</b>
<i>Agreed</i>	To what extent is the ToC agreed or contestable?
<i>Level of effort</i>	Are the activities and outputs of the intervention commensurate with the expected results?
Criteria for Each Result	
<i>Well-defined</i>	Is the results statement unambiguous?
<i>Plausible timing</i>	Is the time frame for the result reasonable?
<i>Logical coherence</i>	Does the result follow logically from the previous result? Is the sequence plausible or at least possible?
<i>Measurable</i>	Is there a need to measure the result? How can the results be measured? What is the likely strength or status of evidence for the result being realized?
<i>M&amp;E implications</i>	What are the implications for monitoring and evaluation?
Criteria for Each Assumption	
<i>Well-defined</i>	Is the assumption unambiguous?
<i>Logical coherence</i>	Is the assumption a precondition or event for the effect sought?
<i>Justified</i>	What is the justification for the assumption as being necessary or likely necessary?
<i>Realized</i>	Is it plausible that the assumption will be realized? Are there at-risk assumptions that should be addressed?
<i>Sustainable</i>	Is the assumption sustainable?
<i>Measurable</i>	Is there a need to measure the assumption? How can the assumption be measured? What is the likely strength or status of evidence for the assumption being realized?
<i>M&amp;E implications</i>	What are the implications for monitoring and evaluation?
Criteria for Each Causal Link	
<i>Independence</i>	Are the assumptions for the link independent from each other?
<i>A sufficient set</i>	Are the set of causal link assumptions along with the prior causal factor sufficient to bring about the effect? Is the link plausible?
<i>Strength/Status of evidence</i>	What is the strength or current status of evidence for the causal link being realized?



*Stap 4: Verzamel wat aan bewijsmateriaal ('empirische evidentie') te vinden is over de ToC. Beoordeel de sterke en zwakke punten van de verbanden in deze veranderings-theorie.*

Toelichting:

- Verzamel het materiaal uit eerdere onderzoeken en evaluaties voor inzichten over (1) de waargenomen resultaten, (2) elk van de schakels in de resultatenketen, (3) andere beïnvloedende factoren, en (4) rivaliserende verklaringen.
- Wees duidelijk over waar het bewijsmateriaal uit bestaat en waar het vandaan komt. Een bron kan uiteraard het eigen (empirisch) onderzoek zijn, maar dat is lang niet altijd voldoende; daarom wordt ook en vooral gekeken naar probleemgerichte overzichten van onderzoek; 'systematic reviews' in lijn met Cochrane of andere platforms/clearing houses en instituten) en soms, als het om een eerste scanning gaat, meer informele literatuuroverzichten. Besef dat er altijd beperkingen aan deze zoekacties zijn.

*Stap 5: Stel de contributie-claim samen, beoordeel deze en betwist deze.*

Toelichting:

- De contributie-claim is het verhaal dat aangeeft waarom redelijkerwijze verwacht kan worden dat de beleidsinterventie die centraal staat, heeft bijgedragen aan het optreden van de waargenomen gevolgen (i.c. effecten).
- Geef de sterke en zwakke punten in de veranderingstheorie in het licht van het beschikbare bewijsmateriaal aan en doe dat ook voor de relevantie van de andere beïnvloedende factoren en het bewijsmateriaal dat verzameld is om rivaliserende verklaringen te ondersteunen. Verfijn of actualiseer de veranderingstheorie, indien nodig.
- Om *confirmation bias* (het al dan niet impliciet maar eenzijdig op zoek gaan naar en vinden van 'bevestigingen' van de contributie-claim) te voorkomen is het nodig niet alleen op zoek te gaan naar wat de sterke, maar vooral ook de zwakke punten in de theorie zijn. Daartoe is het nodig kennis te krijgen van (empirisch) materiaal dat de theorie ondersteunt maar bovenal van materiaal dat wijst op 'weerleggingen' of 'ernstige vraagtekens' bij de theorie (resp. claim). Als dat is gebeurd, is het mogelijk dat de veranderingstheorie bijgesteld of geactualiseerd moet worden.
- Pijnlijk is het als de theorie zo zwak is dat deze niet kan functioneren als onderbouwing van de contributie-claim. Om beleidsfiasco's te voorkomen rest dan soms de contributie-claim op te bergen in het *nationaal beleidshomeopathisch museum*. In ieder geval kan er geen '*contributie-narrative*' geproduceerd en gecommuniceerd worden.
- Voor het overzicht van de te zetten stappen in een CA laten we deze laatste twee problematische *mogelijkheden* buiten beschouwing. Daarom verder met stap 6.

