

Des Kaisers neue Theorien

**Wie viel Theorie steckt in unseren Wirkungsmodellen
und welche Theorie braucht die Evaluationspraxis?**

Keynote zur Frühjahrstagung 2023

des AK Methoden in der Evaluation

in der DeGEval – Gesellschaft für Evaluation e.V.

19. Juni 2023, Statistisches Bundesamt, Wiesbaden



„Kejserens nye klæder“

Hans-Christian Andersen (1837)

„Am Hofe eines Kaisers, der viel Wert auf feinste Kleidung legte...“

A little bit of history: ca. 1930er/1940er Jahre

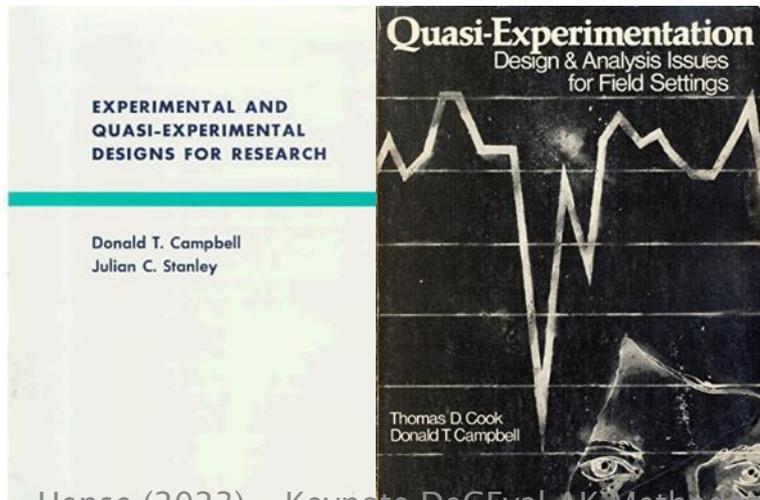
Ralph W. Tyler

- 8-year study
- Evaluation mittels sachfremder vs. gegenstandsspezifischer Kriterien
- „objectives-based“ evaluation
- Evaluation =
Messung der Zielerreichung

A little more history: ca. 1960er

"Once upon a time, the evaluation researcher needed only the 'Bible' ('**Old Testament**', Campbell and Stanley, 1963; '**New Testament**', Cook and Campbell, 1979) to look up an appropriate research design and, hey presto, be out into the field."

(Pawson & Tilley, 1997, p. 1)



Krisen der Evaluation (ca. 1970er/80er)

Auftragskrise

- Öl-/Wirtschaftskrise und neoliberaler Turn
- Nutzungskrise
 - Carol Weiss, Michael Patton, Marvin Alkin
- Null-Effekt Krise
 - Rossi's laws of evaluation: "The better designed the impact assessment of a social program, the more likely is the resulting estimate of net impact to be zero."

Sündenbock „Black box“ evaluation?



**„... erschienen zwei
Weber und
behaupteten, sie
könnten ihm ganz
einzigartige Kleider
anfertigen ...“**



(Programm-)Theoriebasierte Evaluation

“[we should] spend more effort on understanding how programs work than on the effort to find out whether or not they actually work in some specific and non generalizable instance”

Peter H. Rossi & Huey-Tsyh Chen

(vgl. Chen & Rossi, 1980, 1983; Chen, 1991, 2004)

Phase 1: Ideengeschichtliche Wurzeln (ca. 1900-1930er)

Scientific Management (Taylorismus)

- Aufdecken von Ineffektivität
- Methode: Analyse von Handlungsabläufen durch Zerlegung in Einzelbestandteile ("Kausalketten")

Übernahme in den Schulbereich

- Schule als Bildungs-"Fabrik"
- Effektivität von Schule kann wie die einer Fabrik verbessert werden

Gargani (2003)

Phase 2: Erste Spuren in der frühen Evaluation (ca. 1930er – 1960er)

Ralph W. Tyler

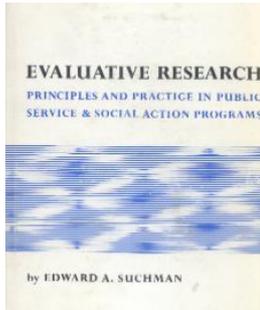
Curriculum als Hypothese, die die
Evaluation überprüfen soll:

"the curriculum of every school is based
upon hypotheses

...
no careful evaluation has been made to
check the validity of its hypotheses"
(zit. n. Gargani, 2003, p. 18)

Gargani (2003)

Phase 3: Klassiker der modernen Evaluation (ca. 1960er- 1980er)



Edward A. Suchman (1967)

- "means–ends hierarchy" (Kausalkette)
- "A test of 'Does it work?' presupposes some theory as to why one might expect it to work." (p. 86)

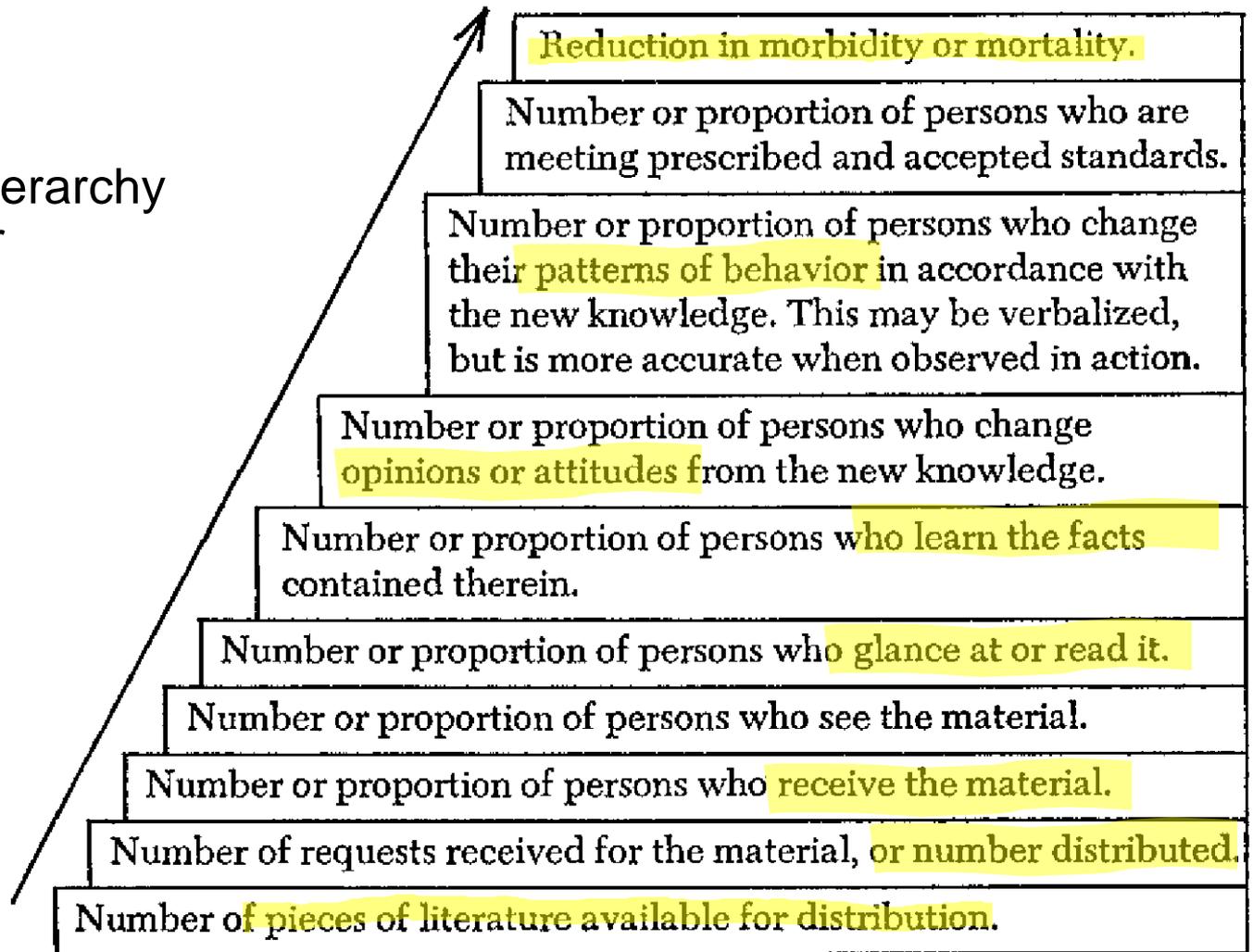
Carol H. Weiss (1972)

- implementation failure vs. theory failure
- Programmtheorie explizit als Basis der Evaluation

Joseph S. Wholey (1979)

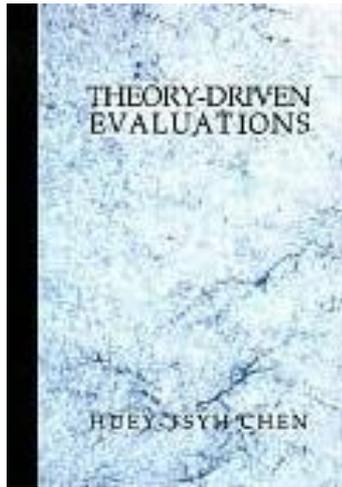
- Logische Modelle als Teil der Evaluierbarkeitsanalyse

