

# **Evaluierung – neue Methoden und Prozesse**

Sigrid Stagl & Peter Kaufmann  
Science and Policy Research (SPRU)

Praesentation am  
37. AWI-Seminar  
20. Juni 2007

## Ex-post Evaluierung

- Ökonometrische Methoden: propensity score matching & double difference methodology

## Ex-ante Evaluierung

- Agenten-basiertes Modellieren
- Multikriterienanalyse zur Prozesssteuerung und Involvierung von Experten/Akteuren

- Ökonometrische Methoden: propensity score matching & double difference Methoden

# Qualitätsprobleme bei Programm- evaluationen

- **Mangel an verlässlichen Daten.**
  - Monitoring nicht ausreichend mitbedacht
- **Mangel an geeigneten Methoden.**
  - Auseinanderhalten von Programmeffekten und externen wirtschaftlichen Entwicklungen -> meist qualitative Bewertung
- **Zu wenig Zeit und Geld.**
  - Restriktionen der Geldgeber erlaubt nicht den Einsatz von differenzierteren Methoden (daher auch ein Mangel an Wissensakkumulation)

Quelle: CEAS, 2003; Forstner and Plankl, 2004

# Problematik fuer Evaluatoren von laendlichen Programmen

- Formulierung der Null-Option
  - Erfordert Vergleich von gefoerderten mit nicht-gefoerderten Subjekten (-> Matching mit gleichen Charakteristika)
- Vermeidung einer Verzerrung bei der Auswahl der Kontrollgruppe (selection bias)
  - Programme sind auf bestimmte Zielgruppen gemuenzt
- Auseinanderhalten von Programmeffekten und exogenen Effekten
- Bewertung nicht-gehandelter Gueter (Umwelt, Infrastruktur, laendliche Kultur)
- Methoden fuer die Bewertung von Nebeneffekten

- **Deadweight loss** (Nettowohlfahrtsverlust. Waere auch ohne Foerderung passiert)
- **Verdraengungseffekt** (Ressourcen werden nur zwischen Regionen verschoben, z.B. Beschaeftigung)
- **Substitutionseffekt** (zum Nachteil von nicht-foerderbaren Subjekten, z.B. Weizen- vs. Roggenproduzenten)
- **Nebeneffekte**, die nicht spez. in Programmzielen erfasst wurden.

# Propensity Score Matching

1. Annahme: alle relevanten Unterschiede zw. den beiden Gruppen wird durch die beobachtbaren Variablen beschrieben.
  2. Kontrollgruppe, in der die Verteilung dieser Variablen so aehnlich wie moeglich der Foerdergruppe ist.
  3. Passende Subjekte (gefoerdert und nicht-gefoerdert) werden zusammengespannt.
  4. Der 'average treatment effect' (ATE) wird berechnet.
- => Propensity score ist die geschaeetzte Wahrscheinlichkeit der Teilnahme am Foerderprogramm

# Wie wird der 'average-treatment effect' kalkuliert (double difference)?

- Noetig sind Daten beider Gruppen zu Beginn und am Ende der Foerderzeit.
- Zu jedem gefoerderten Subjekt  $i$  wird eine Gruppe von 'vergleichbaren' nicht gefoerderten Subjekten gesucht, und dann wird...
  - das Ergebnis des gefoerderten Subjekts  $i$ , mit den Ergebnissen der 'Nachbar-subjekte'  $j$  in der Kontrollgruppe verglichen;
  - Gewichte fuer die nicht-gefoerderten Subjekte  $j$  sind in Relation zur Aehnlichkeit der beobachtbaren Variabeln in  $i$  und  $j$ .
  - Dies wird fuer Haupteffekte sowie fuer Nebeneffekte durchgefuehrt, und der Nettoeffekt berechnet.



- Bestehende Evaluierungspraxis basiert auf Methoden, die systematische Verzerrung der Ergebnisse toleriert.
- Es gibt eine Möglichkeit, dies zu vermindern, aber nur mit Investitionen in Monitoringsysteme. Dann wären auch differenzierte ökonomische Evaluierungen mit relativ geringerem Aufwand möglich.
- Herausforderung: Wissen über diese Methoden bei Praktikern und Evaluatoren zu verbreiten.

