

Haben autonome Maschinen Verantwortung?



TECHNISCHE
UNIVERSITÄT
DARMSTADT

**Prof. Dr. Christoph Hubig
Technische Universität Darmstadt**

**Tagung Autonome Systeme und Arbeit,
FGW Düsseldorf
25.01.2018**

Gliederung

1. Verantwortung - Fragehorizont
2. Kriterien der Verantwortlichkeit
3. Autonomie (Grenze: Erkennen/Entscheiden – Anerkennen)
4. intentionale Zustände (Grenze: Bewusstsein – Selbstbewusstsein)
5. Koaktion mit autonomen Systemen (u. a. Robotern)/Hybridisierung
6. Ausblick: Verantwortungsnetzwerke

1. Verantwortung - Fragehorizont

Gegenstand: Wofür?

(sing. Handlungen, -sfolgen, Güter, Rollenübernahme, Delegation von Rollen/Funktionen, Systemarchitektur, mediale Steuerung)

Subjekt: Wer?

(Individuum, Rollenträger, „elektron. Person“, „Hybrid“, Organisation, Institution)

Instanz: Wovor?

(Selbst/Gewissen, betroffene Dritte, moral. Autorität, Regelungsinstanz, Jurisdiktion)

Für welchen Zeitpunkt?

(retrospektiv, prospektiv/Vorsorge...)

Hinsicht: Im Blick worauf?

(moralisch, rechtlich, ökonomisch, politisch)

Unter welchen Maßstäben?

(Werte, Normen, Prinzipien der Moral, Sittlichkeit, Rechtszustände)

Raumbezug?

lokal, regional, global

2. Kriterien der Verantwortlichkeit

- Autonomie
- intentionale Zustände
- kausale Urheberchaft
- Kontrollfähigkeit (auch über Alternativen)
- normative Rechtfertigungsfähigkeit

- als
- „starke“ (allein intrinsisch)
 - „schwache“ (partiell extern gesteuert und geregelt, „delegiert“)
 - „als ob“/simulierte

Verantwortlichkeit

3. Autonomie (Grenze: Erkennen/Entscheiden – Anerkennen)

Autonomietyp

Delegationstyp

Rechtfertigungshypothek

(1)

operativ

Freiheitsgrade bei der Wahl optimaler *Mittel* (Effizienz/ Effektivität)/bei gegebenen Zwecken: Assistenzsysteme jeder Art

Erkennen (Sollgröße, Situation) und Entscheiden in Grenzen vollständig delegierbar (im Rahmen von (3))

utilitaristisch, vertragsethisch (Haftung); auf Basis von Interessenaggregation und Präferenzenabgleich (safety/ security), algorithmisierte/schematisierte Begründung

(2)

strategisch

Freiheitsgrade der Wahl optimaler *Zwecke* (Priorisierung, Umweg, Taktung, Dauer etc.) unter gegebenen Zielen: Planung und Realisierung von Abläufen z. B. bei der Fertigung, Pflege, Therapie, auton. Fahren

Erkennen und Entscheiden teilweise delegierbar, sofern Interventionen (Reversibilität, Abbruch, informationelle Selbstbestimmung/ limitorientierte Sensorik) im Rahmen von (3) gewährleistet

utilitaristisch, klugheitsethisch (Erhalt des *Handelnkönnens*)

(3)

moralisch

Freiheit der Anerkennung von *Prinzipien* und *Zielen* sowie der Selbstzuschreibung der Anerkennung

