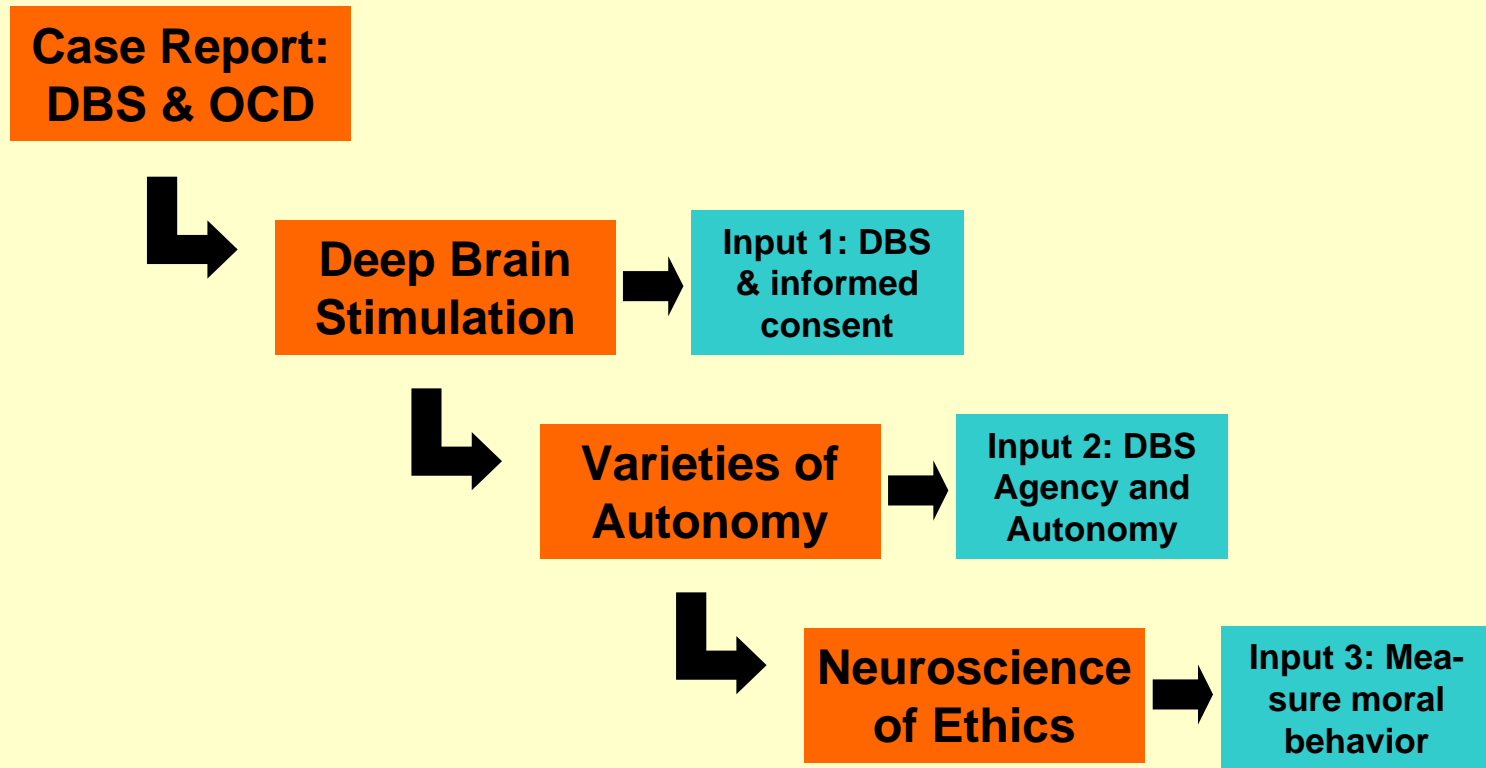# Autonomy and Moral Agency

Is our Brain More than a Complex System?