Frans L. Leeuw

*Stap 6: Zoek naar aanvullend (empirisch) materiaal om de bewijskracht van de ToC en de contributieclaim te vergroten.*

Toelichting:

- Bepaal ten eerste welk soort aanvullend materiaal nodig is om de geloofwaardigheid/bewijskracht te vergroten. Gaat het om informatie van beleidsmakers en stakeholders zelf en is dat voldoende? Het antwoord is doorgaans: neen. Immers, zij zijn belanghebbenden, en slagers die hun eigen vlees keuren (en eventueel opleuken), zijn niet welkom.
- Informatie uit andere bronnen, waaronder – opnieuw – wetenschappelijke literatuur, zal dan nodig zijn. Ook kan een poging worden gedaan tot *subsummatie* (Leeuw, 2012; 2023). Daarbij gebruikt de onderzoeker kennis uit eerdere evaluaties en (ander) onderzoek door de interventie of het programma (en bij voorkeur de mechanismen ervan) onder een *algemenere ‘paraplu’* of *‘noemer’* te brengen. In plaats van te proberen nieuwe gegevens over de *specifieke interventie* op te sporen (meestal met de conclusie dat ze (nog) niet beschikbaar zijn), maakt subsummatie gebruik van wat bekend is over de mechanismen die *ten grondslag liggen aan gelijksoortige interventies in look-alike contexten*. Pawson (2002a, 2002b) heeft een dergelijke benadering toegepast in een artikel over subsidies en in een artikel over de impact van naming & shaming. Deze kunnen worden gezien als ‘generieke programma-ideeën’ (Pawson, 2013, p. 92).

*Stap 7: Herzie en versterk het bijdrageverhaal.*

Bouw (als de bevindingen uit stap 6 daar aanleiding toe geven) een geloofwaardiger bijdrageverhaal. En: beoordeel de sterke en zwakke punten ervan opnieuw en trek conclusies.

## **Wat komt er niet aan de orde?**

Wat dit stappenplan niet behandelt, zijn de volgende thema’s:

- Ten eerste de vraag wat de ‘afhankelijke variabelen’ zijn waar aandacht naar dient uit te gaan. Uiteraard moeten die sporen met beleidsdoel(en). Echter, die kunnen proximaal of distaal zijn, vaag of expliciet geformuleerd zijn en simpel dan wel complex zijn.
- Ten tweede de vraag naar de tijdshorizon. Mayne (2018) hield een pleidooi om met name ook de langetermijneffecten te betrekken in een contributieanalyse en niet te volstaan met het vergaren van kennis en feiten over de werking op korte termijn (om ‘short term-ism’ te voorkomen).
- Ten derde de vraag om welke soort gedragsmechanismen het gaat. Schmitt (2020, p. 15) onderscheidt tussen procesmechanismen (de procedures, het beheer, de organisatie van interventies betreffende) en gedragsmechanismen (de eerdergenoemde cognitieve, gedrags- enz. ‘drivers’). Het is cruciaal beide niet door elkaar te halen. Immers, vaak liggen aan procesmechanismen weer gedragsmechanismen ten grondslag.
- Contributie en attributie staan in een soms gespannen verhouding tot elkaar (zie de uitspraak dat CA een zwakke variant is van causale (attributie)analyse).

Palenberg (2023) heeft betoogd dat er verschillende causaliteitsconcepties zijn waarvan sommige meer met CA van doen hebben dan andere. Om te voorkomen dat in rapportages over CA te gemakkelijk en oppervlakkig met de causaliteitsproblematiek omgegaan wordt, is het dienstig er expliciet aandacht aan te besteden. Tot nu toe gebeurt dat zelden.