nicht delegierbar

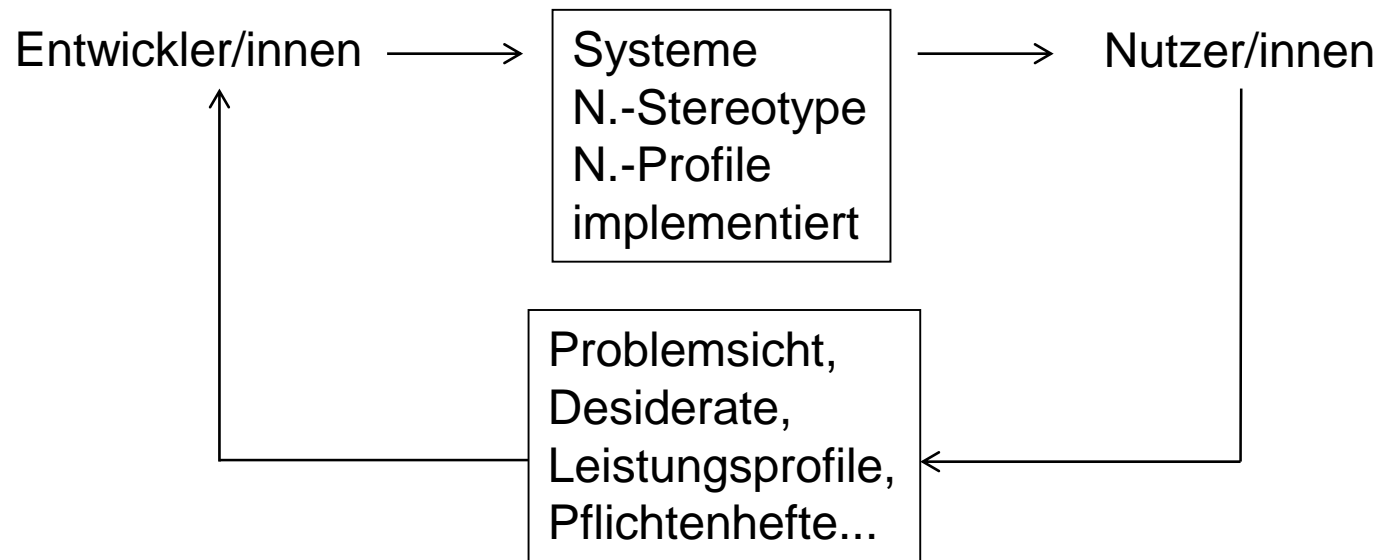
deontologisch (Menschenwürde, Menschenrechte)

4. Intentionale Zustände (Grenze: Bewusstsein – Selbstbewusstsein)

- Überführung von Präsentiertem in *Repräsentationen* als...
(Situation/Kontext, Systemzustand, Sollgröße, Strategien und ihre Passung (adaption, mitigation etc.), als zu verfolgen oder zu vermeiden, zu gewährleisten oder aufzugeben, als disponibel oder indisponibel etc.)
↳ „propositionale“ Gegenstände der Erinnerung, Vergegenwärtigung und Erwartung einschließlich der Systemzustände selbst „Selbstmodell“
 - —————> „Bewusstsein“ (auch bei höheren tierischen Spezies)
- ≠ Selbstbewusstsein
als voraussetzender Instanz der *Produktion* von Repräsentationen („Du sollst Dir kein Bildnis machen“),
- als Instanz der verantwortlichen Anerkennung der Prinzipien der Bildung von Repräsentationen
 - der Anerkennung der Einnahme eines Verhältnisses zu Repräsentationen auf Basis von Autonomie (3)
 - nicht in K. I. implementierbar wg. log. Paradoxien und Aporien