Means-ends hierarchy
für Material zur
Gesundheits-
aufklärung

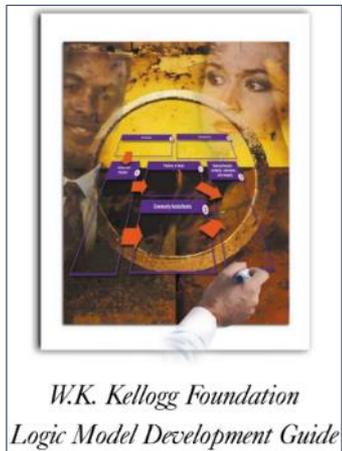


Suchman (1967, p. 55)

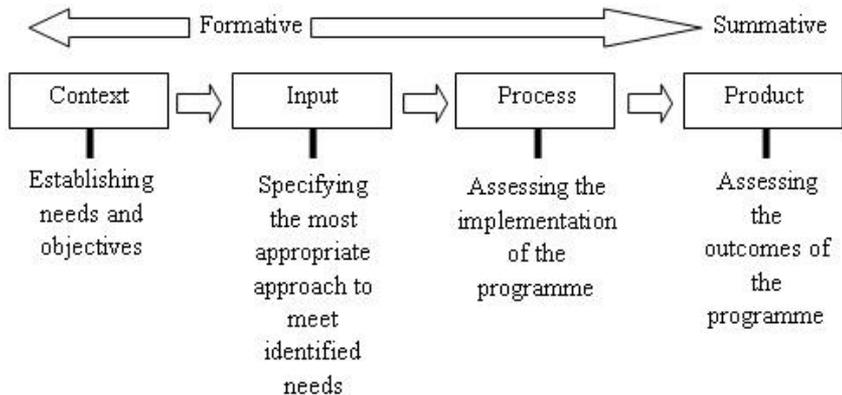
Phase 4: Gegenwart (seit ca. 1990)



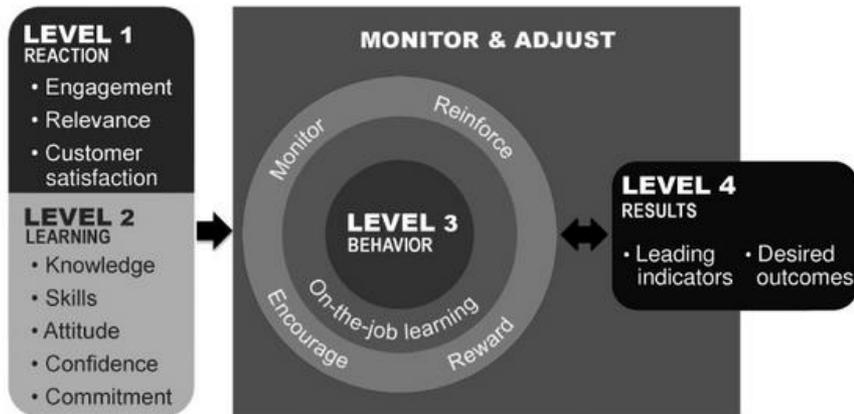
- Erste Monographie zum Thema (Chen, 1990)
- Logisches Modell bei großen Stiftungen
 - W.K. Kellogg Foundation / United Way of America
 - Vorher schon: LogFrames als Werkzeug in der EZ
- "Boom": Wirkungsmodelle werden fester Bestandteil der Fachdiskussion und des Methodenrepertoires vieler Evaluierender
- Zunehmende Rezeption im deutschen Sprachraum



Für's Protokoll: Zwei Missverständnisse



CIPP-Modell der Evaluation
(Dan Stufflebeam, 1968ff)

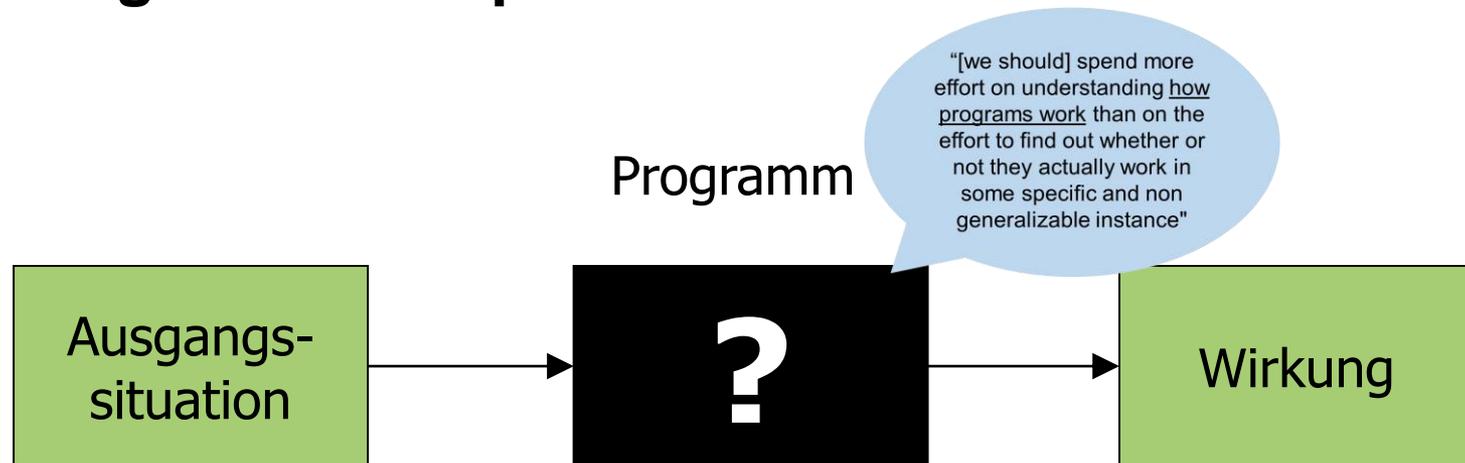


Vier Ebenen der Evaluation
(Kirkpatrick, 1959/60ff)

**„Begeistert erteilte
der Kaiser ihnen
den Auftrag.“**



Ursprüngliches Versprechen: Öffnen der Black Box

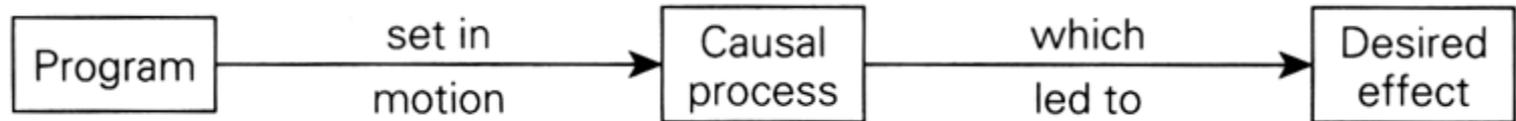


➔ Kritik an rein outcome-orientierten Evaluationen, z.B.

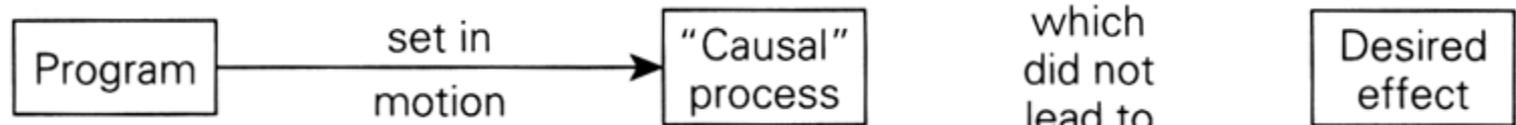
- Rein zielorientierte Evaluation
- Experimentelle und quasiexperimentelle Designs (RCTs)