- Ton et al. (2019) maken het punt dat ‘most contribution analysis studies do not quantify the “share of contribution” that can be attributed to a particular support intervention’; in hun working paper doen ze daartoe een poging door CA en process tracing op een zodanige manier te verbinden dat enige vorm van kwantificatie daaruit voort kan komen (zie verderop, bij voorbeeld 1).
- Opteren voor het fallibilistisch perspectief in plaats van aandacht te besteden aan het onderbouwen van de ToC, soms tegen beter weten in. Rescher (2024) omschrijft deze benadering als volgt: ‘Fallibilisme is een filosofische doctrine die stelt dat onze wetenschappelijke kennisclaims altijd kwetsbaar zijn en onjuist kunnen blijken te zijn. Wetenschappelijke theorieën kunnen niet categorisch als waar worden beschouwd, maar alleen als met enige waarschijnlijkheid om waar te zijn.’ Alhoewel Karl R. Popper niet genoemd wordt, is het evident dat zijn werk eveneens een fallibilistisch karakter heeft.
- Complexiteit. Hoe CA’s te doen als overheden en semi-overheden (maar ook de private en quasi-private sector) gelijktijdig diverse, soms tegenstrijdige interventies vanuit hun beleidsfabrieken afvuren op de samenleving, resp. op een zelfde ‘gewenst’ gedrag? Voeg daaraan toe dat mens en maatschappij herinneringen hebben aan eerdere (en ook weer soms tegenstrijdige) interventies waarmee ze te maken hadden, en de complexiteit neemt verder toe. Dán wordt het uitvoeren van een CA een complexe maar vooral hachelijke opdracht (McDavid, 2023).
- Hoe criteria uit kwalitatief onderzoek, zoals geschetst door Biradavolu (2024), te gebruiken en met welke gevolgen bij het beoordelen van een CA, is niet a priori duidelijk.<sup>9</sup>

## Voor als men er meer over wil weten

Twee voorbeelden:

**Voorbeeld 1:** *Uitzoeken wat de bijdrage(n) is/zijn van de CDC (de voormalige Commonwealth Development Corporation, nu het Britse instituut voor financiering van de private sector in ontwikkelingslanden) aan het realiseren van ontwikkelingsdoelen.*

Ton et al. (2019; 2022) beschrijven het proces hoe tot inzichten te komen over de bijdrage(n) van CDC aan het realiseren van bepaalde doelen; welke stappen gezet worden (in een iteratief proces) om tot een *robuuste beleidstheorie* te komen, hoe die theorie *getoetst* wordt, welke data nodig zijn, tegen welke problemen de onderzoekers aanlopen en welke ‘ways out’ zij vervolgens presenteren.

9 Het gaat om criteria zoals ‘cognitive empathy’ van de evaluator, ‘heterogeneity’ en ‘palpability’. Zie: <https://iaes.cgiar.org/spia/news/criteria-judge-qualitative-research>.

Frans L. Leeuw

Tegen de achtergrond van inzichten uit de gedrags- en cognitiewetenschap (m.n. het COM-B-model,<sup>10</sup> dat verwantschap vertoont met het in ‘realist evaluations’ gebruikte CMO-model (Context Mechanism-Outcome)) ontwikkelden de onderzoekers een *generieke ToC*. Deze ToC wijst op veranderingen in capaciteit en gedrag die nodig zijn om de verwachte uitkomsten te realiseren; ook wordt aandacht besteed aan de aannames achter deze veranderingen. De gedragsveranderingen zullen naar verwachting resulteren in directe voordelen voor mensen, die op hun beurt zullen leiden tot een verbeterd welzijn.

De onderzoekers geven aan dat niet volstaan kan worden met het opstellen en presenteren van een theory of change, zeker als het gaat om het realiseren van de bijdrage van CDC-financiering aan ‘hogere’, d.w.z. meso- en macrodoelen (dan aan meer concrete gedragsveranderingen op microniveau). Daarom breiden ze de toepassing van CA uit door een ‘system map of the non-linear dynamics and interactions between outcomes’ te ontwikkelen. Verder maken ze gebruik van inzichten uit process tracing studies.

Een belangrijk onderdeel van het werk is ‘ranking contributions and discarding alternative explanations’. Dit sluit aan op een van de in dit artikel eerdergenoemde aspecten van CA, namelijk voorkomen dat ‘alternatieve verklaringen’ over het hoofd gezien worden. Omdat CA-vragen niet met ja/nee beantwoord kunnen worden, maar eerder door een reeks logische stappen, die elk ‘het vertrouwen vergroten dat de interventie een impact heeft gehad’ (Befani & Mayne, 2014, p. 17), worden de veronderstelde verklaringen (assumptions, theorie, narratief) over de bijdrage van CDC afgezet tegen (concurrerende) alternatieve verklaringen. Daarbij gebruiken ze verschillende tests waaronder de hoop-test.<sup>11</sup> Het resulteert in een narratief met gevolgtrekkingen van deze soort: ‘In land X CDC is de belangrijkste financiële bijdrage aan verandering Y door Z’ (p. 5).

