



Smart Robots – Smart Ethics?



Klaus Kornwachs

Universität Ulm, acatech München, Berlin,
Büro für Kultur und Technik, Argenbühl



Smart Robots – Smart Ethics?

1. Erzählung: Konvergenz und Universalisierung
2. Ethik: Grundsätze, Werte, Verantwortung
3. Roboter: Subjekte und/oder Objekte der Verantwortung?
4. Bedingungen verantwortlichen Handelns: Smart Ethics



Smart Robots – Smart Ethics?

1. Erzählung: Konvergenz und Universalisierung

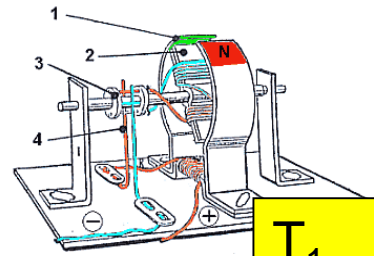
2. Ethik: Grundsätze, Werte, Verantwortung

3. Roboter: Subjekte und/oder Objekte der Verantwortung?

4. Bedingungen verantwortlichen Handelns: Smart Ethics

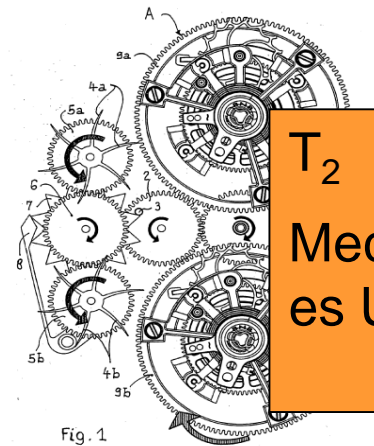


Roboter sind “Kinder” von Konvergenzprozessen



T_1

Elektromagnetischer Motor



T_2

Mechanisches Uhrwerk

Elektromagnetisches Uhrwerk

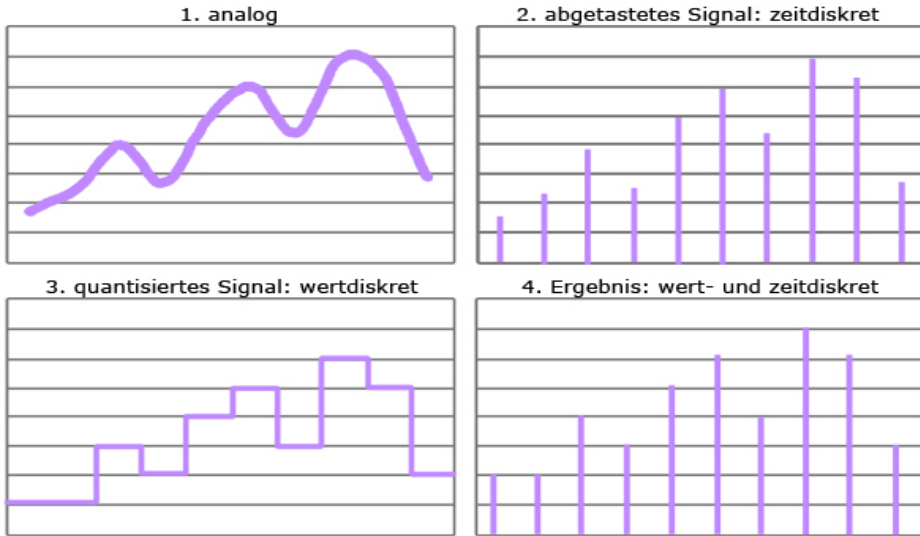
$T_1 \text{ conj } T_2$



Konvergenz:

Kohärente Technologien bilden eine gemeinsame Technologie

Convergences on the Technological Level



Digitalisierung (Nyquist 1928)

Computer Technology
Incl. Software Technology



Communication
Technology
Analog Circuits and Wires



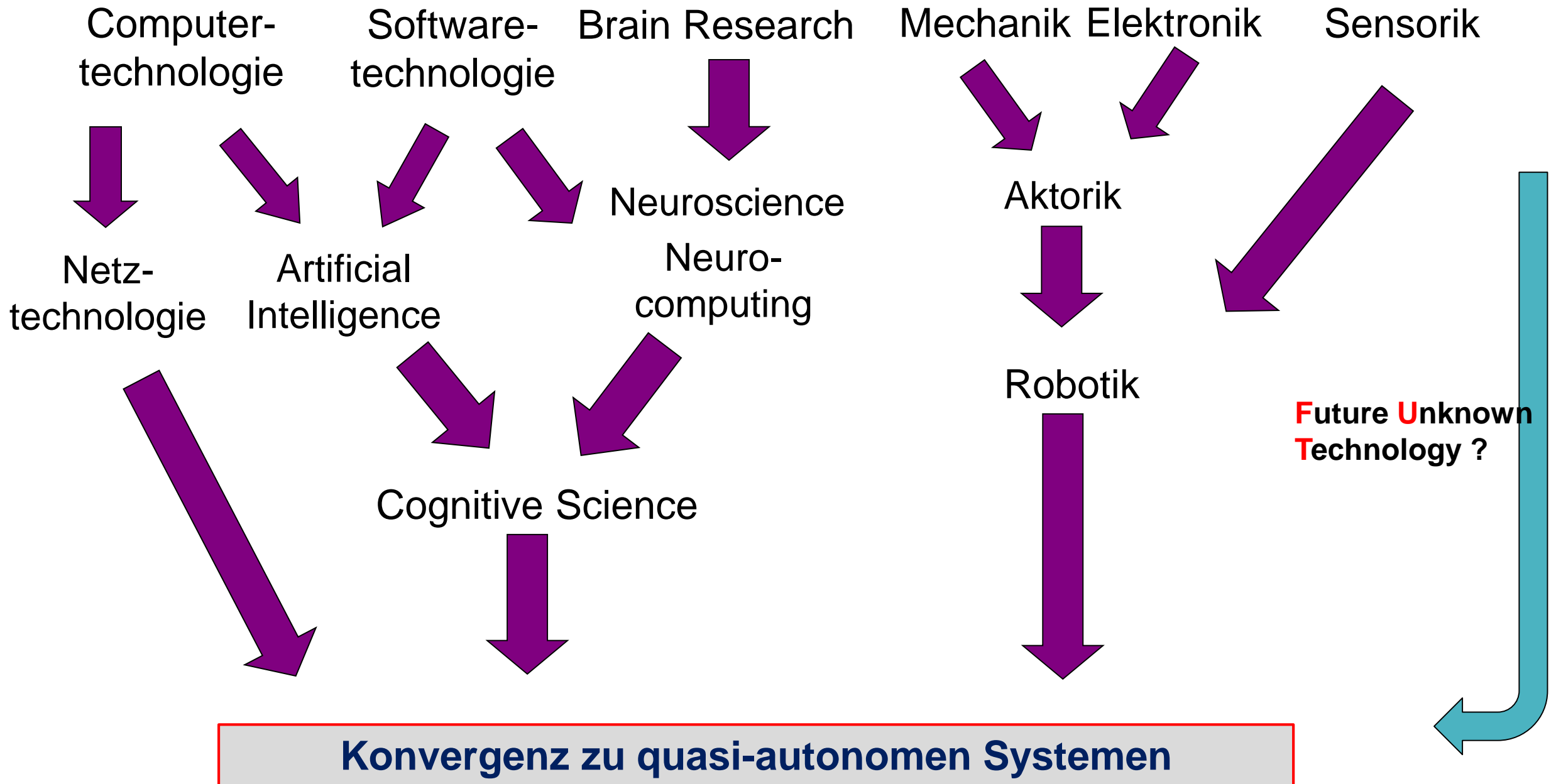
High Frequency
Technology
Electromagnetic Waves