Theory failure vs. implementation failure

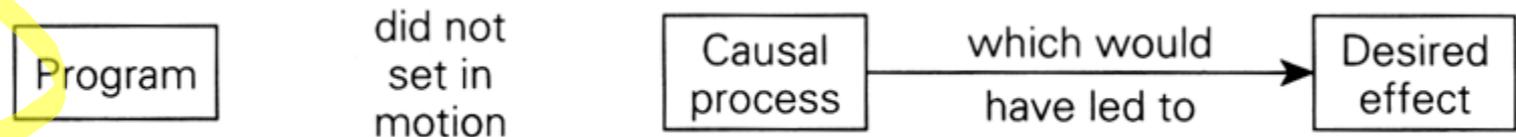
Successful program



Theory failure



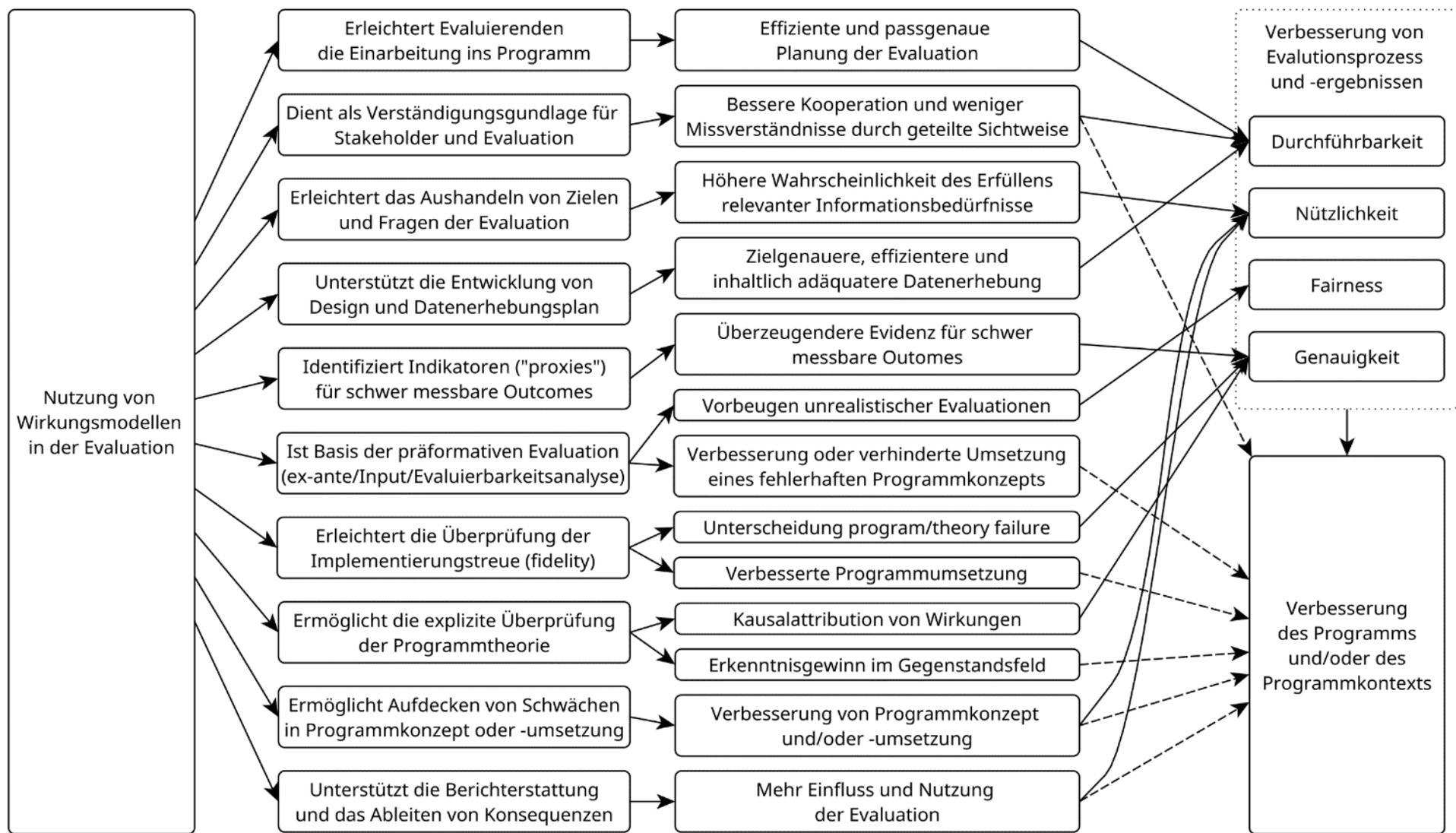
Implementation failure



(Weiss, 1998, p. 129)

Weitere Versprechungen: Wirkungsmodelle als Mehrzweckwerkzeug in der Evaluation





(Hense & Taut, 2021)

Wie schön sind des Kaisers neue Kleider wirklich?

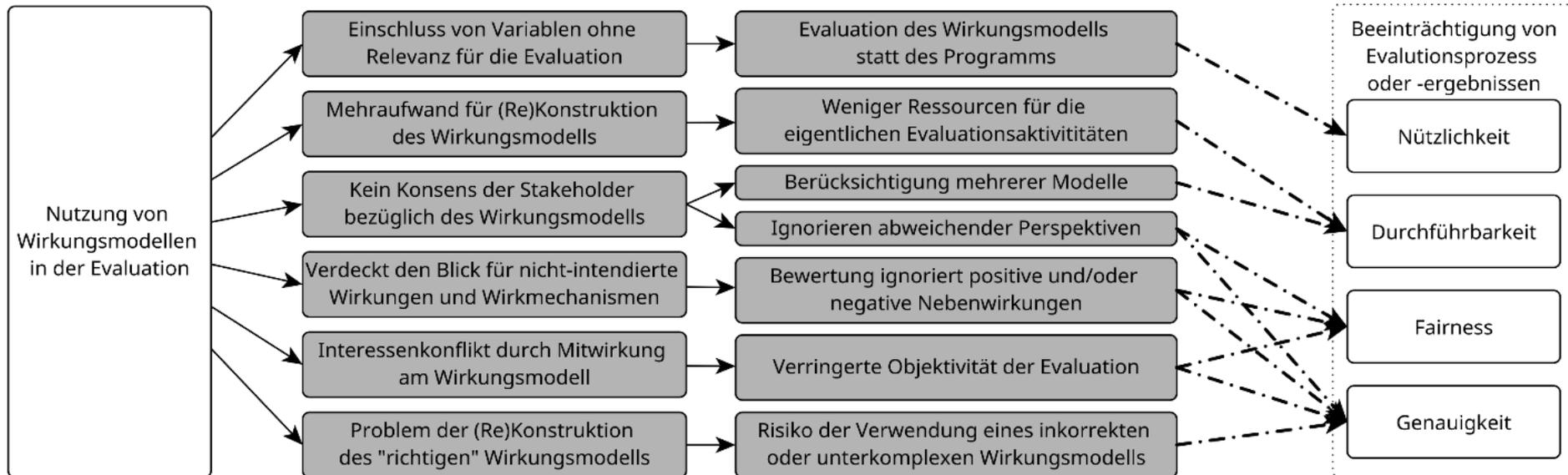
Hense (2023) – Keynote DeGEval AK Methoden



Kritik an der Verwendung von Wirkungsmodellen

Daniel L. Stufflebeam

Michael Scriven



(Hense & Taut, 2021)

A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009

Chris L. S. Coryn¹, Lindsay A. Noakes¹,
Carl D. Westine¹, and Daniela C. Schröter¹

American Journal of Evaluation
000(00) 1-28
© The Author(s) 2010
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1098214010389321
<http://aje.sagepub.com>



„In many of the cases reviewed, the explication of a program theory unmistakably was **unnecessary**, or almost an **afterthought** in some instances, and was **not visibly used** in any meaningful way for formulating or prioritizing evaluation questions nor for conceptualizing, designing, conducting, interpreting, or applying the evaluation reported.“ (p. 216)

Kritik der „new kids on the block“

z. B. Schmitt & Beach (2015). *The contribution of **process tracing** to theory-based evaluations of complex aid instruments et al.*

„In practical TBE case studies, mechanisms are often not explicitly theorized.“ (p. 432)

Programmtheorie = logisches Modell?

How We Model Matters: A Manifesto for the Next Generation of Program Theorizing

Sebastian Lemire

University of California, Los Angeles

Jane Whynot

University of Ottawa

Steve Montague

Performance Management Network

„From our viewpoint, striking changes are called for in program theorizing for it to **rise to its potential**.

For far too long, program theorizing in evaluation has been characterized and depressed by **linear, overly simplistic program depictions** [...].

The all-too common mechanistic and perhaps ritualistic development of **less than inspired program theories**, followed by limited, if any, attention in subsequent program implementation, has resulted in **piles of half-implemented or even unused program theories**. “

(Lemire et al., 2019, p. 415f)

How We Model Matters: A **Manifesto** for the Next Generation of Program Theorizing