- **Agenten-basiertes Modellieren**
- Multikriterienanalyse zur Prozesssteuerung und Involvierung von Experten/Akteuren

# Was ist agenten-basiertes Modellieren?



University of Sussex

---

- ‘Agenten’ (Software Objekte) interagieren in einem Computer
  - Autonomes Verhalten (wie Roboter)
  - Interaktionen zwischen Agenten ueber die Zeit verursachen Dynamik
  - Die representierten Systeme (Oekonomien, Staedte, Firmen, Kulturen, Oekosystems, etc.) adaptieren, da Agenten lernfaehig sind
  - Agenten koennen auf unterschiedlichen Ebenen, in verschiedenen Settings etabliert werden
- Strategisches Verhalten und institutionelle bzw. Makroeinfluesse koennen entweder explizit, oder mittels Verhaltensregeln eingefuehrt werden. -> Bottom-up

- Praeferenzen koennen sich veraendern (exogen oder endogen).
- Individuelles Verhalten is beeinflusst von der persoenlichen Geschichte, Eigenheiten (e.g. Habitus), dem Verhalten anderer, und Institutionen.
- Verhalten von Individuen auf aggregierter Ebene kann oft nicht ganz durch die individuellen Charakteristika erkluert werden (Emergenz).

# Wie implementieren?

## **Konstruieren einer virtuellen Welt**

- Bevoelkern mit verschiedenen Agententypen mit z.B. Werte, Handlungsweisen, Lernmechanismen, Unsicherheiten
- Agenten mit Ressourcen ausstatten – verschiedene Arten von Kapital und sozialen Netzwerken

**Anfangsbedingungen festsetzen** – aus Experimenten und der Realitaet

**Interaktionen/Aktivitaeten** – Zuruecklehnen und warten was passiert

**Ergebnisse** – Muster von Handlungsweisen, die auch auf Individuen zurueckverfolgt werden koennen.

Programm mehrmals laufen lassen um Sensitivitaetstests durchzufuehren, z.B. fuer konsistente Reaktionen auf Politikaenderungen

# ABM als interdisziplinäres Modellierungswerkzeug



---

Etablierte Methode in Biologie

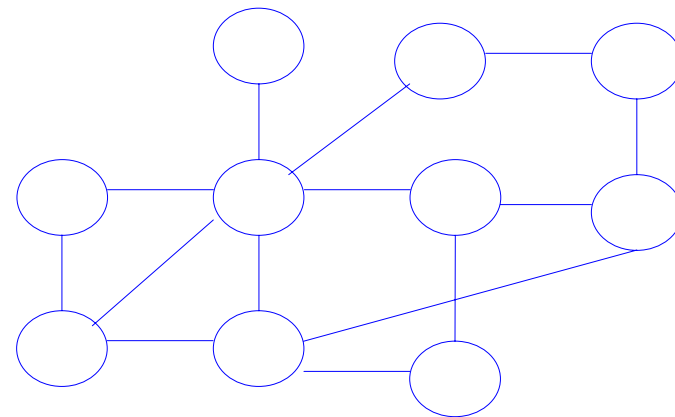
Zunehmend in Sozialwissenschaften angewendet

- Geographie: Wachstum von Städten, Fußgängerströme, etc. können räumlich explizit modelliert werden

Volkswirtschaft: interagierende Agenten in einem gegebenen Kontext mit Lernfähigkeiten erklären Marktdynamiken

- Verhalten bei Auktionen; technologischer Wandel & Wachstum, etc.
- Evolution von Normen; sozialer Einfluss, z.B. von Arbeitsmarktnetzwerken
- Veränderungen von Landnutzung; Technologiediffusion in der Landwirtschaft