Aparte aandacht besteden Ton et al. (2022) aan de vraag hoe de contributie te kwantificeren is, met andere woorden: wat het aandeel van de ‘contributie[factor]’ is aan de ontwikkelingsdoelen. Voor het ‘Private Enterprise Program Ethiopia’ zetten ze de stappen uiteen hoe zij tot (enige vorm van) kwantificering zijn gekomen, in de vorm van onder andere ‘contribution scores’.

**Voorbeeld 2:** *Uitzoeken wat de bijdrage(n) is/zijn van het ‘Crunch&Sip® nutrition primary prevention program that aims to increase fruit, vegetable (FV) and water consumption among primary school children in New South Wales, Australia’ (Biggs et al., 2014, p. 215-216).*

De auteurs pasten het stappenplan van CA toe, maar beginnen met op te merken dat sinds het ontstaan van het programma Crunch&Sip in 2008 zo’n vijftien

10 Het COM-B model van gedragsverandering (Michie et al., 2011) zegt dit: ‘behaviour (B) [change] occurs as the result of interaction between three necessary conditions, capabilities (C), opportunities (O) and motivation (M)’.

11 ‘A hoop test prescribes that a certain CPO [causal process observation] must be present for a hypothesis to be valid. Failing a hoop test rejects a hypothesis but passing a hoop test does not corroborate a hypothesis’ (Podesta, 2023, p. 56). Naast deze test zijn er andere, zoals de *smoking gun test*. Daarbij geldt het omgekeerde: ‘if a given piece of evidence is present, then the hypothesis must be valid’ (Mahoney, 2012, p. 576).

gezondheidsvoorlichtingscampagnes erover zijn gevoerd, gericht op zo'n 2000 basisscholen. Alhoewel er naar de effecten van FV-programma's (in dit geval in Australië) verschillende experimentele evaluaties uitgevoerd zijn, bleek dat zij 'do not always provide information to understand why the program has worked, nor do they provide adequate information about the assumptions and underlying mechanisms that influence program outcomes' (Biggs et al., 2014, p. 216). Vandaar dat Biggs et al. een beroep deden op de CA-aanpak. In hun artikel laten de auteurs zien hoe systematisch de verschillende (zes) stappen van de contributie-analyse behandeld zijn, waarbij ze in een overzicht ('rating system to assess the degree of influence') de vraag beantwoorden welke factoren van 'Crunch&Sip have contributed to increased consumption of fruit and vegetables' (p. 223 ff). De assumpties die er aan ten grondslag liggen, zijn gearticuleerd, mechanismen zijn benoemd, alternatieve verklaringen en factoren behandeld (waaronder 'rivaliteit' tussen het narratief en de alternatieve verklaringen: betreft het rivaliteit op implementatiegebied, of op theoretisch gebied bijvoorbeeld) en hoe de 'contributie' uiteindelijk te karakteriseren is in termen van zekerheid, robuustheid, reikwijdte enz.). Biggs et al. (2014) keken ook naar het proces van implementatie en welke stappen gezet kunnen worden om de 'contribution story' te verstevigen.

### Leestips

- Special issue *Canadian Journal of Evaluation* 2023, Vol. 37, No. 3. Dat is geheel gewijd aan werk van John Mayne, CA, toepassingen, vergelijkingen met andere benaderingen en methoden.
- Special Issue *Evaluation* 2012, Vol. 18, No. 3. Geheel gewijd aan CA (methodologie, toepassingen, begripsverduidelijking, vijf jaar terugkijken enz.).
- Delahais, 2023
- Ton et al., 2019
- [www.toolboxbeleidsevaluaties.nl](http://www.toolboxbeleidsevaluaties.nl), waar een fiche over de methode is te vinden (met vindplaats)

### Lijst met voorbeelden van CA-toepassingen op verschillende terreinen

Adianto, T. et al. (2019). Contribution analysis of the state Defense awareness programme towards the national character building. *Journal Pertahanan*, 5(2), 26-37.

Globalisering en de negatieve gevolgen voor de eigenheid van het nationale karakter van Indonesië; literatuuronderzoek om de bijdrage van een communicatie- en awareness-programma aan het beleidsdoel in kaart te brengen.

Choi, T. et al. (2023). Using contribution analysis to evaluate health professions and health sciences programs. *Frontiers in Medicine*, 10, Article 1146832. <https://doi.org/10.3389/fmed.2023.1146832>

Frans L. Leeuw

Gezondheidszorg en onderwijs(evaluaties); bijdrage ervan aan de ‘impact and outcomes of these programs on developing health graduate capabilities’.