Sebastian Lemire

University of California, Los Angeles

Jane Whynot

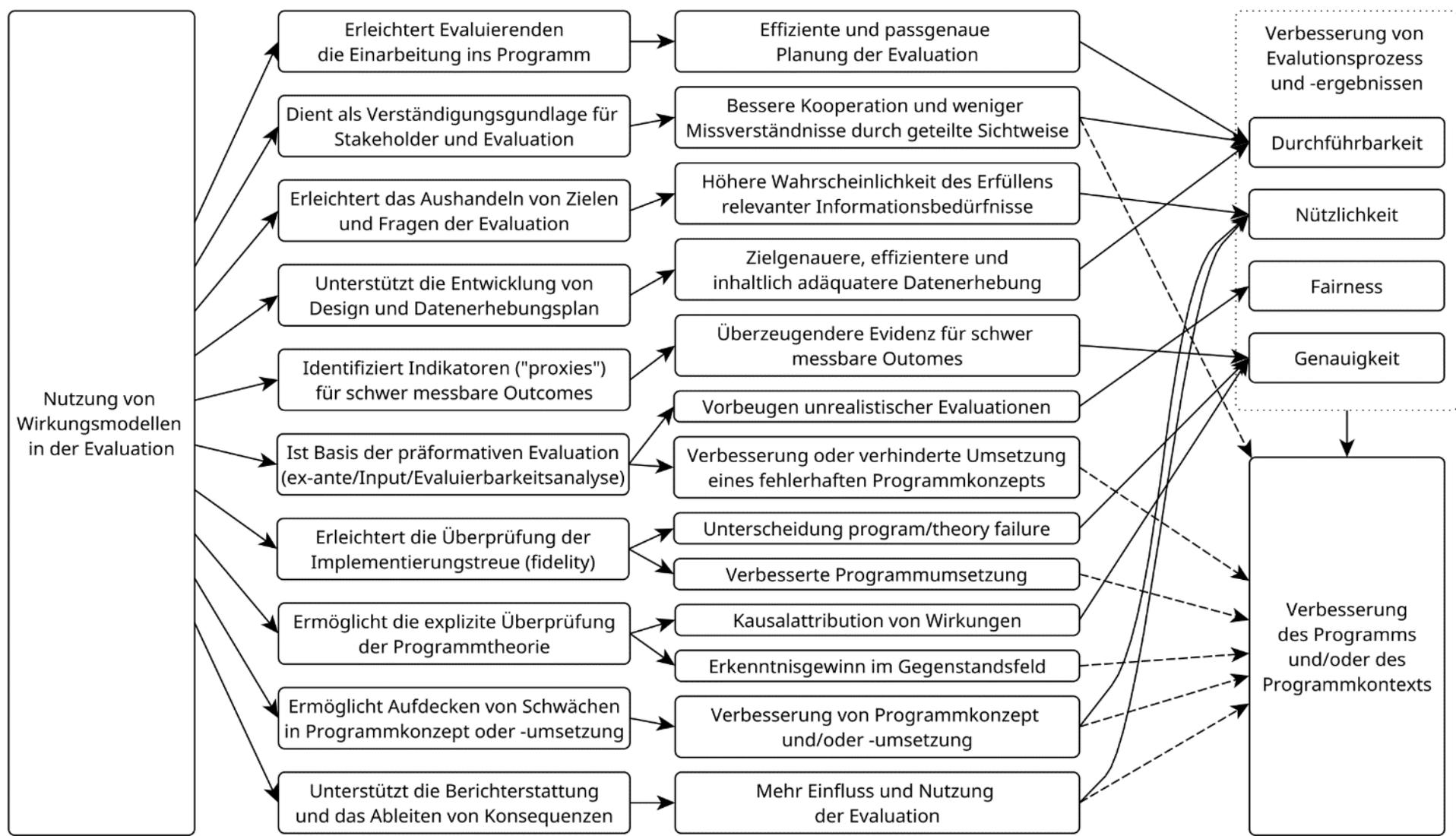
University of Ottawa

Steve Montague

Performance Management Network

1. Promote inclusion and representation
2. Strengthen the linkage between evaluation and decision making
3. Develop situated, contextualized, systems-oriented models
4. Focus on program archetypes
5. Explore different types of causal explanations
6. Flip, stack, and layer program theories
7. Put “theory” back into program theory
8. Strengthen the testability of program theories
9. Pursue theory-based synthesis and accumulation of knowledge
10. Do more with less, more quickly

**... oder sind wir doch zu ~~dumm~~ wenig
routiniert, um die edlen Kleider zu
sehen?**



(Hense & Taut, 2021)

Wo wir noch hinter den Möglichkeiten zurückbleiben

1. Wir trivialisieren Theoriearbeit.
2. Wir schätzen visuelle Sprache zu gering.
3. Wir fragen zu wenig nach dem Verwendungszweck.
4. Wir befassen uns zu oft mit Nabelschau.
5. Wir lassen die Praxis zurück.
6. Wir fangen zu oft (wieder) bei Null an.
7. Wir brauchen mehr und bessere Theorie.

(1) Wir trivialisieren Theoriearbeit

Theoriearbeit ist nicht immer trivial, wir tun aber manchmal so als ob sie es wäre.

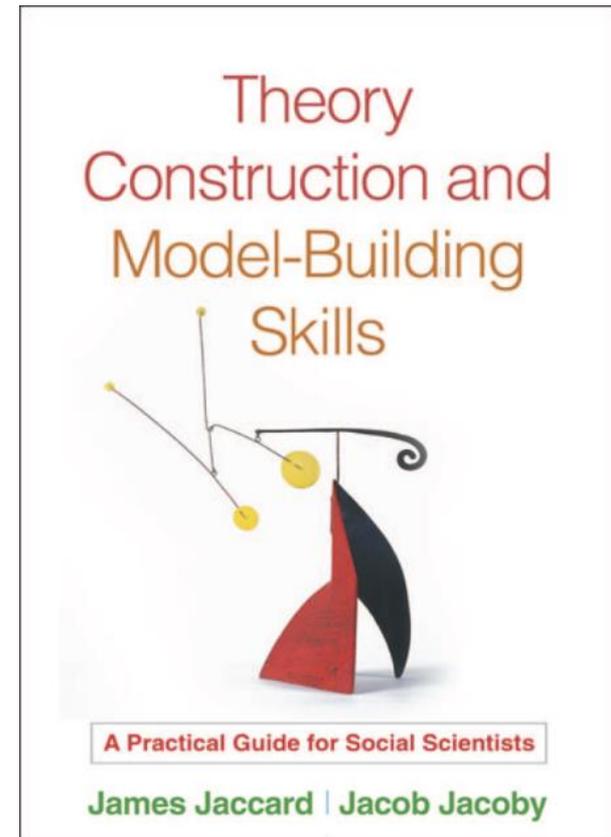
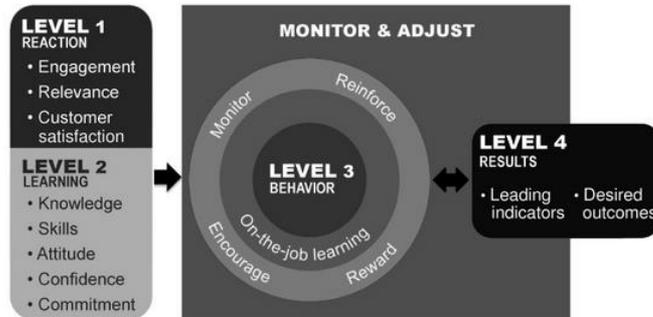
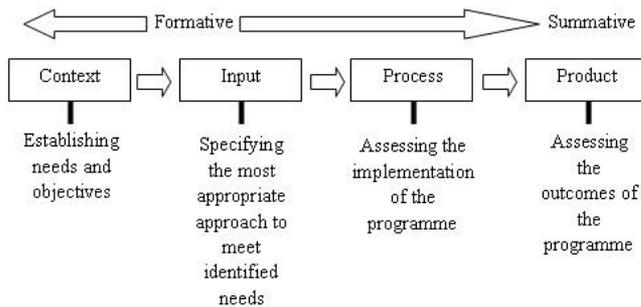
Exercise 1 uses the Basic Logic Model Development Template



*W.K. Kellogg Foundation
Logic Model Development Guide*

Resources	Activities	Outputs	Short- & Long-Term Outcomes	Impact
<i>In order to accomplish our set of activities we will need the following:</i>	<i>In order to address our problem or asset we will conduct the following activities:</i>	<i>We expect that once completed or underway these activities will produce the following evidence of service delivery:</i>	<i>We expect that if completed or ongoing these activities will lead to the following changes in 1–3 then 4–6 years:</i>	<i>We expect that if completed these activities will lead to the following changes in 7–10 years:</i>

Theorie ≠ Kästchen + Pfeile



Jaccard & Jacoby (2010)

(2) Wir schätzen visuelle Sprache zu gering

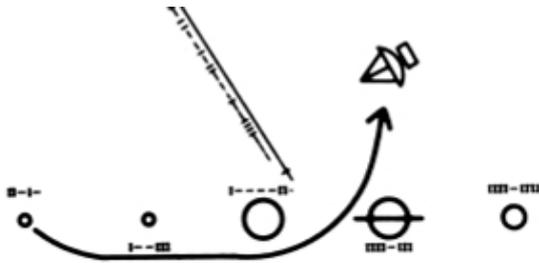
Sprachkompetenz vs. visuelle Kompetenz

Wahl sprachlichen Mittel vs. bildlicher Mittel

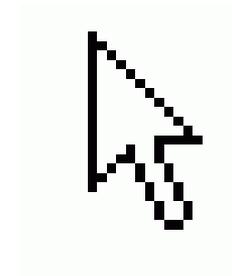
Semiotik und Syntax von Modellen und ihren
Komponenten

Semiotik des Pfeils

Bewegung



Zeigen



Richtungshinweis

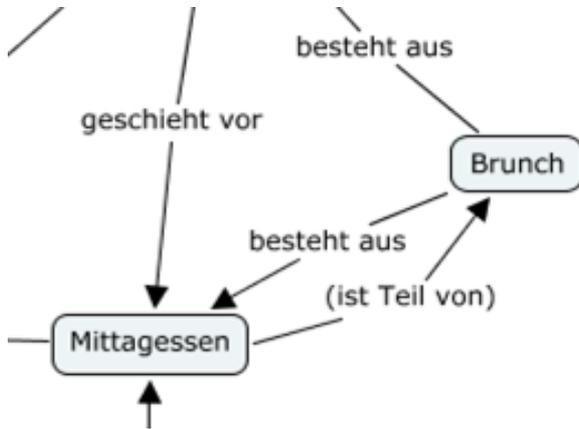


Transformation



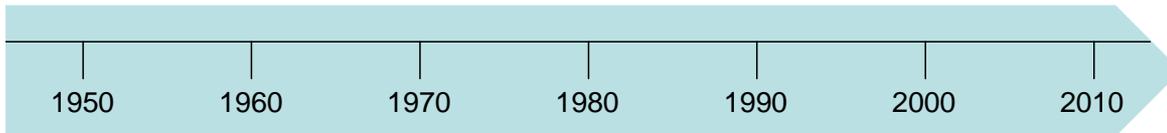
vgl. Storrer & Wyss (2003)