# Sozialer Einfluss: Beispiel

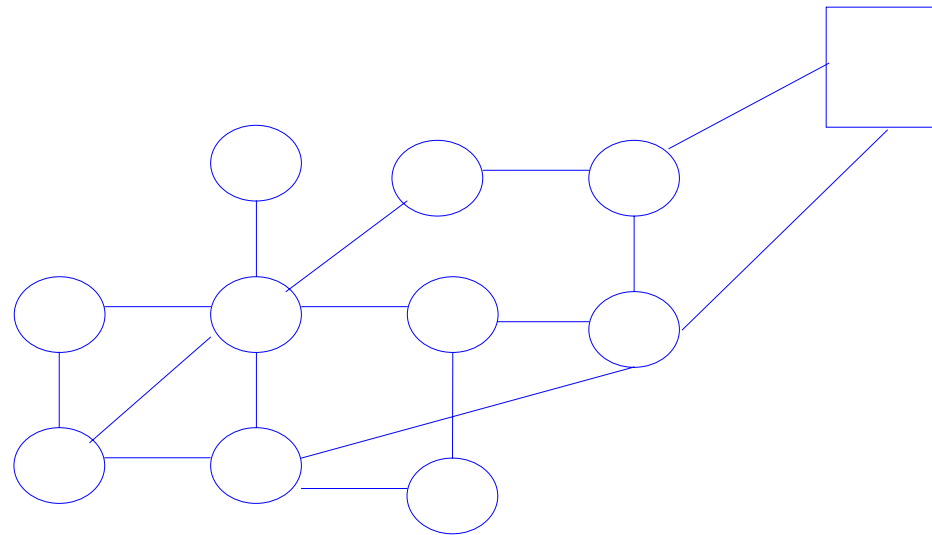


Soziales Netzwerk



# Ebenen der Darstellung

- Individuen
- Soziale Gruppen
- Institutionen
- Länder



Heterogenität aufgrund verschiedener Agententypen innerhalb einer Ebene und auf verschiedenen Ebenen, z.B. Institutionen interagieren mit Landwirten



# „Emergenz“

„Komplexe Muster auf Systemebene resultieren aus Handlungen auf einer niedrigeren Ebene, koennen aber nicht auf diese reduziert werden.“

In ABM koennen sich komplexe Muster auf Systemebene aufgrund von diskreten (nicht stetiger) und expliziten Interaktionen und Handlungen auf Agentenebene ergeben. Beispiel: Verbreitungsmuster einer Technologie. Reaktionen auf Schocks.

# Analytische vs. Agenten-Basierte Modelle

## Analytische Modelle

- Vorteile: Transparenz der Ergebnisse; Allgemeinguetigkeit; Annahmen expliziter.
- Nachteile: Annahmen koennen uebermaessig beschraenkend sein; ignorieren Individuen; emergentes Verhalten schwierig abzubilden; oft nicht leicht verstaendlich.

## ABM

- Vorteile: Flexibel; Fokus auf mehrere Ebenen; emergentes Verhalten; Heterogenitaet; fuer Nichtmathematiker leichter verstaendlich; Agenten koennen mit Verhaltensweisen ausgestattet werden.
- Nachteile: Ergebnisse koennen etwas wage sein; weniger allgemein gueltig, da potentiell einige Spezifika eingebaut; passiert leicht dass zuviel Komplexitaet eingebaut wird; manchmal schwer zu analysieren.

# So einfach wie moeglich halten

Zweck: Ein System zu *verstehen*. Es geht nicht um Vorhersage.

- Zu komplexe Modelle erschweren dies.

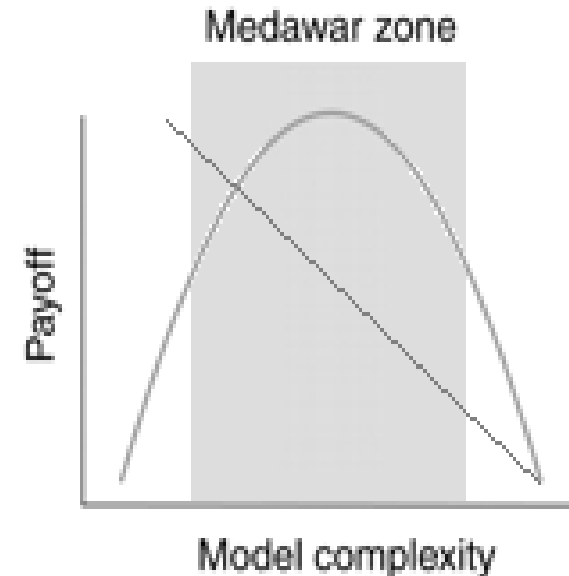
Analytisch + Simulation

= transparent und rigoros

Einfach beginnen und langsam Komplexitaet integrieren.