World Wide
Communication

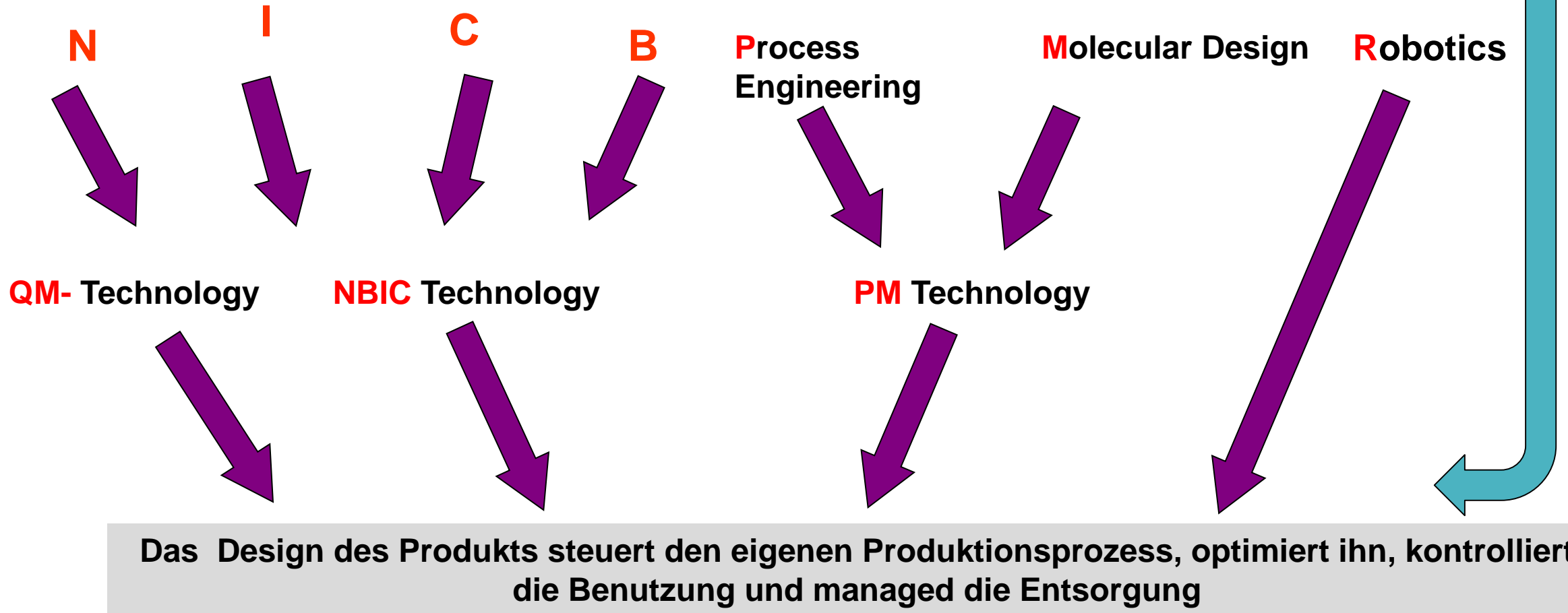


ICT or I



Convergences on the Technological Level

Future **U**nknown
Technology ?



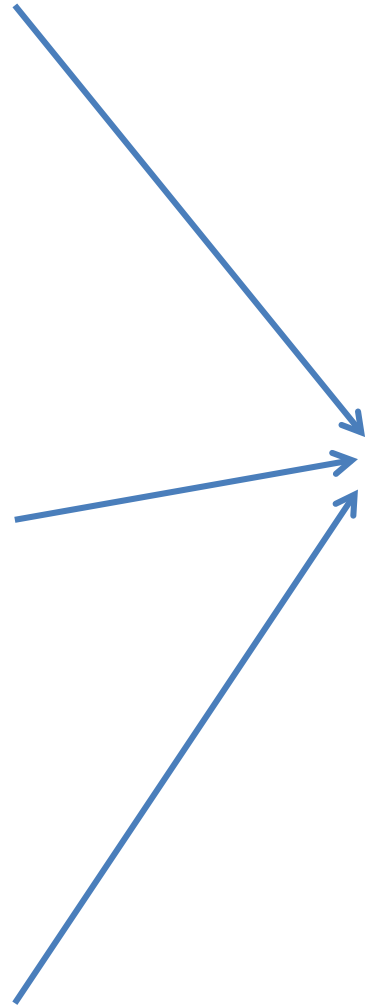
Das Produkt ist Teil eines Systems aus interagierenden wertschöpfenden intelligenten Robotersystemen

Konvergenz führt zur Universalisierung



Foto: thinkgeek.com

Konvergenz führt zur Unversalisierung



CNC Universal
Machining Center



Drehbank

<http://commons.wikimedia.org/wiki/File:RMB3-2yuan-A.jpg>



<http://commons.wikimedia.org/wiki/File:MoriSeikiLathe.jpg>

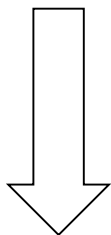
Smart Robots – Smart Ethics?

1. Erzählung: Konvergenz und Universalisierung
- 2. Ethik: Grundsätze, Werte, Verantwortung**
3. Roboter: Subjekte und/oder Objekte der Verantwortung?
4. Bedingungen verantwortlichen Handelns: Smart Ethics



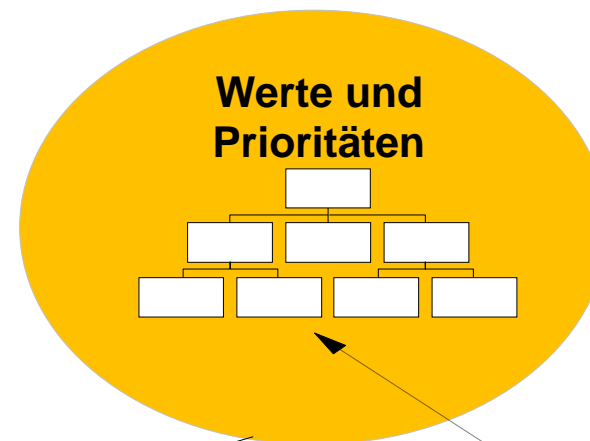
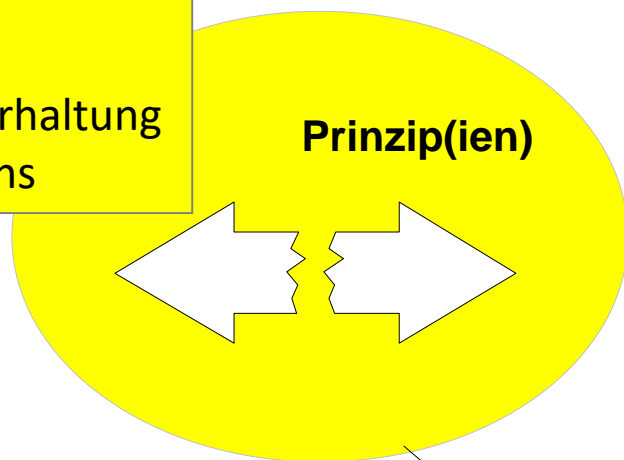
- ✓ Goldene Regel
- ✓ Kat. Imperativ von Kant
- ✓ Wenn das alle täten ...
- ✓ Prinzip der Bedingungserhaltung verantwortlichen Handelns

Gesellschaft



Rechte

Grundrechte sind die jedem Menschen in jeder Rechtsgemeinschaft zustehenden Rechte, die er nicht aufgrund seiner Zugehörigkeit zu einer Rechtsgemeinschaft und deren rechtskonstituierenden Akte beanspruchen kann, sondern aufgrund seines Menschseins.



Normen \Rightarrow Moralische Sätze



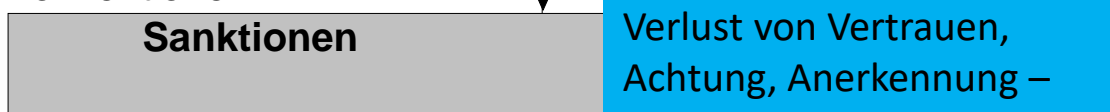
Tugenden



Sitten



Konventionen



- ✓ Leben
- ✓ Nachhaltigkeit
- ✓ Freiheit
- ✓ Gerechtigkeit
- ✓ Ehre
- ✓ Ruhm
- ✓ Wahrheit
- ✓ Persönl. Wohlergehen
- ✓ Liebe
- ✓ Vertrauen
- ✓ Sozialer Status
- ✓ Anerkennung
- ✓ Macht
- ✓ Schönheit

Ordnung, Sparsamkeit, Reinlichkeit, Pünktlichkeit, Höflichkeit, Fleiß

Fleiß, Tapferkeit, Besonnenheit, Gelassenheit, Bescheidenheit, Wahrhaftigkeit, Treue, Vertrauen, Gerechtigkeit, Toleranz

Unmittelbarkeit, Ursprünglichkeit, Lebendigkeit, Unbedingtheit, ggf. Spontaneität