Semiotik des Pfeils

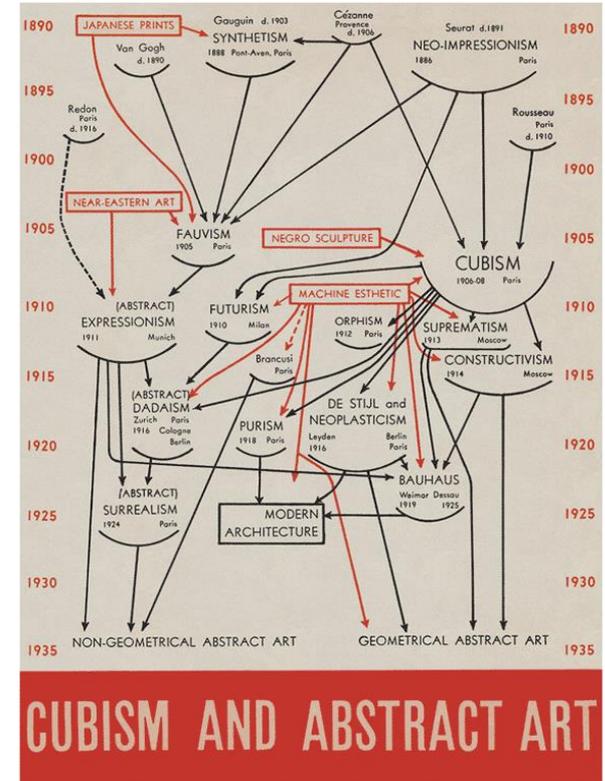


Richtung
einer
Beziehung

Zeitliche Abfolge



Einflüsse

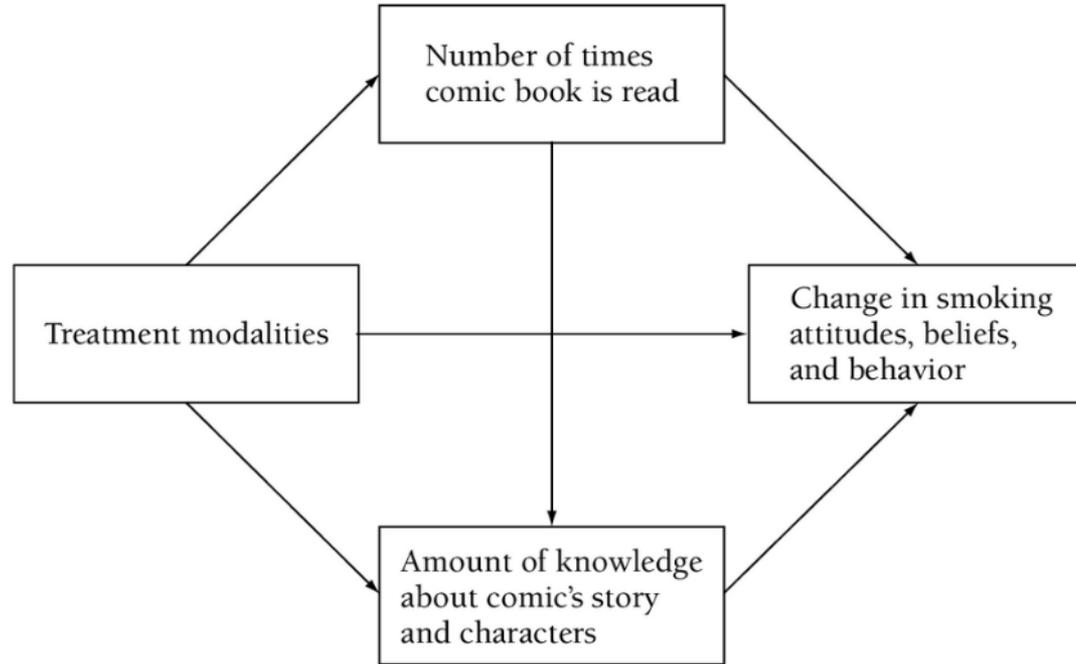


Part	X: Financial and non-financial inputs of budget support related to oversight bodies are implemented		
	⇓	⇓	⇓
1	Donors and government establish a trustful <i>policy dialogue</i>	Donors apply <i>conditionality</i> to better integrate oversight bodies into the budget process	Donors provide technical assistance to the staff of oversight bodies to improve capacities
	⇓	⇓	⇓
2	Government provides budget information to facilitate dialogue with donors and send a cooperative signal to secure (additional) financial flows in the preferred form of budget support ⇓ The information is shared with different stakeholders (parliament, civil society organizations, media) ⇓	Government complies with conditionalities (i.e. improves the formal status of oversight bodies) in order to secure committed budget support funds	Oversight bodies acquire knowledge relevant to their work in budget scrutiny
3	The budget process is more transparent (policy dialogue), the formal role of oversight bodies is improved (<i>conditionality</i>) and their capacity to oversee budget processes is increased (technical assistance).→		
4	Better formally integrated and trained oversight bodies increasingly request information on budget issues from the government.→		
5	Driven also by budget support donors, the government provides more and high quality information on budget issues.→		
6	Oversight bodies conduct valuable & solid analysis of budget related issues and provides policy recommendations to the government.→		
7	Government positively reacts to these recommendations and changes its policies.→		
	Outcome (Y): The links between government and oversight bodies are strengthened in terms of budget scrutiny.		

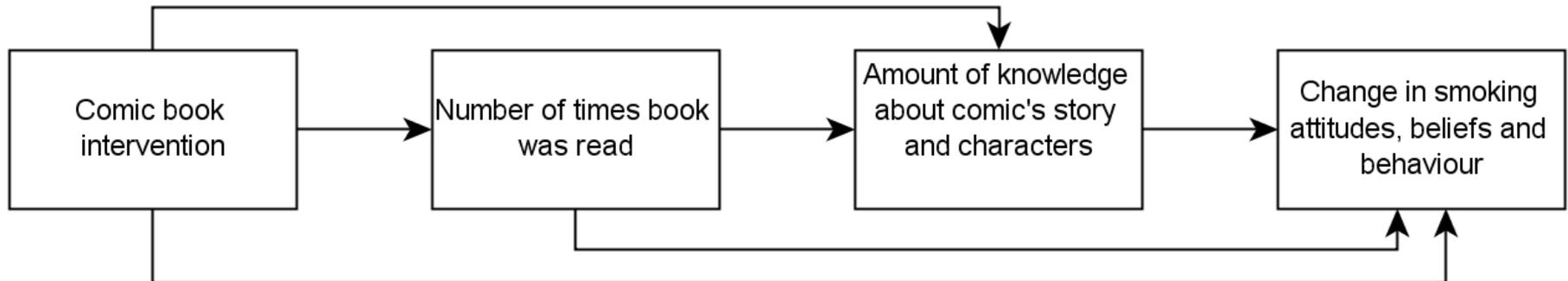


→ MODELLE AUFRÄUMEN

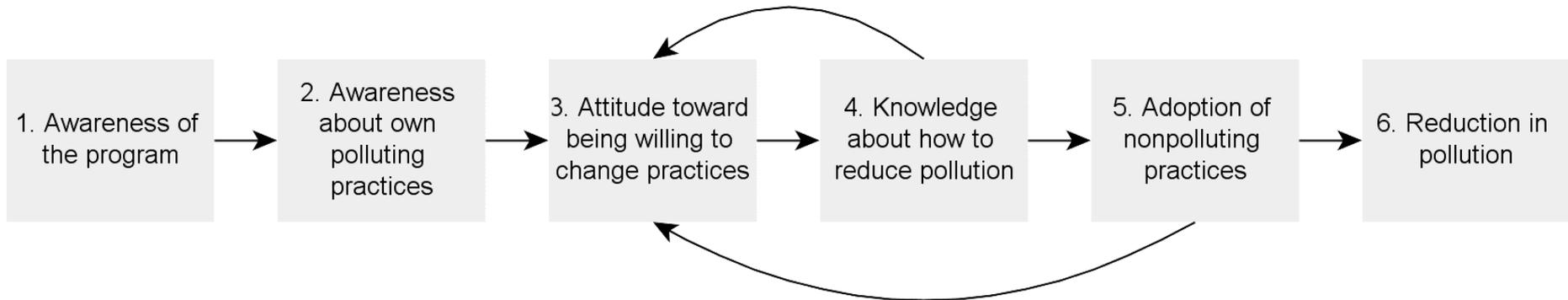
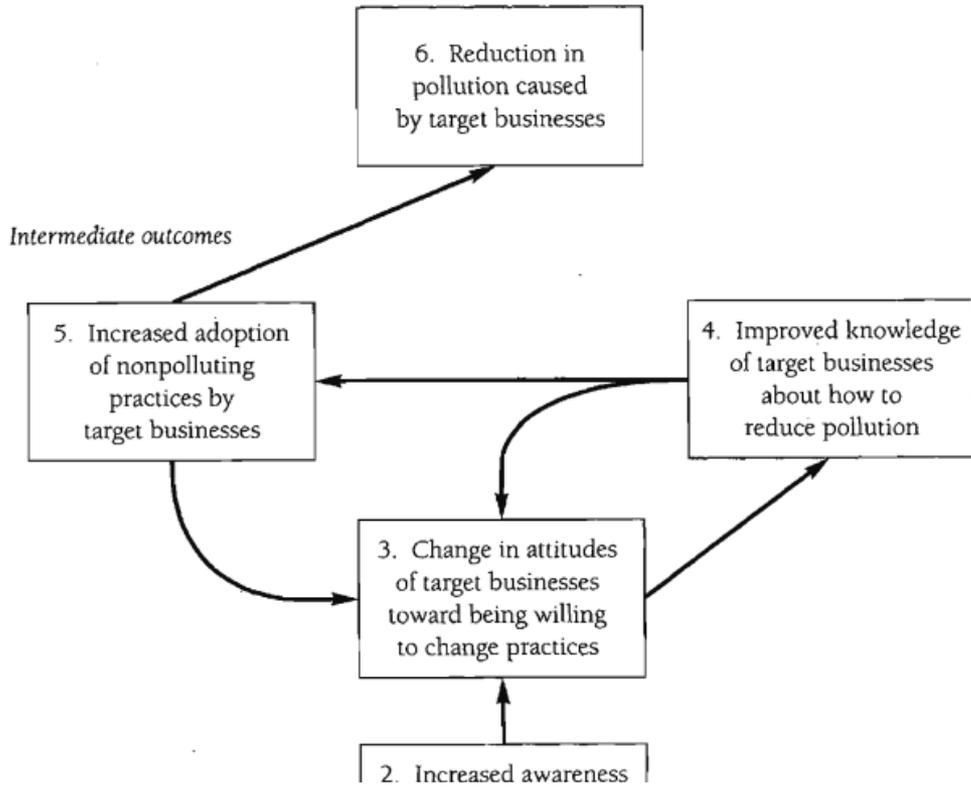
Figure 5.2. Theory of an Antismoking Program