An Investigation at the Borderline of Philosophy, Neuroscience and Medicine

Markus Christen,
University Research Priority Program Ethics
University of Zürich

Roadmap

Case Report: DBS & OCD

↳

Deep Brain Stimulation → **Input 1: DBS & informed consent**

↳

Varieties of Autonomy → **Input 2: DBS Agency and Autonomy**

↳

Neuroscience of Ethics → **Input 3: Measure moral behavior**

Case report (1)

Step 1: Technical brain interventions that result in remarkable behavioral and personal chances:

A point "pro autonomy" or "contra autonomy"?

Case report (2)

Case 3. Case 3 is a single woman of 38 years with onset of symptoms at 16 years and a family history positive for OCD. She worked in administration until the age of 32 but lost her job because of the severity of her symptoms. Becoming completely non-functional she returned home to live with her parents. The main symptoms are intrusive sexual thoughts and impulses; excessive toilet, washing and counting rituals; compulsive buying and the urge to touch. She was preoccupied by the thought that objects and things might not be real, might not exist, and had a whole series of compulsions to ensure that they were really there. She fulfilled criteria for dependent personality disorder with a pattern of a constrictive, self-pitying, helpless and anxious patient, utterly dependent on significant others.

Case report (3)

Immediately after stimulation her attitude changed remarkably. From the first week, she was less anxious and felt more relaxed and easy-going. Her parents conclude: 'If she had made a pilgrimage, we would have thought a miracle had happened'. Some remaining compulsions hardly interfered with social activities and she was able to control and resist them a lot better. Relief of symptoms was markedly clear and reproducible by switching DBS on. With stimulation she was very active during the day. She reported feeling more cheerful and bustling, experiencing more excitement, talking louder and faster and laughing more. She sometimes seemed disinhibited, but did not engage in risky, reckless or impulsive behaviours.

Gabriëls et al. Acta Psychiatr Scand 107 (2003): 275-282

Deep Brain Stimulation

**Step 2: Learning some more about Deep Brain Stimulation.**

What is it, and how does it work?

DBS – Operation mode

Key elements:

- Based on pacemaker-technology (Medtronics has basically a monopoly)

- Targets: Several subcortical nuclei, dependent on disease (e.g. subthalamic nucleus), which are relatively precisely addressed.

- Stimulus parameters: monopolar cathodic square pulses (1-5 V amplitude, 60-200 ms duration, 120-180 Hz frequency), usually derived in a trial-and-error process.

- Mechanism on cellular level is unclear.

- Chirurgical intervention is relatively safe. Local chirurgic intervention for battery change (after several month, dependent on stimulation)

- Patient may control stimulator to some extend

Source: Medtronic / DANA-foundation
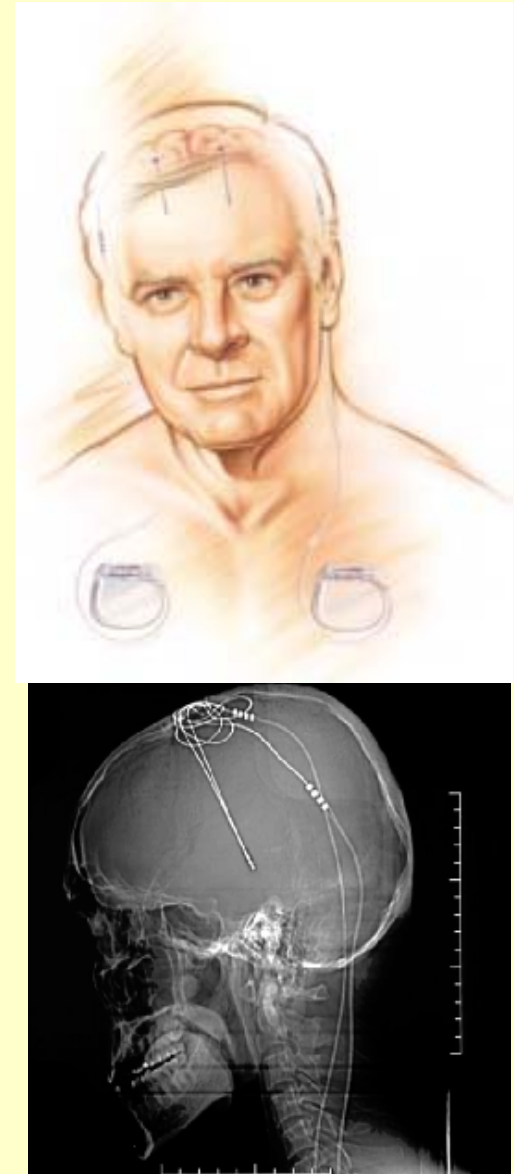
DBS – Applications (1)