Delahais, T. (2024). Setting the stage for contribution claims. In A. Koleros et al. (Eds.), *Theories of change in reality* (pp. 15-25). Taylor & Francis.

Dit hoofdstuk past CA toe op MOPAN (= The Multilateral Organisation Performance Assessment Network). Dit (logic) model wordt gebruikt om ‘organisational performance arrangements of multilateral organisations such as the United Nations Agencies on behalf of its member countries’ te evalueren. Delahais stelt de vraag wat de bijdrage(n) van MOPAN nu eigenlijk is/zijn aan het vergroten van de performance en effectiviteit van MO (Multilateral Organizations).

Gadda, M. et al. (2019). Making children’s rights real: Lessons from policy networks and contribution analysis. *The International Journal of Human Rights*, 23(2), 392-407.

Beleidsnetwerken rondom een VN-conventie inzake de rechten van het kind en de rol van CA ‘to evaluate recent developments in children’s right advocacy’.

Junge, K., Cullen, J., & Iacopini, G. (2020). Using CA to evaluate large-scale, transformation change processes. *Evaluation*, 26(2), 227-245.

CA toegepast op ‘emergent, large-scale and transformation change processes’.

Kitching, M. (2021). A mixed methods contribution analysis of UK students’ unions’ internal communications response to addressing staff motivation during the Covid-19 pandemic. *Tuning Journal of Higher Education*, 9(1), 207-237.

Hoger onderwijs; contributie van de interne communicatie van studentenvakbonden op de motivatie van de staf tijdens de COVID-19-pandemie.

Livingston, W., Madoc-Jones, I., & Perkins, A. (2020). The potential of contribution analysis to alcohol and drug policy strategy evaluation: An applied example from Wales. *Drugs: Education, Prevention and Policy*, 27(3), 183-190.

Drugsbeleid en strategieontwikkeling in Wales.

McLaughlan, T. (2022). Chai Chats: an online teacher-training program of observation and social connectedness. *Educational Technology and Society*, 25(1), 92-107.

Evaluatie via een CA van een ‘international, online, content skills-based teacher education program’.

SEO. (2022). *Durf te leren, ga door met meten*. Ministerie van EZ, Den Haag.

Aandacht voor CA bij het zoeken en vinden van kaders en methoden voor de evaluatie van systeem- en transitiebeleid in Nederland.

Stokes-Cawley, O. et al. (2021). Economic contribution analysis of national estuarine research reserves. *Water*, 13, 1596. <https://doi.org/10.3390/w13111596>

Bijdrage aan de economische waarde van 'protected natural areas for coastal habitats' en rol van het National Estuarine Research Reserve System. Het betreft een kwantitatieve CA.

Verder geeft het special issue van *Evaluation* (2012) onder redactie van John Mayne een aantal voorbeelden van toepassingen.

## Literatuur

- Beach, D., & Pedersen, R. (2011). What is process-tracing actually tracing? The three variants of process tracing methods and their uses and limitations. *APSA 2011 Annual Meeting Paper*. <https://ssrn.com/abstract=1902082>
- Befani, B., & Mayne, J. (2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *Institute for Development IDS Bulletin*, 45(6), 7-36.
- Biggs, J., Farrell, L., Lawrence, G., & Johnson, J. (2014). A practical example of contribution analysis to a public health intervention. *Evaluation*, 20(2), 214-229.
- Biradavolu, M. (2024). *Criteria to judge qualitative research. homestanding panel on impact assessment* [Blog]. CGIAR.
- Bohni, S., Lemire, S., & Montague, S. (2023). Remembering John Mayne – A practical thinker and a thinking practitioner. *Canadian Journal of Program Evaluation*, 37(3), 297-305.
- Carvalho, S., & White, H. (2004). Theory-based evaluation: The case of social funds. *American Journal of Evaluation*, 25(2), 141-160.
- Choi T. et al. (2023). Using contribution analysis to evaluate health professions and health sciences programs. *Frontiers in Medicine*, 10, Article 1146832. <https://doi.org/10.3389/fmed.2023.1146832>
- Delahais, T. (2023). Contribution analysis. *LIEPP Methods Brief 44* (pp. 1-6).
- Delahais, T. (2024). Setting the stage for contribution claims. In A. Koleros et al. (Eds.), *Theories of change in reality* (pp. 15-25). Taylor & Francis.
- Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of practice. *Evaluation*, 18(3), 281-293.
- Delahais, T., & Toulemonde, J. (2017). Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management? *Evaluation*, 23(4), 370-388.
- Dinh, K., Worth, H., Haire, B., & Hong, K. T. (2019). Confucian evaluation: Reframing contribution analysis using a Confucian lens. *American Journal of Evaluation*, 40(4), 562-574.
- Leeuw, F. L. (1993). Doelmatigheidsonderzoek bij de Algemene Rekenkamer: Ontwikkelingen in de afgelopen tien jaar en schetsen voor de toekomst. *Bestuurswetenschappen*, 46(1), 8-25.