Diagnosekompetenz + Urteilskraft

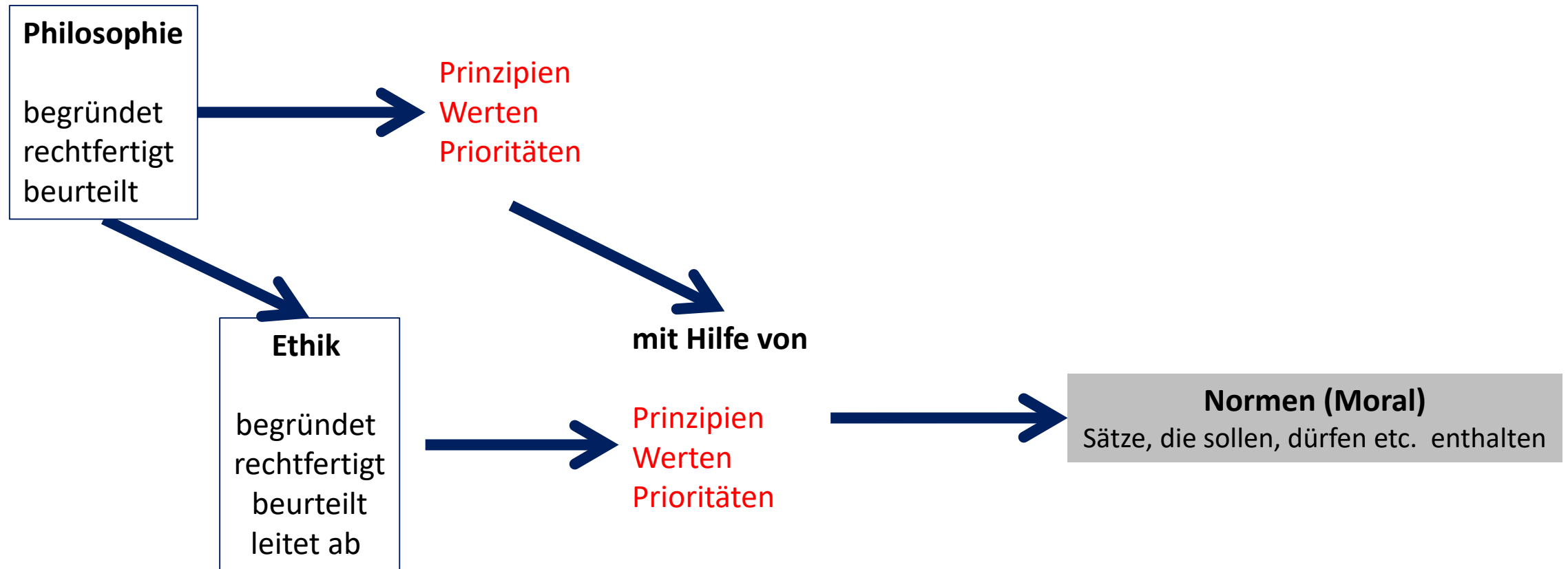
- Was Du nicht willst, das man Dir tu, das füg auch keinem anderen zu.

+

- Das Leben ist für mich ein größerer Wert als die Ehre.
- Ich möchte nicht getötet werden

=

Du sollst nicht töten
Recht auf Leben



Die klassische Quelle der Entstehung von Wertevorstellungen sind:

- Erfahrung von Bedürfnissen,
- Religiös offenbarten Vorstellungen,
- Bestimmungen einer philosophischen Anthropologie,
- Theorien über Evolution und Gesellschaft, Geschichte und Politik,
- Einsichten, die sich in Diskursen als zustimmungsfähig erweisen,
- **Technisch-organisatorische Erfahrungen:**