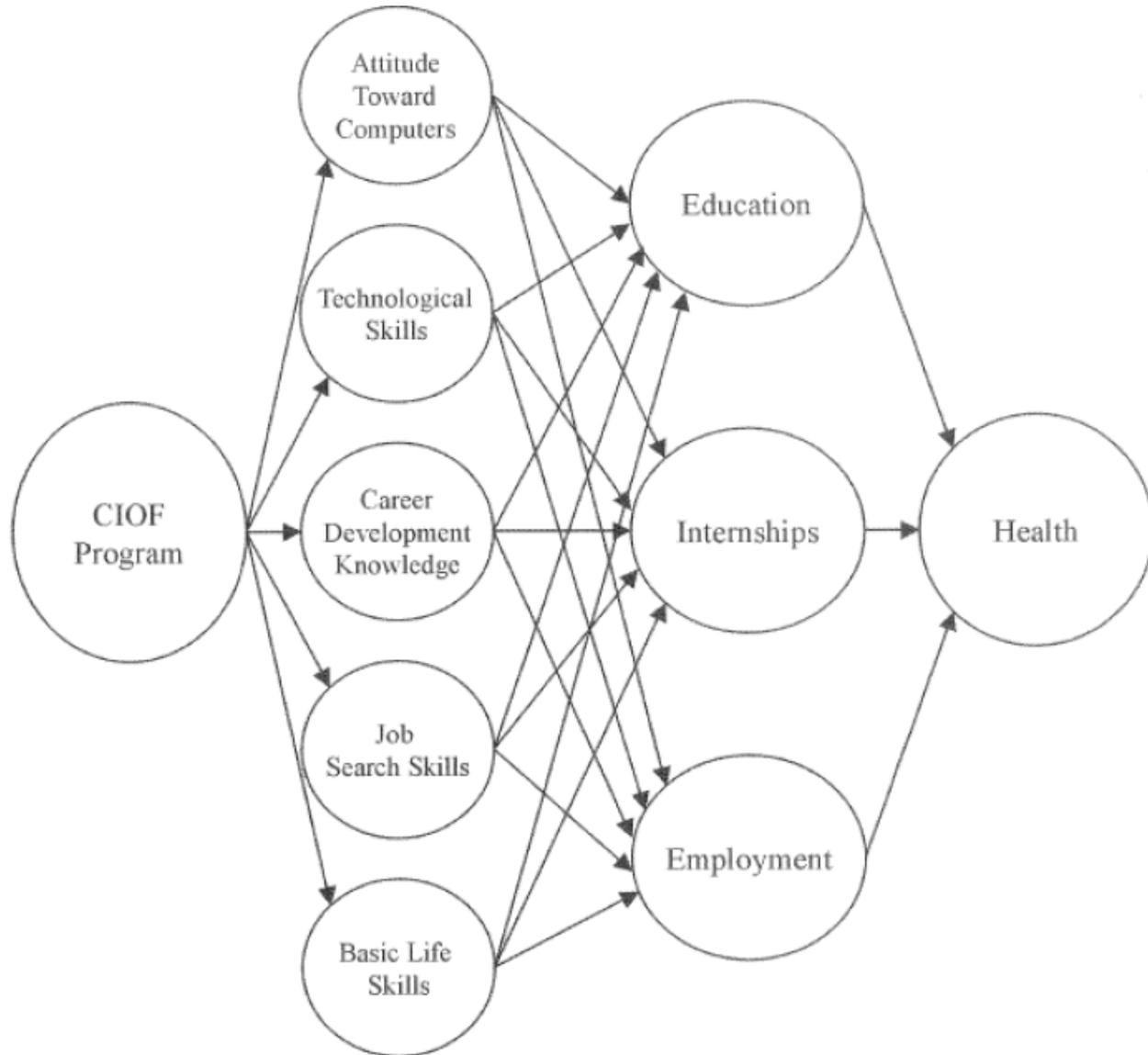


Chen (1990)