Modelle sind nicht die Realitaet

- Hinterfrage immer die zugrundeliegenden Annahmen



# ABM als Ergaenzung zu bestehenden Methoden

Heterogene Agenten, die in Netzwerken kooperieren  
beeinflussen die lokale Wirtschaft

- Landwirtschaftliche und nicht-landwirtschaftliche Firmen
- Verschiedene Arten von Haushalten
- Regierungs- und gesetzgebende Institutionen
- Interessensgruppen, NGOs, politische Parteien

Interaktionen fuehren zu Dynamiken, bestimmen das Niveau  
des Ergebnisses

# ABM als Ergaenzung zu bestehenden Methoden

- Verschiedene Netzwerkstrukturen (lokal und darueber hinaus) fuehren zu unterschiedlichen Transaktionskosten
- Wissen, Werte und Investitionsentscheidungen sind durch die Struktur der Informations- und Kommunikationsnetzwerke beeinflusst
- Durch das Einfuehren von Politiknetzwerkstrukturen wird die Implementierung von F&E Massnahmen endogenisiert.

1. Ein sozial eingebettetes agenten-basiertes Modell einer lokalen Wirtschaft entwickeln.
2. Wie Agenten ihre Werte und Einstellungen bilden
3. Investitionsentscheidungen einzelner Agenten modellieren
4. Handlungen der lokalen Institutionen modellieren

Ziel: NICHT Vorhersagen zu machen, sondern Systemzusammenhaenge zu verstehen.

- Agenten-basiertes Modellieren
- **Multikriterienanalyse zur Prozesssteuerung und Involvierung von Experten/Akteuren**

Kosten-Nutzen-Analyse

Multikriterienanalyse

Partizipative Methoden



# Vergleich verschiedener Richtungen der ex-ante Evaluierung

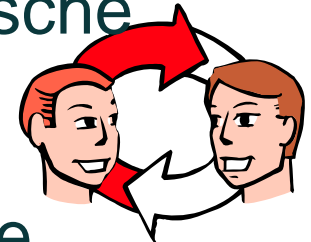
## Monetäre Bewertung

- Zahlungsbereitschaft
- Individuen als wirtschaftliche Akteure
- Gemeinschaftswerte vernachlässigt
- Konflikte werden versteckt
- Ignoriert Unsicherheit