Frans L. Leeuw

- Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation*, 24(1), 5-20.
- Leeuw, F. L. (2012). Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation*, 18(3), 348-363.
- Leeuw, F. L. (2023). John Mayne and rules of thumb for contribution analysis: A comparison with two related approaches. *Canadian Journal of Program Evaluation*, 37(3), 403-421.
- Leeuw, F. L., & Waltermann, A. (2022). On identifying assumptions underlying legal arrangements: Some conceptual and methodological considerations. *Law and Method*, 11 (April-June), 1-21.
- Lemire, S., Bohni, S., & Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation*, 18(3), 294-309.
- Mackie, J. (1965). Causes and conditions. *American Philosophical Quarterly*, 2(4), 245-264.
- Mahoney, J. (2012). The logic of process tracing tests in the social sciences. *Sociological Methods and Research*, 41(4), 570-597.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270-280.
- Mayne, J. (2017). Theory of change analysis: Building robust theories of change. *Canadian Journal of Program Evaluation*, 32(2), 155-173.
- Mayne, J. (2018). *Contribution analysis and the long-term perspective: Challenges and opportunities*. Unpublished paper, July 14.
- Mayne, J. (2019). Revisiting contribution analysis. *Canadian Journal of Program Evaluation*, 34(2), 171-191.
- McDavid, J. (2023). Enduring themes in John Mayne's work: Implications for evaluation practice. *Canadian Journal of Program Evaluation*, 37(3), 355-370.
- Michie, S., Van Stralen, M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(42), 1-11.
- Palenberg, M. (2023). Causal claims in contribution analysis. *Canadian Journal of Program Evaluation*, 37(3), 389-402.
- Pawson, R. (2002a). Evidence-based policy: The promise of realist synthesis. *Evaluation*, 8(3), 340-358.
- Pawson, R. (2002b). Evidence and policy and naming and shaming. *Policy Studies*, 23(3), 211-230.
- Pawson, R. (2013). *The science of evaluation*. Sage.
- Pawson, R., & Sridharan, S. (2010). Theory-driven evaluation of public health programmes. In A. Killoran & M. Kelly (Eds.), *Evidence-based public health: Effectiveness and efficiency* (pp. 43-62). Oxford University Press.
- Pearl, J. & Mackenzie, D. (2018). *The Book of Why: The new science of cause and effect*. Basic Books.
- Podesta, F. (2023). Combining process tracing and synthetic control method: Bridging two ways of making causal inference in evaluation research. *Evaluation*, 29(1), 50-66.
- Rescher, N. (2024). Fallibilism. In *Routledge Encyclopaedia of Philosophy*, <https://doi.org/10.4324/9780415249126-P019-1>
- Richards, R. (2019). *The value of theory of change at the portfolio level in large-scale projects*. K4D Helpdesk Report. Institute of Development Studies, Brighton.
- Schmitt, J. (2020). The causal mechanism claim in evaluation: Does the prophecy fulfil? *New Directions for Evaluation*, 16(October), Special Issue 'Causal Mechanisms in Program Evaluation' (Ed. J. Schmitt), 11-26.
- SEO. (2022). *Durf te leren, ga door met meten*. Ministerie van EZ.



- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. DfID.
- Ton, G. et al. (2019). *Contribution analysis and estimating the size of effects: Can we reconcile the possible with the impossible?* Centre for Development Impact Practice Paper. Institute of Development Studies, Brighton.
- Ton, G., Taylor, B., & Koleros, A. (2022). Assessing the contribution to market system change of the private enterprise programme Ethiopia. *Institute for Development Studies IDS Bulletin*, 53(1), 63-86.
- Van der Knaap, P., Pattyn, M., & Haanemayer, B. (2023). *Beleidsevaluatie in theorie en praktijk*. Boom bestuurskunde.
- White, H. (2010). A contribution to current debates in impact evaluation. *Evaluation*, 16(2), 53-162.
- Yamamoto, T. (2012). Understanding the past: Statistical analysis of causal attribution. *American Journal of Political Science*, 56(1), 237-256.