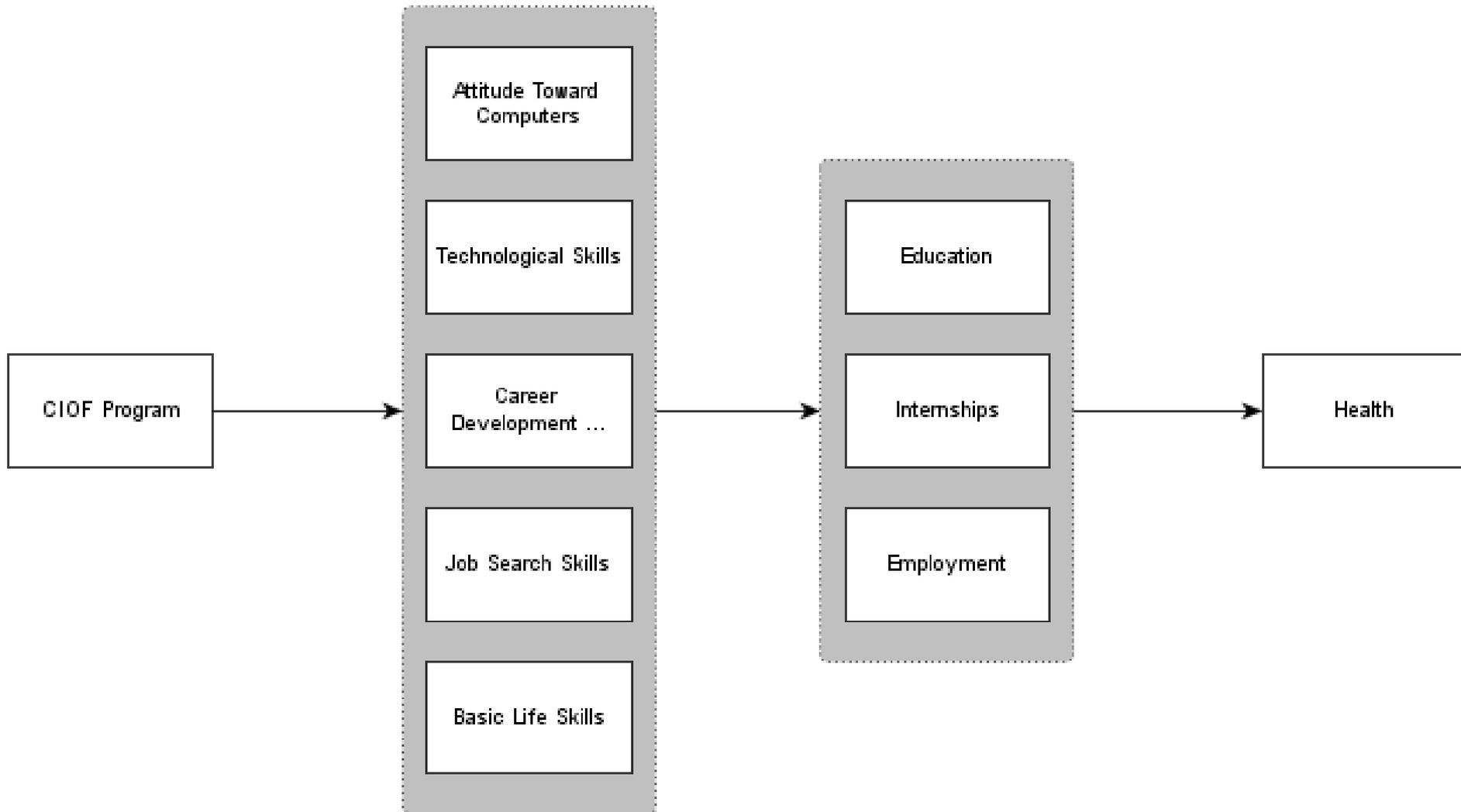


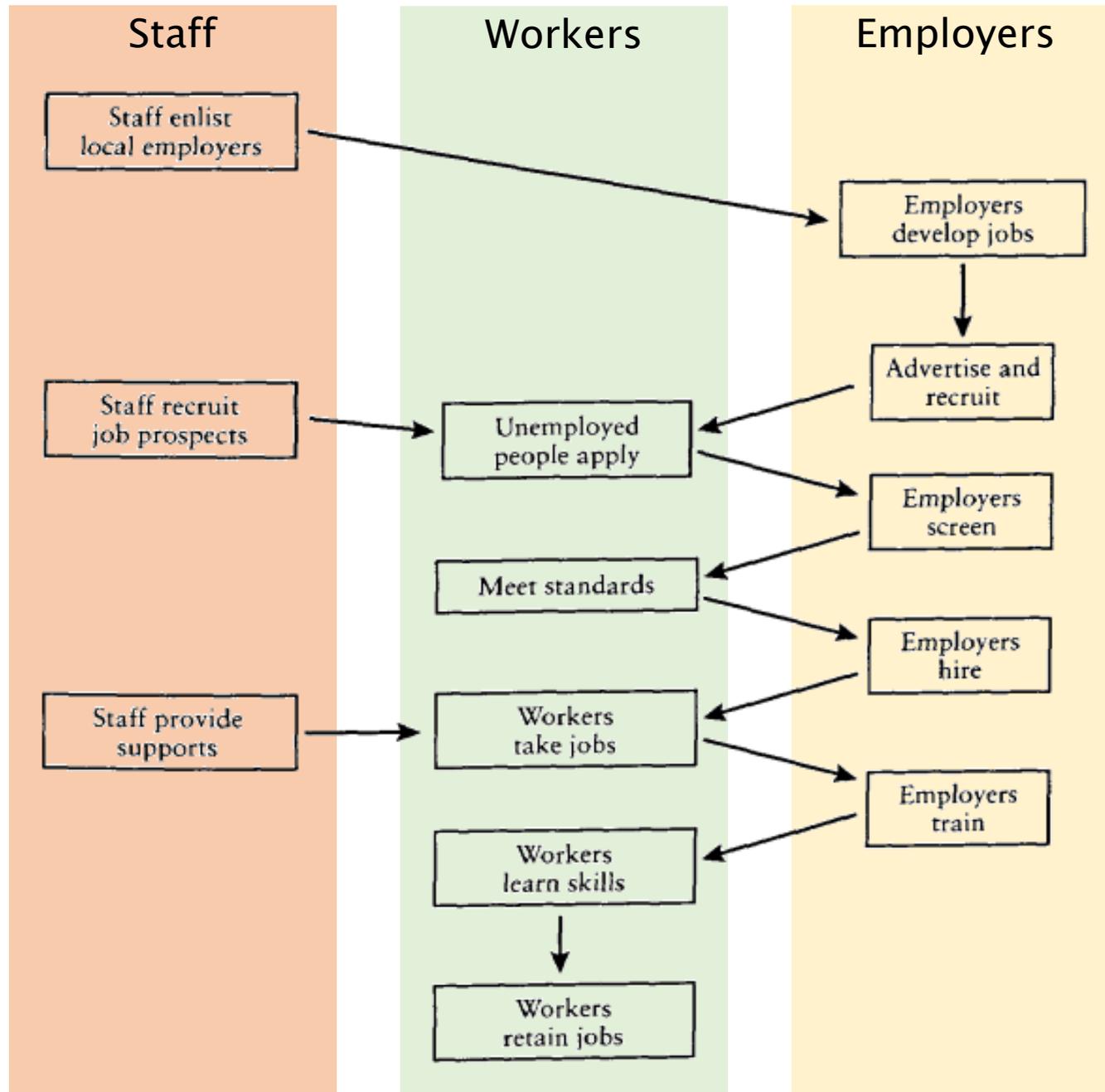
Funnell (2000, p. 93)





Donaldson (2007, p. 215)





(Weiss, 1998)

(3) Wir fragen zu wenig nach dem Verwendungszweck

„Was ist das richtige Modell?“



„Wie falsch darf es sein für den Zweck,
den wir damit verfolgen?“

La trahison des images (René Magritte, 1929)

"Essentially, all models are wrong, but some are useful"

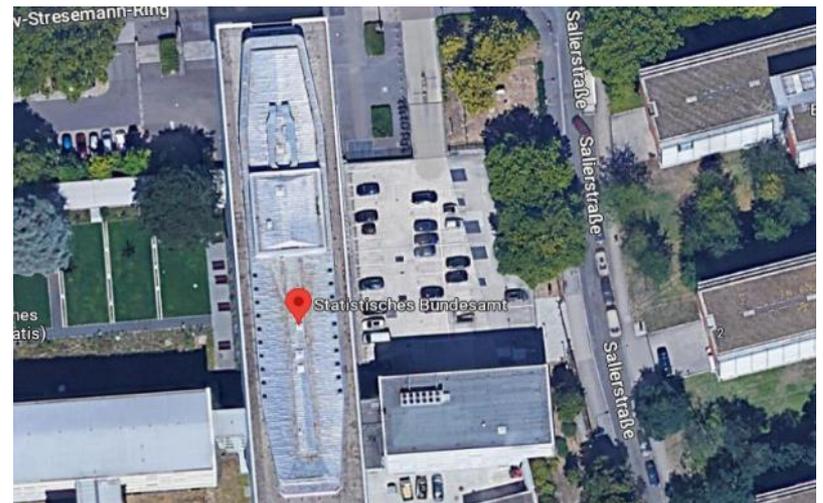
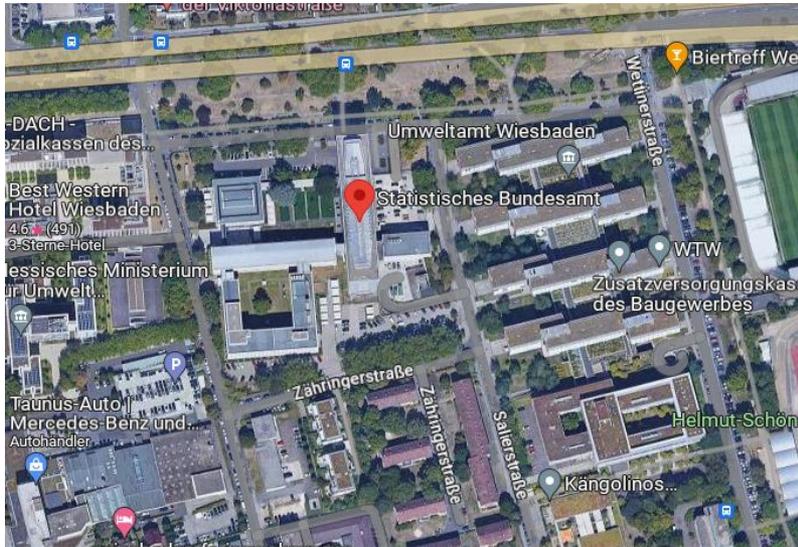
„... the practical question is how wrong do they have to be to not be useful"

Box & Draper (1987)

Modelle sind gleichzeitig

- bewusste **Abstraktionen** von der "Realität" und
- **Konstruktionen**, da sie von der Sichtweise der Konstruierenden geprägt sind.

Form follows function

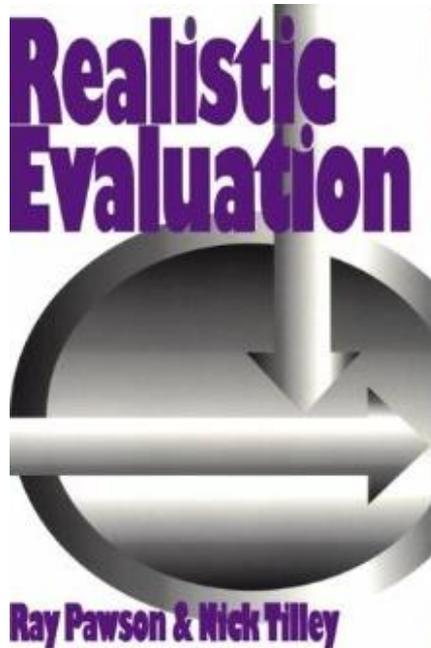


(4) Wir befassen uns zu oft mit Nabelschau.

Was ist „Programmtheorie“, „theory of change“, „change theory“, „intervention logic“, „logic model“, „program model“, „program logic“ ... ?

Was ist „theory based evaluation“, „theory driven evaluation“, „theory-oriented evaluation“, „process tracing“, „contribution analysis“ ... ?

Was ist ein „Mechanismus“?



Evaluation research 1963–1997

“Must do better. Too easily distracted by silly ideas. Ought to have a clearer sense of **priorities** and to work more systematically to see them through. Will yet go on to do great things.”

(Pawson & Tilley, 1997, p. 28)

(5) Wir lassen die Praxis zurück

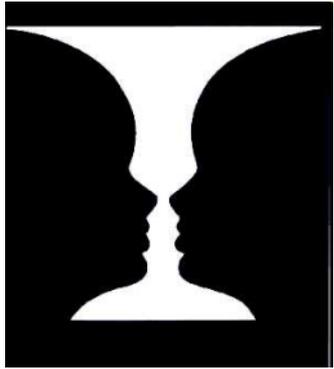
Werkstattssprache vs. Vermittlungssprache

- Was meint „Wirkungsmodell“?
 - Was meint „Wirkung“?
 - Was meint „Input“, „Output“, „Outcome“ ...?
- Hohe Anmutungsqualität des Wirkungsjargons

Die Frage nach der Interventionslogik als doppelte Provokation

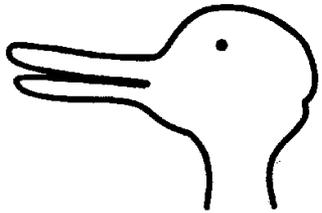
Die Macht der Bilder / Wirkungsmodelle als Vexierbilder

Programmbilder als Vexierbilder (Thomas Widmer)



Was gesehen wird, hängt ab

- von den Betrachtenden und deren Vorkenntnissen, Vorannahmen, Motiven und Einstellungen
- vom Kontext
- von den Erstellenden
- vom Wirkungsmodell selbst



(6) Wir fangen zu oft (wieder) bei Null an

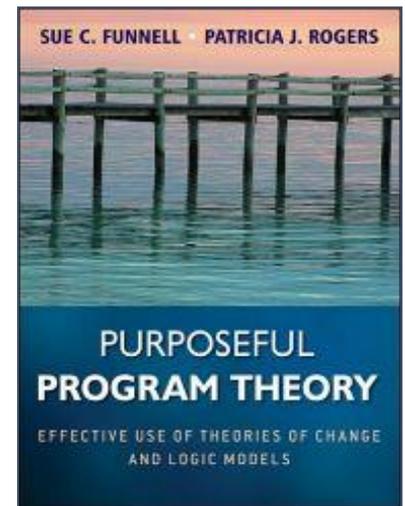
Ständige Neuentwicklungen von Wirkungsmodellen kosten wertvolle Ressourcen und verhindern Vergleichbarkeit.