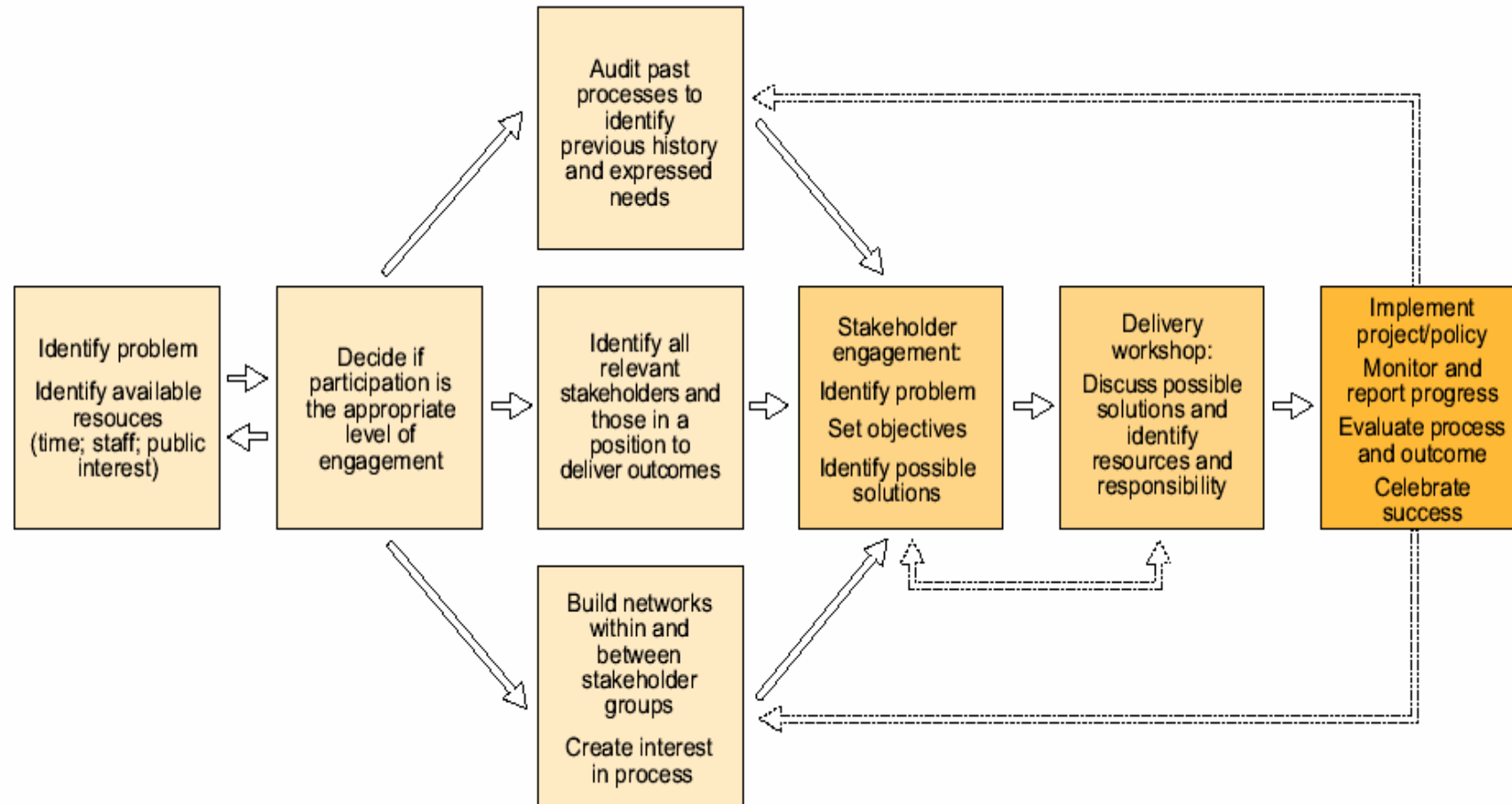
## Bewertung mittels partizipativer Methoden

- Gesprächsbereitschaft
- Individuen als politische Akteure
- Gemeinschaftswerte hervorgehoben
- Konfliktbereiche identifizieren, adressieren und überwinden
- Adressiert Unsicherheit



- Definition des Projekts
- Identifikation der Auswirkungen des Projekts
- Welche Auswirkungen sind oekonomisch relevant?
- Physische Quantifizierung der relevanten Auswirkungen
- Monetaere Bewertung der relevanten Auswirkungen – erforderlich fuer:
  - Vorhersage von Werten, die sich in Zukunft auswirken,
  - Korrektur der Marktpreise wo noetig, und
  - Berechnung von Preisen wo es keine gibt.
- Diskontieren von Kosten und Nutzen
- Kapitalwerttest anwenden  
( $NPV = \sum B_t(1+i)^{-t} - \sum C_t(1+i)^{-t}$ )
- Sensitivitaetsanalyse

# Design von partizipativen Prozessen



Quelle: Richards C, Sherlock K, Carter C (2004): Practical Approaches to Participation, SERP Policy Brief, no. 1, p.20.

Stakeholderpartizipation, beinhaltet Deliberation und Inklusion (Bloomfield et al. 2001), “can initiate social learning processes which translate uncoordinated individual actions into collective actions that support and reflect collective needs and understanding” (Webler et al. 1995).

- World Bank – verschiedene NGO-Stakeholder in der Projektevaluierung
- Rio Deklaration – Agenda 21 (1992)
- Europ. Kommission – 5. Aktionsprogramm der Umwelt (1993) & Aarhus Konvention (1998)

MKA ermöglicht es in Entscheidungen

- grosse Datenmengen, Beziehungen und Ziele,
- Kriterien, die in verschiedenen Einheiten gemessen werden,
- schwache Vergleichbarkeit von Optionen,
- wissenschaftliche Daten von verschiedenen Disziplinen, Werturteile und Interessen

zu berücksichtigen.