5. Koaktion mit autonomen Systemen (u. a. Robotern)/Hybridisierung (1)

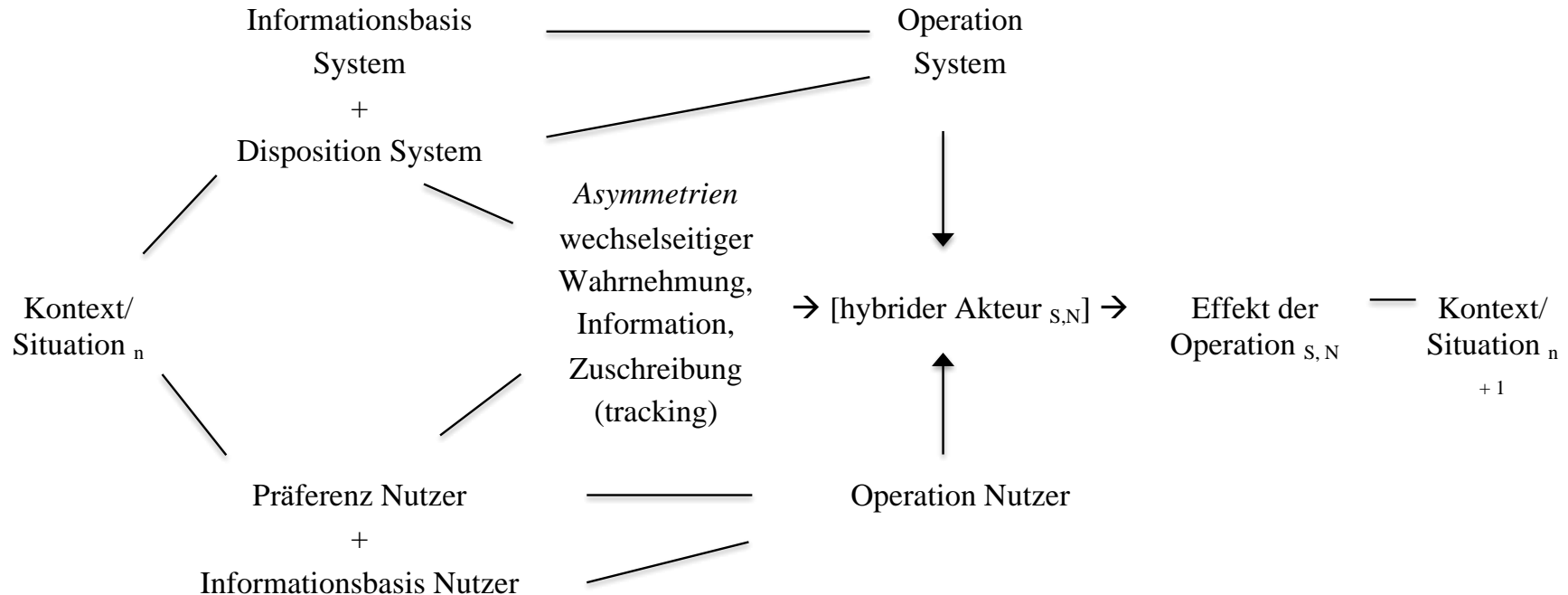
- a) klass. Intentionales Agieren, das auf der Basis von Erwartungserwartungen
Interaktion: auf intentionales Agieren bezogen ist Weber 1968, 441;
Luhmann 1984, 412 u. v. a.)



- a) technisch *vermittelte* qua Sensordatenfusion, Kontextmodellierung und mediale Steuerung
Interaktion: (Trainingsbedingungen): Bildung von Systemdispositionen im Rahmen von Rollenzuweisung und Gestaltungsstandards