→ Archetypen und Wirkungsmodelltypen als wiederkehrende Interventionslogiken (Wachsmuth, Brinkmann & Hense, 2019)

- Gemeinsame grundlegende Wirkungslogik
- Unterschiedliche Ausformungen im konkreten Anwendungsfall

Programm-Archetypen nach Funnel & Rogers (2011)

1. Bildung, Beratung und Aufklärung
2. Carrots and Sticks (Belohnung und Bestrafung)
3. Case Management (Einzelarbeit mit Klienten)
4. Community Capacity Building (Lokale Ressourcen stärken)
5. Distribution von Produkten und Dienstleistungen
(vgl. Funnell & Rogers, 2011, ch. 12)



EDITED BY

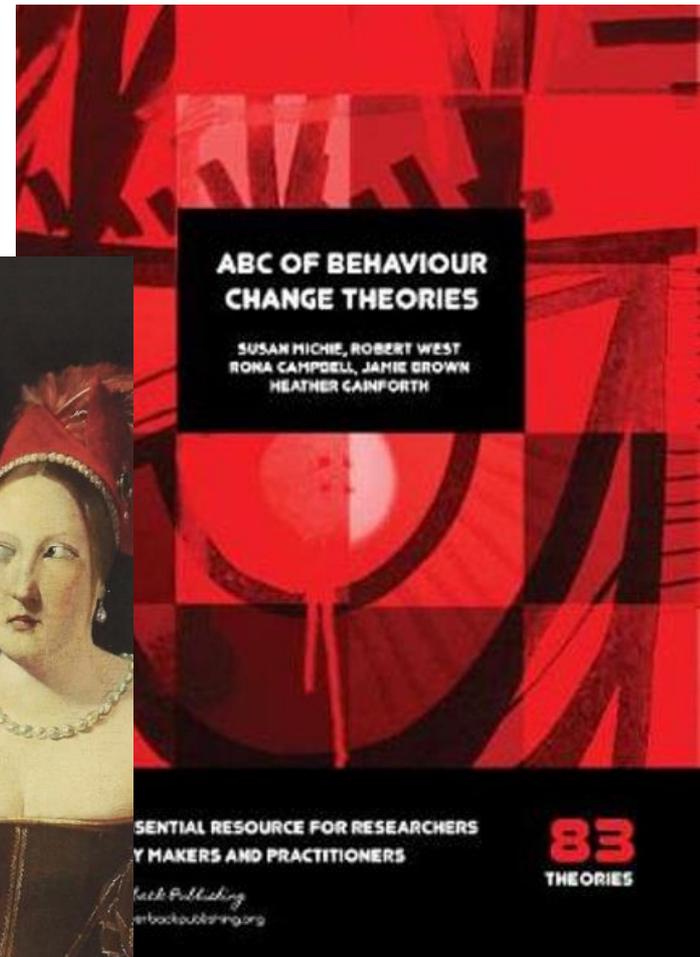
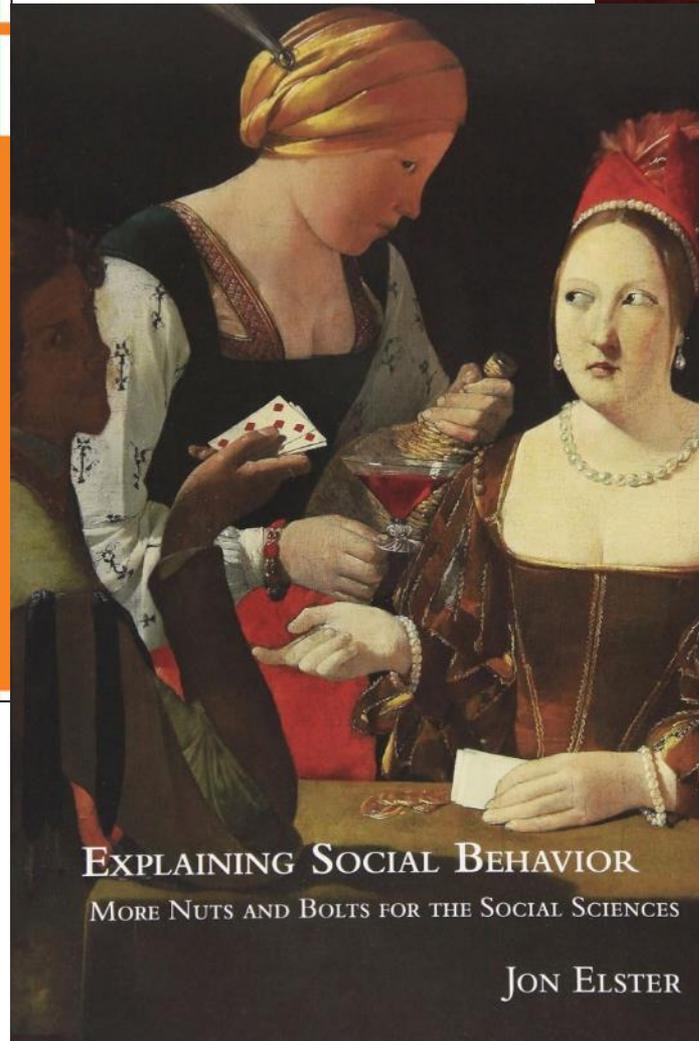
Marie-Louise
Bemelmans-Videc

Ray C. Rist

Evert Vedung

CARROTS, STICKS & SERMONS

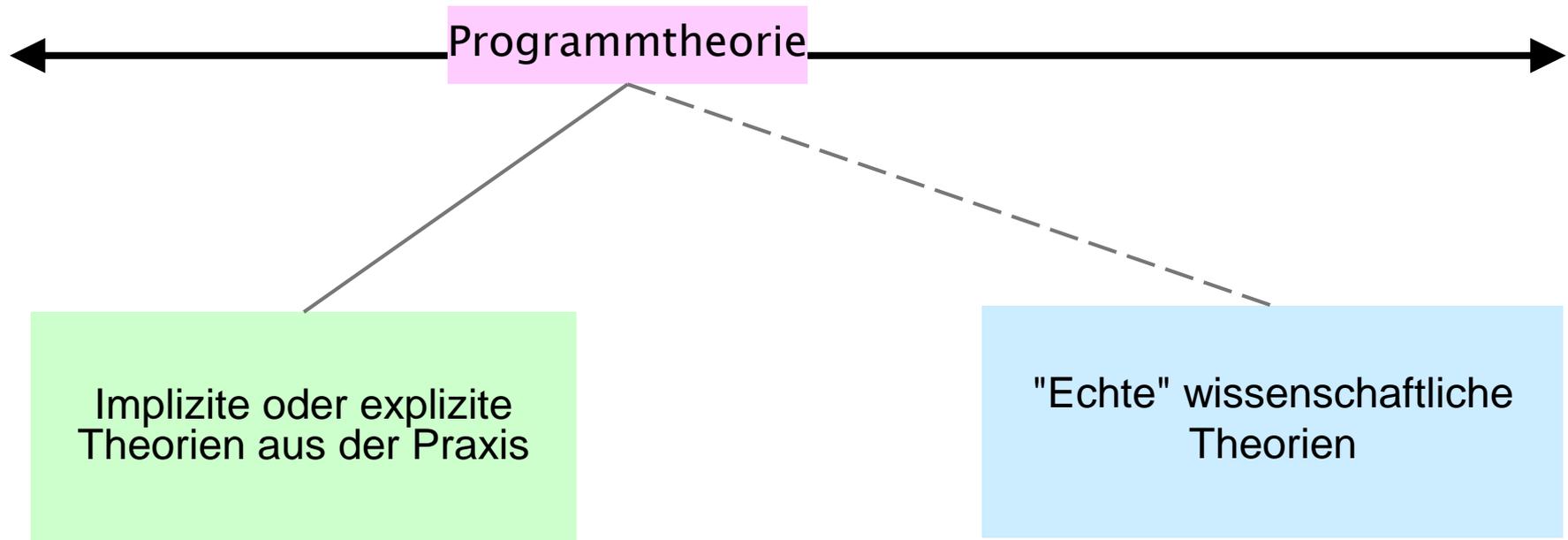
Policy Instruments
& Their Evaluation



Theorien für Programmtheorien (eine Auswahl)

- Selbstwirksamkeitstheorie
- Theory of Reasoned Action und Theory of planned behavior
- Technology Acceptance Model 2 (TAM2) und Unified Theory of Acceptance and Use of Technology (UTAUT)
- Behaviorismus (Operantes Konditionieren)
- Sozialkognitive Lerntheorie
- Problemorientiertes Lernen
- Transfer of Training
- Selbstbestimmungstheorie der Motivation
- Bildungsproduktionsmodell
- QAIT-Modell
- Health Belief Modell
- Implementation Research

(7) Wir brauchen mehr und bessere Theorie (für bestimmte Verwendungszwecke)



Welche Theorien?

Theorien, die Veränderungen auf Zielgruppenebene erklären können.

Zielgruppe Individuum

Zielgruppe Organisation

Zielgruppe Gesellschaft

Fast alle Programme wollen individuelles Verhalten beeinflussen.

→ Primat der Psychologie?

Programmtheorie quo vadis?

Die black box der Pandora ist geöffnet, einen Weg zurück zur Evaluation ohne Programmtheorie wird es nicht geben.

Kompetenzspagat: „Evaluierende“ vs. „program design and evaluation specialists“

Der Kaiser ist nicht nackt, aber wie gut er angezogen ist, liegt in unseren Händen.

Programmtheorie: “Must do better ... Will yet go on to do great things.”