→ MKA erlaubt uns eine beschränkte Anzahl von Alternativen zu reihen, nachdem mehrere teilweise konfligierende Kriterien berücksichtigt wurden.

# Ein typisches MKA Problem

$A$  ... endliche Menge von  $n$  moeglichen Optionen

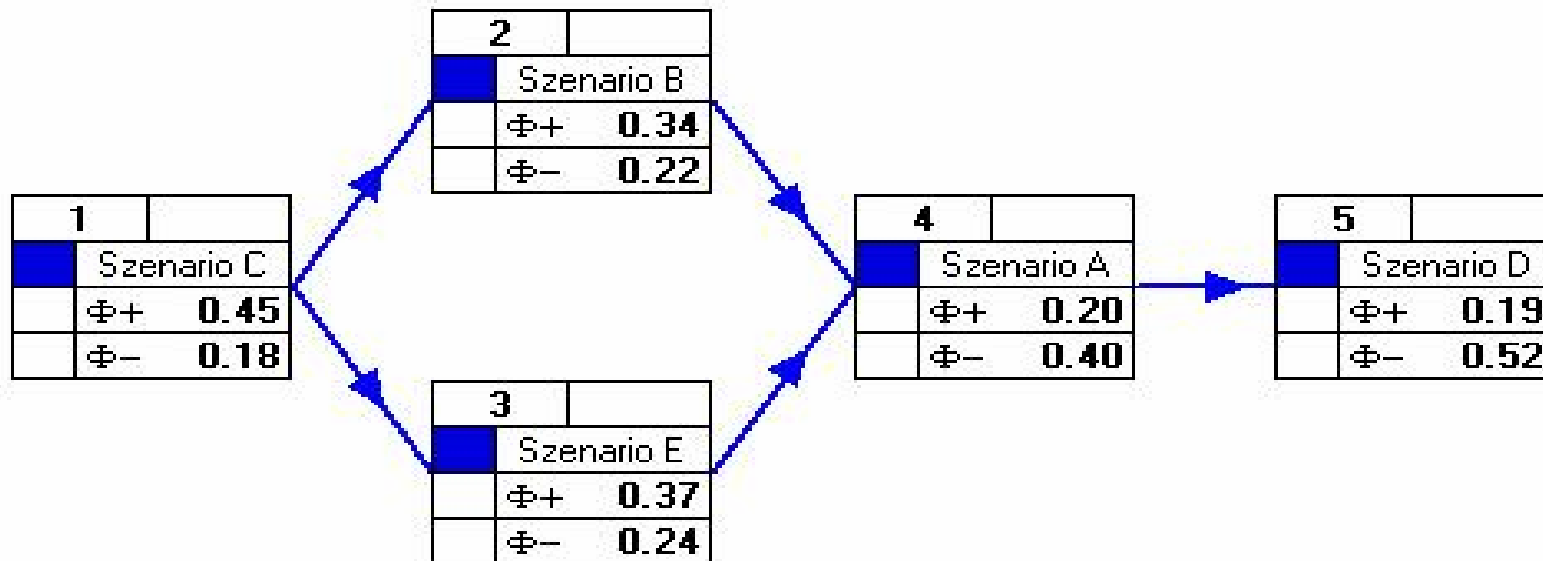
$m$  ... Anzahl der verschiedenen Gesichtspunkte oder Evaluierungskriterien  $g_i$   $i=1,2,\dots,m$ .

Option  $a$  ist besser als Option  $b$  (beide sind Teil von  $A$ )  
gemaess dem  $i$ -ten Gesichtspunkt, wenn  $g_i(a) > g_i(b)$

Auswirkungsmatrix besteht aus Mengen von  $A$  (von Alternativen) and  $G$  (von Evaluierungskriterien).

-> Keine Loesung kann alle Kriterien optimieren;  
Kompromissloesungen muessen gefunden werden.