Konvergenz

Universalisierung

Auseinanderfallen von Merkwelt und Wirkwelt

Invertierung der Zweck- Mittel Relation

Beschleunigung und Disruption

Stabilität

Übersichtlichkeit

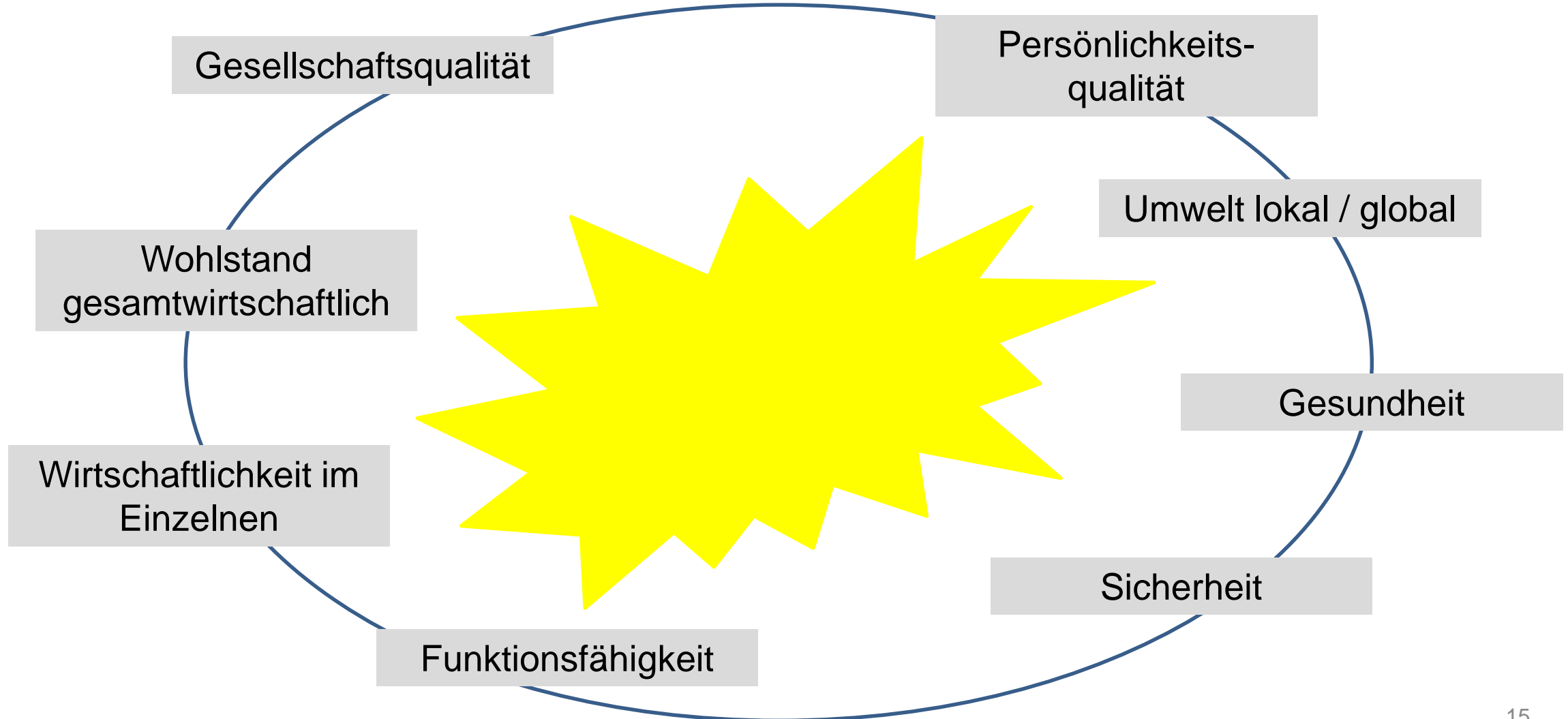
Vorsicht

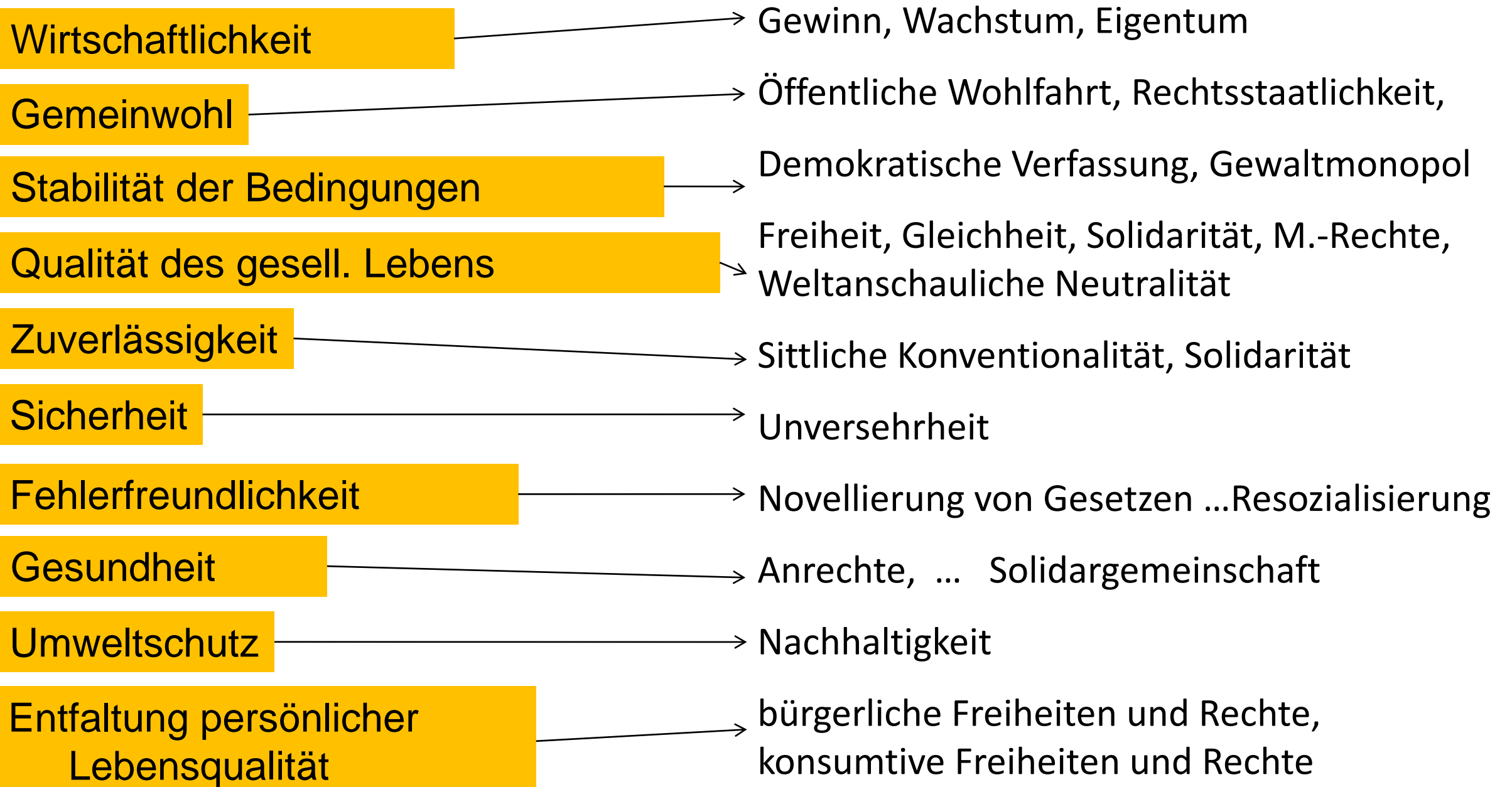
Moralisierung der Mittel

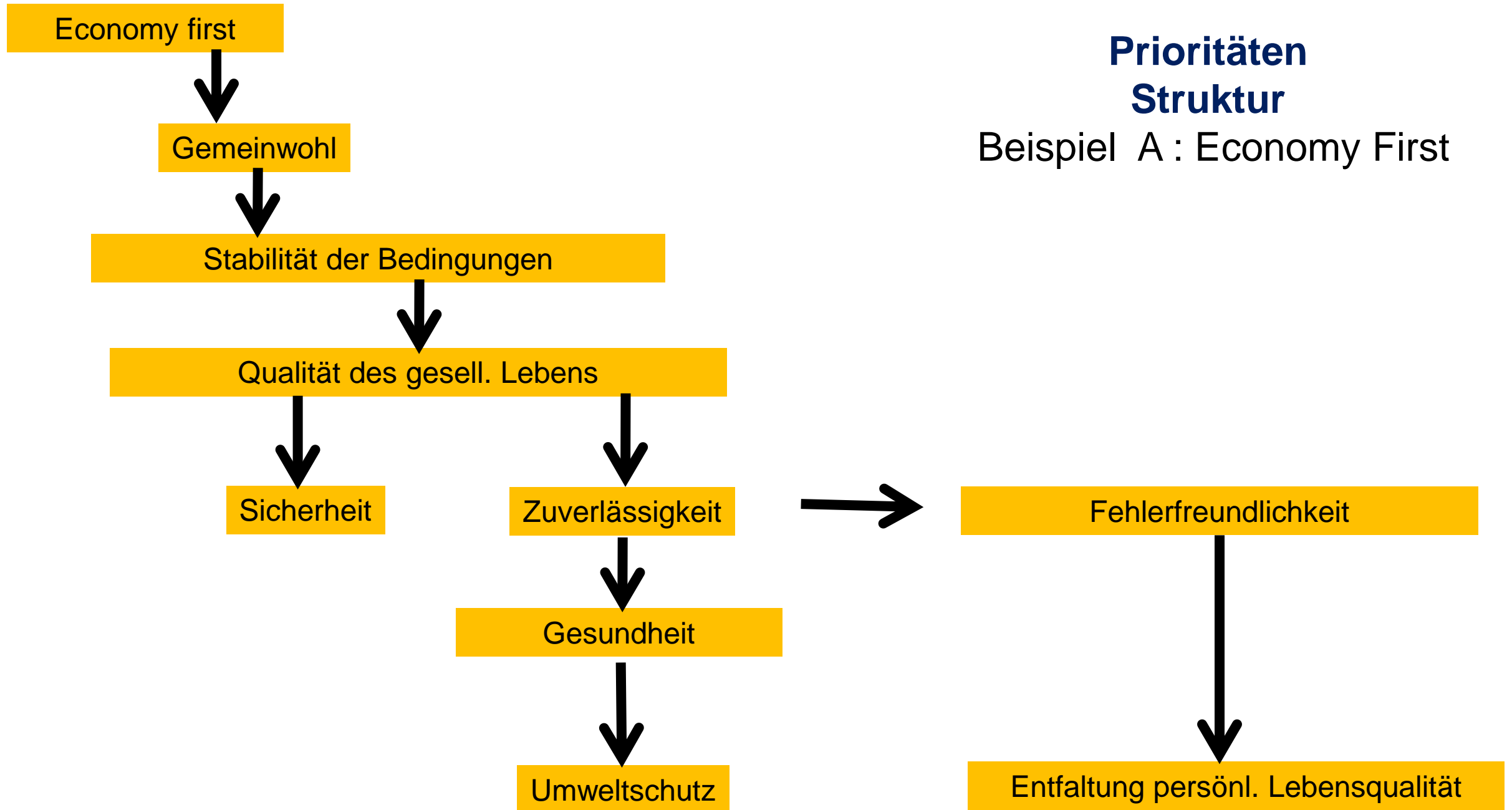
Entschleunigung

Was für eine Technik wollen wir haben?

Das VDI Werteoktogonal (VDI Richtlinie 3780 Technikbewertung (1997))