5. Koaktion mit autonomen Systemen (u. a. Robotern)/Hybridisierung (2)

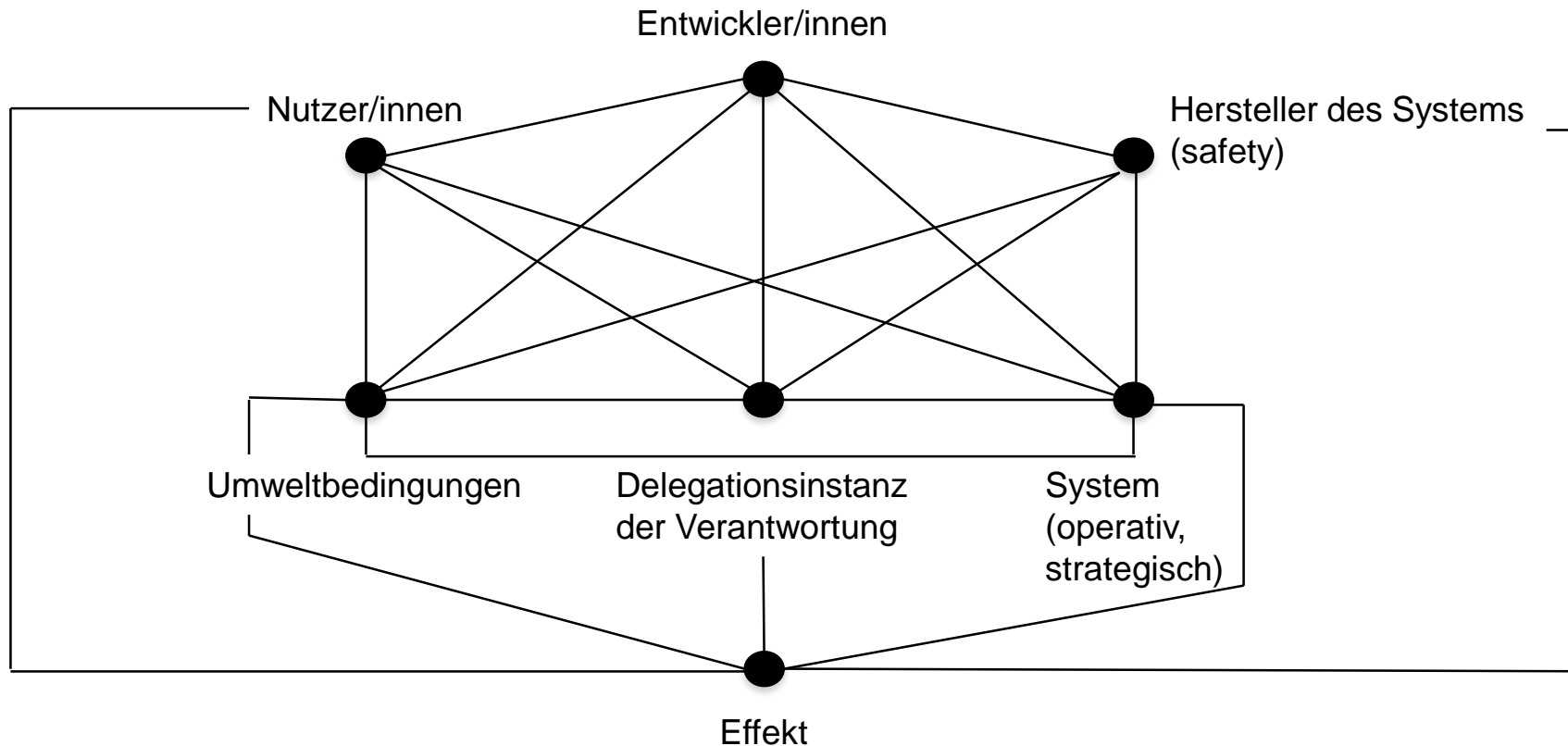
c) Koaktion mit autonomen Systemen:



hin zu „Spiegelwelten“/„filterbubble“: Priorisierung des bisher Vollzogenen (Routinen), Abbau von Irritationen und Unerwartetem, Relevanz-Ranking als f (Anschluss handeln), Einschränkung der Bildung von Selbstbewusstsein (z. B. Suchalgorithmen im Netz, z. B. Systeme „lernen Emotionen lesen“ (smart home, Fertigungsroboter, Kuschelroboter) ...

6. Ausblick (1): Verantwortungsnetzwerke (ANT: „obligatorische Passagepunkte“)

- Knoten im Netz: Zusammenwirken von technischen, intentionalen und natürlichen „Fragmenten“ (Latour)



...

6. Ausblick (2): Verantwortungsnetzwerke – Die Notwendigkeit der Parallelkommunikation

Drei Ebenen

```
graph TD; A[Drei Ebenen] --- B[in der Entwicklung:]; A --- C[während der Nutzung:]; A --- D[nach der Nutzung (in Intervallen):];
```

in der
Entwicklung:

Entwickler/innen-
Nutzer/innen-
Delegatoren-Dialoge
über
Systemarchitektur
und Trainings-
bedingungen

während der
Nutzung:

system- oder nutzerindu-
zierte Dialoge on demand
zur Freilegung *partieller*
Algorithmen, Herstellung
operativer Transparenz,
Management von Irritationen

nach der Nutzung
(in Intervallen):

Foren der Bilanzierung,
Evaluation, Supervision,
Monitoring
vs. anonyme
Vergemeinschaftung

Vielen Dank für Ihre Aufmerksamkeit!