# Exemplarische Reihung von Szenarien mittels MKA



- Evolvierende komplexe Systeme, Unsicherheit, Ambiguitaet und Ignoranz → technische Nicht-Vergleichbarkeit
  - Multiple legitime Perspektiven → soziale Nicht-Vergleichbarkeit
- Rolle von Partizipation
- Rolle von sozialem Lernen



Kosten-Nutzen-Analyse

Partizipative Methoden + Partizipative Methoden

Multikriterienanalyse

Nachhaltigkeitsbewertungsmethoden sind  
mehr als eine technische Frage,  
nämlich 'wert-artikulierende Institutionen'  
(Vatn 2005).

# Nachhaltigkeitsbewertungsmethoden als 'wert-artikulierende Institutionen'



Die Bewertungsmethoden bestimmen -

- wer im Entscheidungsprozess teilnimmt;
- wie sie teilnehmen und in welcher Rolle (KonsumentInnen, Stakeholder, BuergerInnen);
- was zaehlt als Daten;
- welche Datenverarbeitungs- und Aggregationsprozeduren werden verwendet.

Ergebnisse werden beeinflusst durch die Art der Bewertung und die institutionellen Strukturen, in die die Evaluierung eingebettet ist.

Evaluierungsmethoden unterstuetzen  
Entscheidungsfindung fuer nachhaltige Entwicklung  
besser, wenn sie –

- keine monetaere Bewertung aller Aspekte erfordern,
- sehr niedrige oder keine Diskontraten anwenden,
- soziale Lernprozesse unterstuetzen,
- Partizipation der Oeffentlichkeit und / oder von Stakeholdern ermoeglichen,
- transparent sind,
- sowohl quantitative als auch qualitative Daten nutzen koennen, und
- diese systematisch zusammenfuehren.

## Kostennutzenanalyse + Partizipation

- Deliberative monetary valuation

## Multikriterienanalyse + Partizipation

- Social multicriteria evaluation
- Three-stage multicriteria analysis
- Multicriteria mapping
- Deliberative mapping
- Stakeholder decision / dialogue analysis etc.

*Siehe auch Stagl 2007, 'Emerging Methods for Sustainability  
Valuation and Appraisal', [www.sd-research.org.uk](http://www.sd-research.org.uk)*

Hypothese 1: Das Wissen innerhalb eines Referenzrahmens veraendert sich, d.h. neue Fakten werden aufgenommen (single-loop learning).

Hypothese 2: Die Bewertung von Fakten aendert sich aufgrund geaenderter Werte und Annahmen (double-loop learning).

Hypothese 3: Neue Wege werden beschritten um mit komplexen und konfliktbeladenen Themen auf konstruktive Art umzugehen und ein Transfer dieses Wissens auf andere Themen wird ueberlegt (triple-loop learning).

# Resultate – Erneuerbare Energie in Oesterreich

## **Single-loop learning** - TeilnehmerInnen ...

- waren vertrauter mit verschiedenen erneuerbaren Energietraegern
- haben gelernt verschiedene Energiemixes und ihren relativen Beitrag zur Nachhaltigkeit zu vergleichen.

## **Double-loop learning** - TeilnehmerInnen ...

- aenderten ihre Annahmen (gaben sozialen Parametern und Werten mehr Bedeutung)
- aenderten nicht ihre Wahrnehmung ueber Komplexitaet und Unsicherheit; Wahrnehmung der Rolle zukuenftiger Generationen blieb hoch
- blieben am selben Niveau bezueglich gegenseitigem Verstaendnis

## **Triple loop learning** - TeilnehmerInnen ...

- sahen partizipative Prozesse als Basis fuer konstruktive Diskussionen nach dem Workshop
- waren optimistisch ob der Moeglichkeit fuer gemeinsames Handeln
- zogen in Betracht integrierte Bewertung und Partizipation auf andere Themen / Situationen zu uebertragen.

Die neuen Nachhaltigkeitsbewertungsmethoden werden von methodischen Weiterentwicklungen und weiteren Experimenten profitieren, aber sie erwiesen sich bereits bei verschiedenartigen Bedingungen als effektiv.

Keine einzelne Methode ist geeignet um alle Typen von Massnahmen, Programmen und Projekten zu evaluieren → ein differenzierterer Ansatz produziert bessere Ergebnisse.

Danke fuer Ihre Aufmerksamkeit!

Kommentare? Fragen?