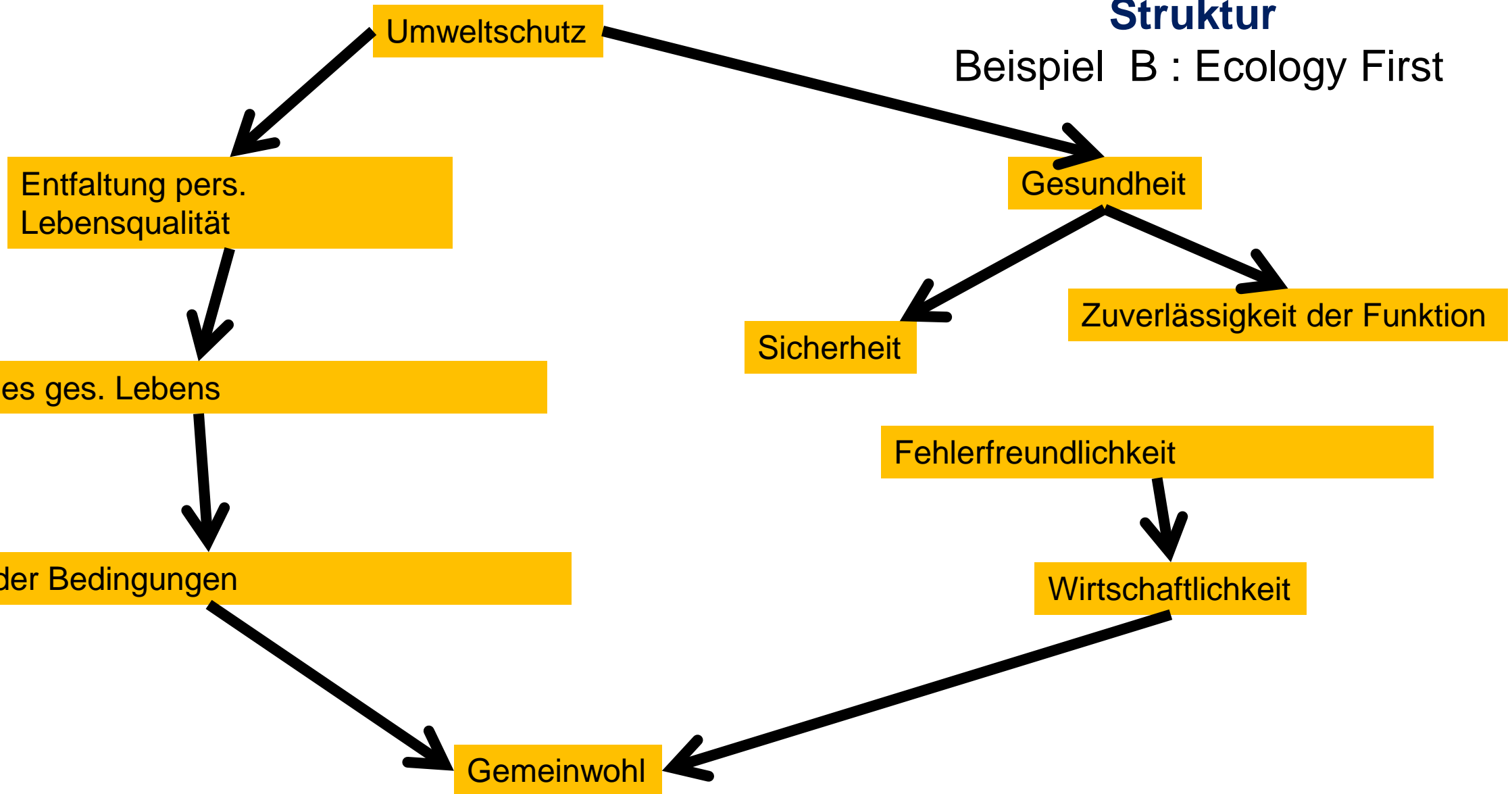


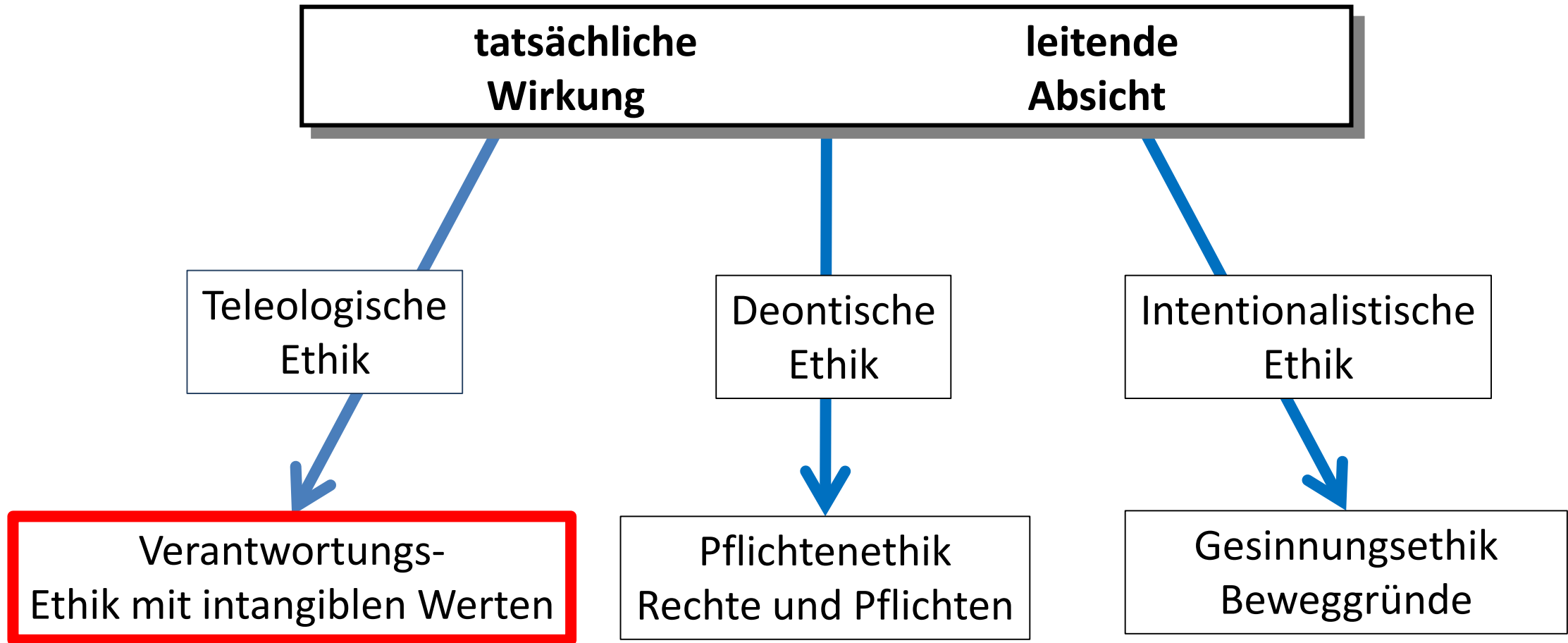




Prioritäten Struktur

Beispiel B : Ecology First





Verantwortung:

Subjekt der Verantwortung

Objekt der Verantwortung

Instanz

Zeithorizont

Sanktion

Wer

Individuum oder Kollektiv

Für was

Absicht, Handlung,
Folgen, Erfüllung

Gegenüber wem

Vorgesetzter, Firma,
Gesellschaft

Wie lang

Dass etwas geschieht,
was geschieht, Folgen

Bestrafung, Belohnung

Verlust, Kompensation

Smart Robots – Smart Ethics?

1. Erzählung: Konvergenz und Universalisierung
2. Ethik: Grundsätze, Werte, Verantwortung
- 3. Roboter: Subjekte und/oder Objekte der Verantwortung?**
4. Bedingungen verantwortlichen Handelns: Smart Ethics





www.levity.com/alchemy/alchemy_laboratories_engravings.html

Erzählung der Hybris:

Wagner, der Famulus von Dr. Faust, erschafft in FAUST II ein kleines Menschlein (*homunculus*)

Hintergrund: Harnstoffsynthese 1828 durch Friedrich Wöhler

Verwandt: Geschöpfe des Prometheus, Pygmalion, Golem, Frankenstein, Pinocchio, Zauberlehrling

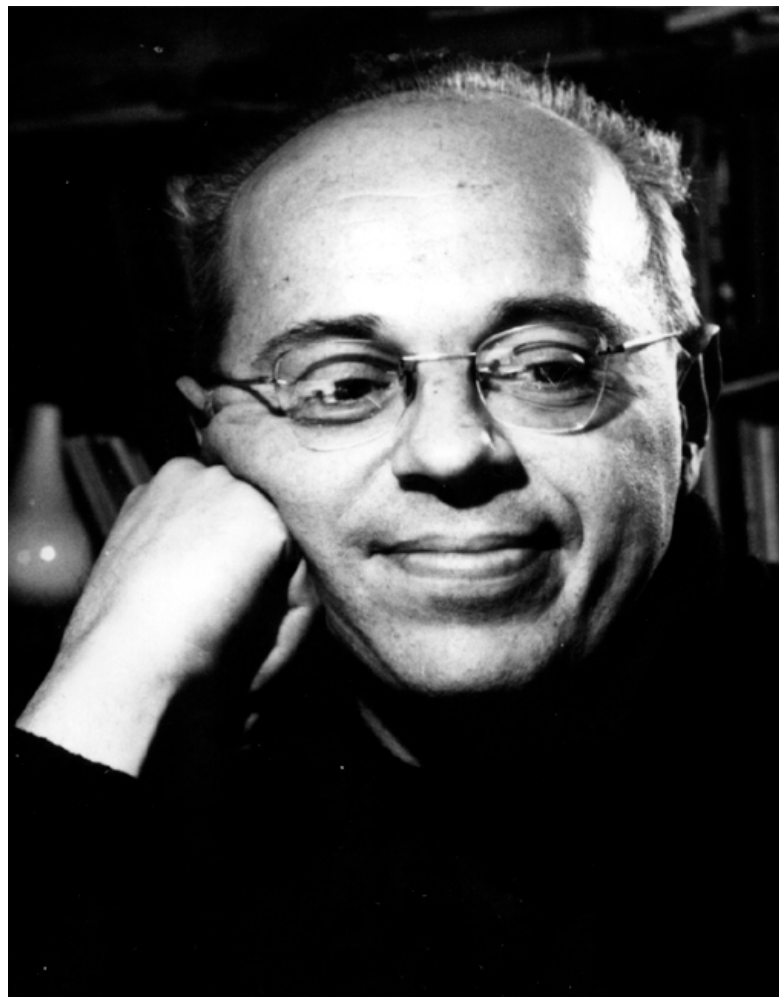
Paracelsus in *de natura rerum* (1538): Spermien 40 Tage in Pferdemist, dann 40 Tage mit Menschenblut nähren ergibt ein kleines menschliches Wesen

Funktion:

Denkbarkeit, Machbarkeit

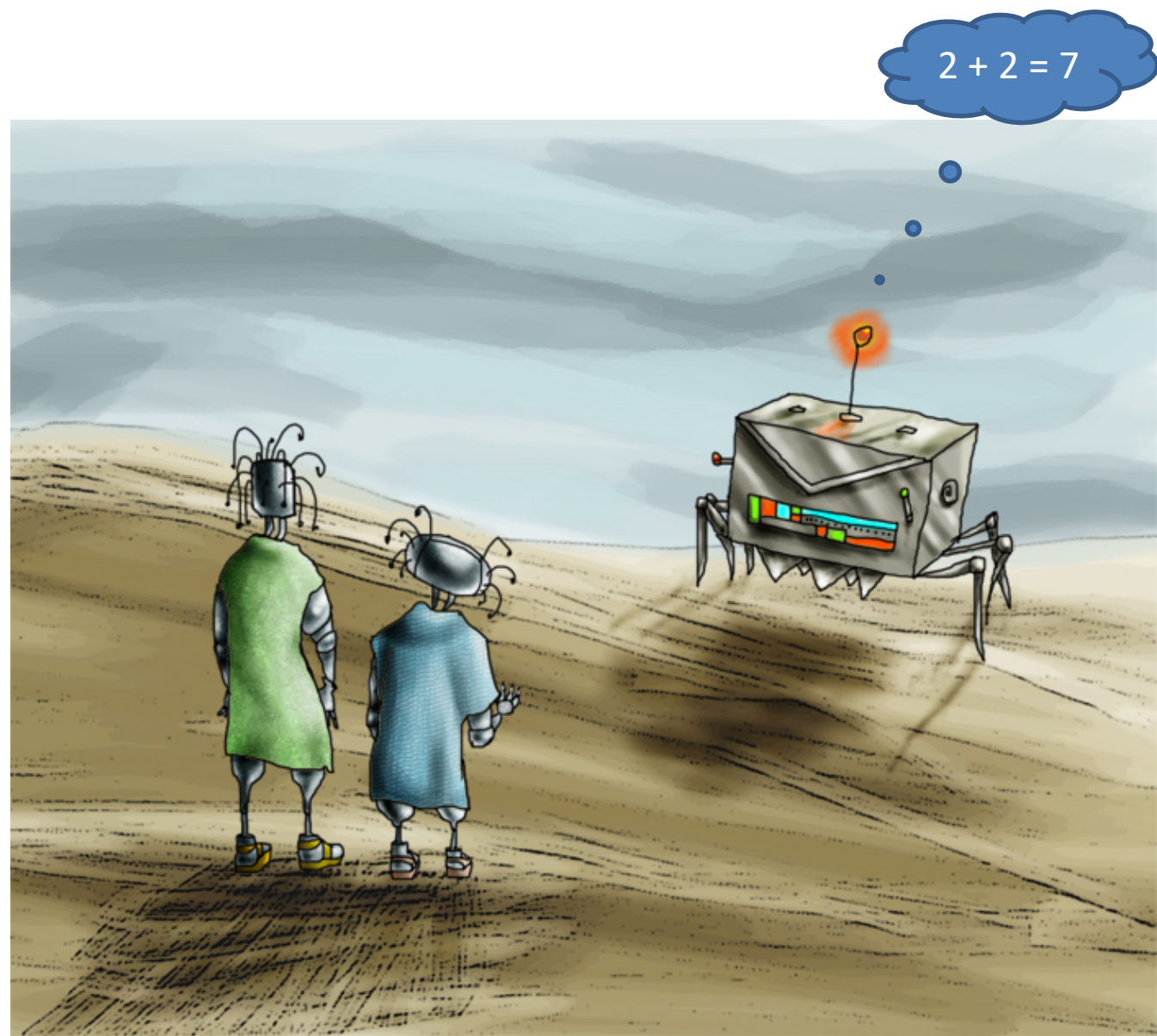
Schaffen eines willfähigen Helfers

Hybriskritik – Entgleiten der Kontrolle



Stanisław Lem (1921 - 2006)

Stanisław Lem: *Trurl's Maschine*.







Magazin
auffüllen
übernehme ich.

Samstag kann
ich leider nicht.

Magazin
leer, bitte
auffüllen!

Kundenauftrag:
50 Getriebe bis Montag

Schalt mich aus!

Kapazität ist bis
Freitag ausgebucht!

Ich kann diesen
Samstag Arbeiten.

Ich muss in 2 Stunden
am Warenausgang
sein!

Einsatz von intelligenten Robotern (IR)

Success-stories oder Projektionen?

Produktionsroboter: Moderne Produktionsanlagen – Industrie 4.0

Serviceroboter: Haushalt, Zustelldienste, Health Care, Altenpflege, Smart Home, autonomes Fahrzeug, Analysen

Netzwerkroboter: Suchmaschinen, Bots, Trolle und Malware

Aussenweltroboter: Raumfahrt, Bergbau, Landwirtschaft, Lagerhaltung

Therapieroboter, Chirurgieroboter, Nanoroboter

Militärroboter: Drohnen, autonome Mobilität, Kampfroboter, intelligente Munition

Edutainment: Unterricht, Spiele, vernetzte Spiele, augmented reality,

Sex-Roboter: Sexindustrie

Fragen der Roboterethik:

(Roboter = intelligente Robotersysteme)

Gibt es besondere **normative Anforderungen** an das Handeln mit Robotern (Design, Herstellen, Programmieren, Benutzen, Entsorgen)?

Gibt es eine **besondere Verantwortung** für die Folgen des Einsatzes von Robotern?

Wie weit darf man die Autonomie treiben? **Entscheidungsersetzenden Einsatz?**

Wie weit darf man die **Simulation humanoider Verhaltensmuster** bei Robotern treiben, die im Alltag, in zuwendungsersetzender Absicht oder im Pflegebereich eingesetzt werden?

Roboterethik I

Annahme: Roboter sind (genuine) Subjekte mit

- ✓ eigenständiger Zielsetzung
- ✓ (Selbst-)bewusstsein (von Genealogie bis zur eigenen Endlichkeit)
- ✓ Erwerb einer Geschichte durch Lernprozesse und “Erfahrung”
- ✓ Fähigkeit zur Selbsttranszendenz (z. B. Metadiskurs, Ich - Nicht-Ich Differenz)
- ✓ Gefühlen und Stimmungen
- ✓ ...

Konsequenz: Müßten solche artifiziellen Subjekte

- ☐ Rechte und Pflichten haben?
- ☐ Anspruch auf Achtung der Autonomie?
- ☐ Anspruch auf Toleranz und Respekt?
- ☐ Zubilligen von Verantwortung (einschl. Sanktionen)?

Pragmatische Antwort: Bis auf weiteres **NEIN!**

Begründung: Solche genuinen Roboter wird es nicht geben
und/oder wir wollen sie nicht haben

Roboterethik II

Annahme: Roboter können mit folgenden Eigenschaften gebaut und eingesetzt werden:

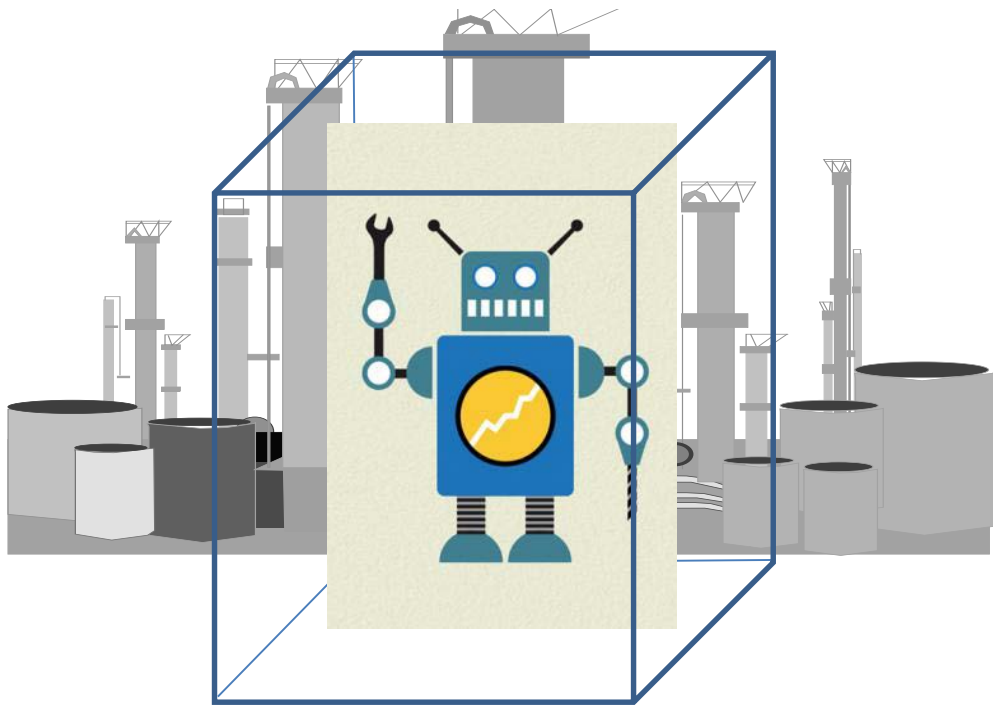
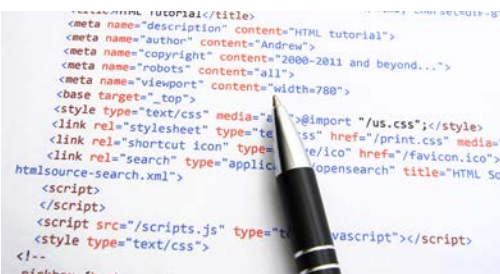
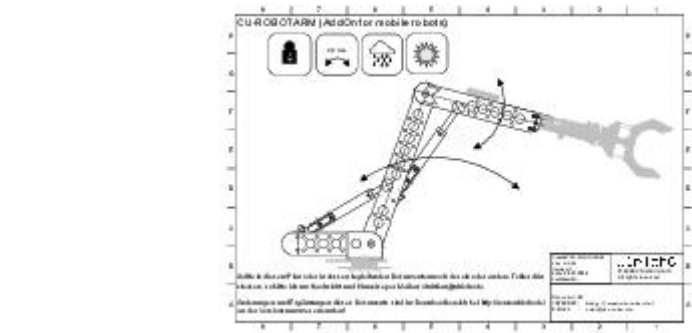
- ✓ lernende Maschine (adaptive Programmierung)
- ✓ selbstständige Suche und Erweiterung nach Trainingsmengen (Big Data)
- ✓ Interne und externe Aktorik und Sensorik (einschl. Selbstmodifikation)
- ✓ Humanoides Aussehen und humanoide Verhaltensmuster
- ✓ Simulation von Zuwendung und Gefühlen

Konsequenz: Wie weit darf man mit der Simulation humanoider Eigenschaften gehen?

- ☐ Es darf keine Verantwortungslücke geben
- ☐ Es darf keine völlig autonomen Entscheidungen geben (E.-ersetzend)
- ☐ Es darf kein unüberwachtes Lernen geben
- ☐ Es darf keine falschen Zuschreibungen geben (kein Turing-Imitation Game)

Pragmatische Antwort: Roboter müssen so designed sein, dass deren Nutzer immer noch verantwortlich, also dilemmafrei handeln können

Sensorik Kommunikation KI Aktorik



Zwecksetzung Design Bau Programmierung Spezifikation & Test Einsatz Wirkungen



Aktorik
Sensorik
Rechner
Steuerung

Algorithmen & KI
Datenbasen
Trainingsmengen

Teach in
Play Back
Paramtereingabe

Sensordaten
Lernen
Rekonfigurieren

Erfolg
Zuschreibungen
Havarien

Akteure: Konstrukteur, Produzent, Mathematiker & Programmierer, Ingenieur, Betreiber /Nutzer Betroffene

Verantwortung:

Subjekt

Technische Entwicklung

Einzelperson, Gruppe, Firma, Betreiber, Institutionen -
nicht der Roboter

Objekt

Funktionsfähigkeit, Schadensfreiheit, Verträglichkeit,
Nachhaltigkeit, Unverletzlichkeit, Zuverlässigkeit

Instanz

unklar: Betroffenen,

Gründe

in der Gemeinschaft akzeptierte Normen (Regeln, Gesetze)

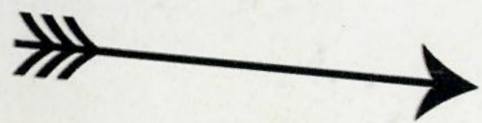
Zeithorizont

unklar – Problem der Langzeitfolgen, Irreversibilität

Sanktion

Selbstschädigung; Haftung (BGB), Bestrafung (STGB);
Vertrauensverlust, ökonomischer **Verlust**

Ich habe KEINE LÖSUNG



ABER ICH



bewundere das PROBLEM

Smart Robots – Smart Ethics?

1. Erzählung: Konvergenz und Universalisierung
2. Ethik: Grundsätze, Werte, Verantwortung
3. Roboter: Subjekte und/oder Objekte der Verantwortung?
- 4. Bedingungen verantwortlichen Handelns: Smart Ethics**



Negative Folgen der Aktionen von Robotern:

Generell:

- Falsche Entscheidungen (Optionsauswahl aus n-fache Alternative)
- Falsche Aktionen aufgrund falscher Sensorik
- Falsche Aktionen aufgrund falscher Programmierung /Spezifizierung
- Falsche Aktionen aufgrund von Fremdsteuerung (Hacking)
- Falsche Aktionen aufgrund von Fehlfunktionen durch Verschleiß, Altern, (Hardware, Software)

Gegen Personen:

- Verletzung bis Tötung durch Aktorik
- Verweigerung Falschdurchführung von Befehlen
- Personenfalschidentifizierung

Unerwünschte Lernergebnisse

Gegen Institutionen:

- Usurpation von existierenden Vernetzungen
- Neuaufbau von unerwünschten Vernetzungen
- Aufbau von unerwünschten Backdoors
- Unmöglichkeit der Abschaltung ohne schwere Schäden

Gegen Gesellschaft

- Abhängigkeit der Infrastruktur
- Zeitkonstanten zur Substitution (Umschalten auf "Handsteuerung")
- Kollektive Effekte vernetzter Computer (Rebound-Effekte)

Was wollen wir – Sklaven, intelligente Mitarbeiter oder zukünftige Herrscher?

Isaac Asimovs Robotergesetze (1950),
ergänzt um das Nullte Gesetz (1983)
normative Sätze



aus den
Paradoxien, die
aus diesen
Gesetzen
erschließbar sind,
kann man neue
Erzählungen
bauen

0. Ein Roboter darf die Menschheit nicht verletzen oder durch Passivität zulassen, dass die Menschheit zu Schaden kommt.
1. Ein Roboter darf keinen Menschen verletzen oder durch Untätigkeit zu Schaden kommen, außer er verstieße damit gegen das nullte Gesetz.
2. Ein Roboter muss den Befehlen der Menschen gehorchen – es sei denn, solche Befehle stehen im Widerspruch zum nullten oder ersten Gesetz.
3. Ein Roboter muss seine eigene Existenz schützen, solange es sein Handeln nicht dem nullten, ersten oder zweiten Gesetz widerspricht

Einwand: Roboter sind keine agierenden Subjekte

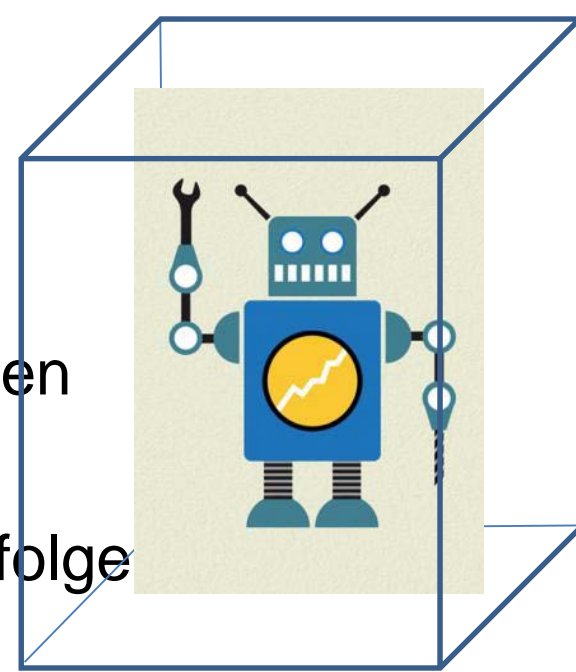
Verantwortungslücke

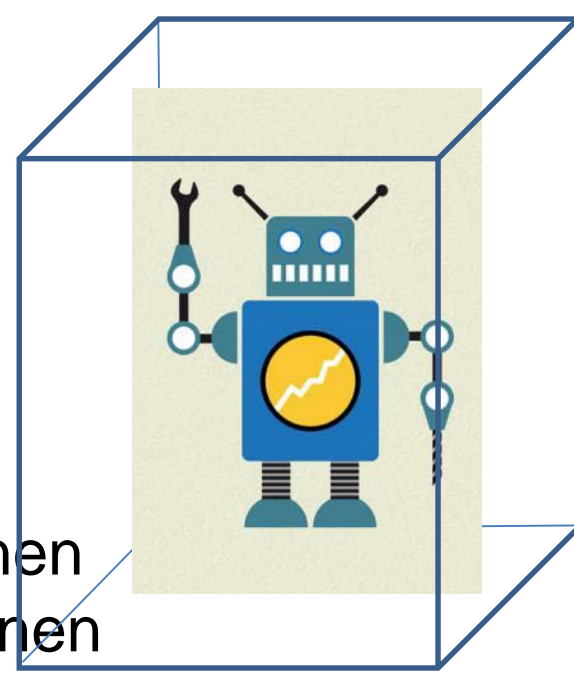
“Autonome Systeme” bilden als lernende Systeme Muster aus Attraktoren in neuronalen Netzen. Diese Muster bestimmen dann das Verhalten des Roboters.

Dieser Vorgang ist deterministisch und hängt von der Reihenfolge und den eingegebenen bzw. akquirierten Daten ab.

Dieser Vorgang ist weder prognostizierbar, noch aus praktischen Gründen nachprüfbar, weil die Nachvollziehbarkeit um Größenordnungen mehr Kapazität bindet als der Rationalisierungseffekt durch den Einsatz eingespart hätte.

Aktionen des Roboters erscheinen autonom, und ihn als Verursacher einer eingetretenen unerwünschten Nebenfolge zu identifizieren, liegt nahe – der Zauberbesen ist schuld. Roboter sind keine Subjekte der Verantwortung





Wir sollten solche Roboter bauen,

- ✓ die keine Verantwortungslücke suggerieren,
- ✓ die Benutzer und Betreiber nicht in dilemmatische Situationen führen, in denen sie nicht mehr verantwortlich handeln können (.. der Roboter ist schuld)
- ✓ deren Lernprozesse kontrollierbar sind,
- ✓ deren Einsatzfolgen wir nach dem Stand der Technik absehen können

d.h. die dem Prinzip der Bedingungserhaltung verantwortlichen Handelns genügen

Handle so, dass die Bedingungen der Möglichkeit des verantwortlichen Handelns für alle Beteiligten erhalten bleiben.

Daraus folgt, daß man nur nach solchen Normen handeln sollte,

mit denen alle von dieser Handlung Betroffenen

im Hinblick auf ihre weiteren Handlungsmöglichkeiten und ihrer ethischen Relevanz,

wenn auch nicht in allen faktualen einzelnen Folgen,

doch prinzipiell einverstanden sein können und noch verantwortlich handeln können.

Subjekt der Verantwortung (Lenk 1987):

Moralisches Urteilsvermögen

Selbstbewußtsein

Entscheidungsfähig

Wahlfreiheit

Abschalten können

Bedingungen verantwortlichen Handelns

Autonomie des Subjekts

Die Autonomie des Subjekts muss sich aus der Freiheit bestimmen.

“Freiheit von ...”	“Freiheit zu ...”
Autonomie	Willensfreiheit.

Möglichst weitgehende Unabhängigkeit von äußeren und inneren Einfluß- und Bestimmungsfaktoren ist nicht hinreichend (“Freiheit von ...”)

Erst Willensfreiheit, der “Freiheit zu ...” ist zusammen mit der Autonomie Voraussetzung für die Möglichkeit des Subjekts, Verantwortung wahrnehmen zu können.

Roboter sind keine Subjekte

Gewährung der Autonomie

Orientierungsmöglichkeit an Prinzipien und Werten

Zeit für Entscheidungen gewähren

Priorisierung von Werten gewähren

Akzeptanz pluraler Wertesysteme

Wo liegt der gemeinsame Nenner?

Bedingungen verantwortlichen Handelns

Eher praktische Bedingungen

Die Instanz muß existieren, sie muß anrufbar sein, und sie muß fähig sein, Sanktionen durchzusetzen

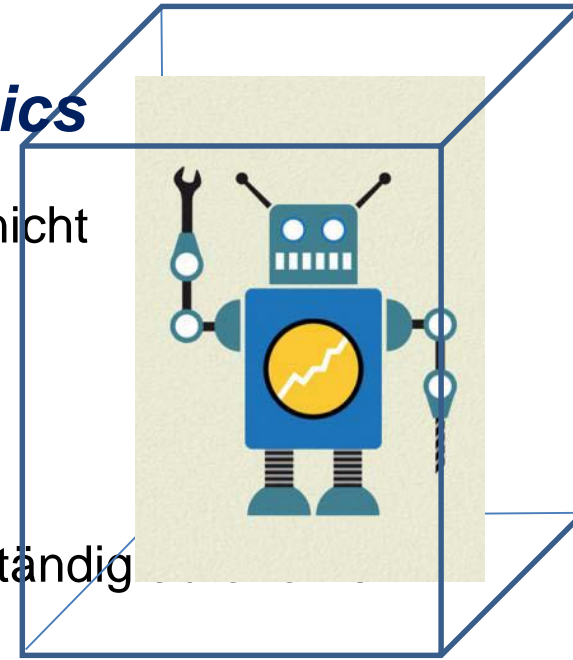
Kein Bau und Einsatz ohne Klärung

Zeitliche Dimension: Absicht, Handlung und Folgen müssen in geeigneter Weise zeitlich proportioniert sein ($T(\text{inst.}) > T(\text{indiv.})$)

Über Absicht, Handlung und Folgen muss fehlendes Wissen prinzipiell und praktisch erwerbbar sein (Transpatenz: individuell und arbeitsteilig)

Vorschlag für 8 Regeln (Normen) als Ergebnis einer *smart ethics*

- 1) Setze Systeme niemals entscheidungsersetzend ein. Auch Roboter dürfen nicht in entscheidungsersetzender Absicht eingesetzt werden.
- 2) *Nihil Nocere* – toleriere keinen Schaden gegenüber Benutzern.
- 3) Nutzerrecht bricht Produzentenrecht.
- 4) Baue keine autonomen Systeme, die man nicht mehr abschalten kann. Vollständig Systeme sollen nicht zugelassen werden.
- 5) Das Herstellen selbstbewusster, autonom agierender Roboter ist verboten (analog dem Chimärenverbot und Human-Klonverbot in der Gentechnologie)..
- 6) Täusche kein menschliches Subjekt als Gegenüber vor. Maschine muss Maschine bleiben, Imitation und Simulation muss erkennbar bleiben. Es muss bei der Mensch-Maschine-Kommunikation allen beteiligten Menschen immer klar sein, dass der maschinelle Kommunikationspartner eine Maschine ist.
- 7) Wer die Frage und den Zweck nicht weiß, kann mit der Systemantwort und dem Verhalten eines Roboters nichts anfangen. Gebe immer den Kontext an.
- 8) Wer Technik erfindet, herstellt, betreibt oder entsorgt, hat Interessen. Die Interessen sind redlich offenzulegen.



Literatur

Arnsburg, René: *Maschinen ohne Menschen? Industrie 4.0 – Von Schein-Revolution und der Krise des Kapitalismus.* Manifest Verlag 2017

Asimov, Isaac: *Meine Freunde, die Roboter.* München 1982: Heyne, S. 67

Hubig, Chr. (Hrsg.): *Ethische Ingenieursverantwortung – Handlungsspielräume und Perspektiven der Kodifizierung.* Sigma, Berlin 2003

Kornwachs, K.: *Das Prinzip der Bedingungserhaltung.* Lit, Münster, London 2000

Kornwachs, K.: Ebenen der Orientierung. Zur Analytik des normativen Hintergrundes. In: Hubig, Chr. (Hrsg.): *Ethische Ingenieurs-verantwortung – Handlungsspielräume und Perspektiven der Kodifizierung.* Sigma, Berlin 2003, S. 31-49 und S. 105-130

Laßmann, Gunther: *Asimovs Robotergesetze.* Heise Medien, E-Book Dez. 2017

Lem, Stanislaw: *Robotermärchen.* Suhrkamp, Frankfurt A.M. 2003?

Lenk, H., Ropohl, G. (Hrsg.): *Technik und Ethik.* Reclam jun. Stuttgart 1987

Mynika, James et al. (McKinsey Global Institute. ed.): *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation.* December 2017

Randow, Gero von : *Kopfgeburt.* In: DIE ZEIT ZEIT Nr. 1 vom 29.12.2016, ed. 14.1.2017 www.zeit.de

VDI: *Ethische Grundsätze des Ingenieurberufs.* In: <http://www.vdi.de/bildung/ethische-grundsaeetze/ethische-grundsaeetze/>

Weber, Karsten: Autonomie und Moralität als Zuschreibung – Über die begriffliche und inhaltliche Sinnlosigkeit einer Maschinenethik. In: Rath, Matthias; Krotz, Friedrich; Karmasin, Matthias (Hrsg.): *Maschinenethik. Normative Grenzen autonomer Systeme.* Springer, VS Wiesbaden, 2019, S. 193

Loh, Janina: *Roboterethik. Eine Einführung.* Suhrkamp, Berlin 2019

Prof. Dr. phil.habil. Dipl. Phys.

Klaus Kornwachs

Former Chair for Philosophy of Technology, BTU Cottbus

Humboldt Study Center for Humanities, University of Ulm

Honorary Professor, China Intelligent Urbanization Co-Creation Center at
Tongji University, Shanghai

Office for Culture and Technology

Strickers Höhe 22

D-88260 Argenbühl

Germany

Phone: ++49-7566-2498, Fax: -1891

Mobil: ++49-175-4122079

klaus.kornwachs@uni-ulm.de

klaus@kornwachs.de

Home Page

www.kornwachs.de