General notes:

- Historical roots: mid 20th century (Penfield, Delgado)

- Today's technology has been developed in the 1980s, first for the therapy of movement disorders (Parkinson, Dystonia, etc.)

- Up to end 2006: more than 35'000 patients are equipped with DBS (Hardesty & Sackeim 2007)

- In the last few years, further applications have been tested in experimental studies:
  - Alzheimer (memory enhancement)
  - Anxiety disorders
  - Autism
  - Depression
  - Epilepsy
  - Multiple sclerosis
  - Obesity
  - Obsessive-compulsive disorders
  - Tourette syndrom

DBS – Applications (2)

Problems discussed in the scientific literature:

- Generally: good results for pharmaceutically resistant movement disorders.

- DBS failures (for movement disorders) are mainly caused by misplacement (48% of cases).

- Sudden "on-off-effects" appear, as well as changes on a longer timescale (several weeks to months, indicating lack of knowledge on physiological mechanisms).

- Dispute on increased suicide-risk (Burkhard et al. 2004; Albanese et al. 2005; Foncke et al. 2006).

- Dispute on validation of „quality-of-life" after intervention (Diamond & Jankovic 2005).

- **Anecdotic reports on complex behavioral changes in patients.**

Input 1 (9 Slides)

The Problem of Informed Consent Regarding Technical Interventions in the Brain

(together with Jan Heilinger, Berlin)

Input 1

The patient had to be at least 18 years, and no more than 60 years of age. He must be able to understand, comply with instructions and provide his own written informed consent. The patient and a close family member were repeatedly and fully informed on both procedures (capsulotomy and DBS). The standard risks for DBS, known from its use in Parkinson's disease (PD) were explained. Criteria excluded a current or past psychotic disorder, any clinically significant disorder or medical illness affecting brain function or structure (other than motor tics or Gilles de la Tourette syndrome) or current or unstably remitted substance abuse. If the patient did not improve after 1 year of DBS, the option of anterior capsulotomy would be reconsidered.

Input 1

Informed consent
is the *process* by which a patient can participate in choices about his healthcare (surgical or medical procedure, inclusion in a clinical study). It originates from the legal and ethical rights the patient has to direct what happens to his body and from the ethical duty of the physician to involve the patient in decisions about his healthcare.

The physician provides to the patient:
- Information about the procedure and the decision at hand
- Explanation of alternatives
- Facts about risks, benefits, and uncertainties of all alternatives

$\rightarrow$ *acceptance/rejection of the intervention by the patient*

Input 1

- Patient is understanding
- Patient is competent to make decisions
- Patient makes his choice voluntary and autonomously – not forced

$\rightarrow$ ***Risk*** *of forced choice, subtle coercion by unfavorable circumstances*

Input 1

In some (psychiatric) cases, the patient is:

- not able to understand his situation
- not able to understand the risks and benefits
- not able to decide rationally
- not able to communicate his decision

$\rightarrow$ ***The capacity of decision-making is variable, as lucid states and disordered states alternate.***

Input 1

- Presumed consent has to be figured out with a surrogate decision maker (near relatives, care giver, medical attendant)

- Legitimization of an intervention has to be postponed, until ability to decide is restored

    $\rightarrow$ *Analogy to the debate about the use of psycho-pharmaceutics*

Input 1

Normative
level

Empirical
level
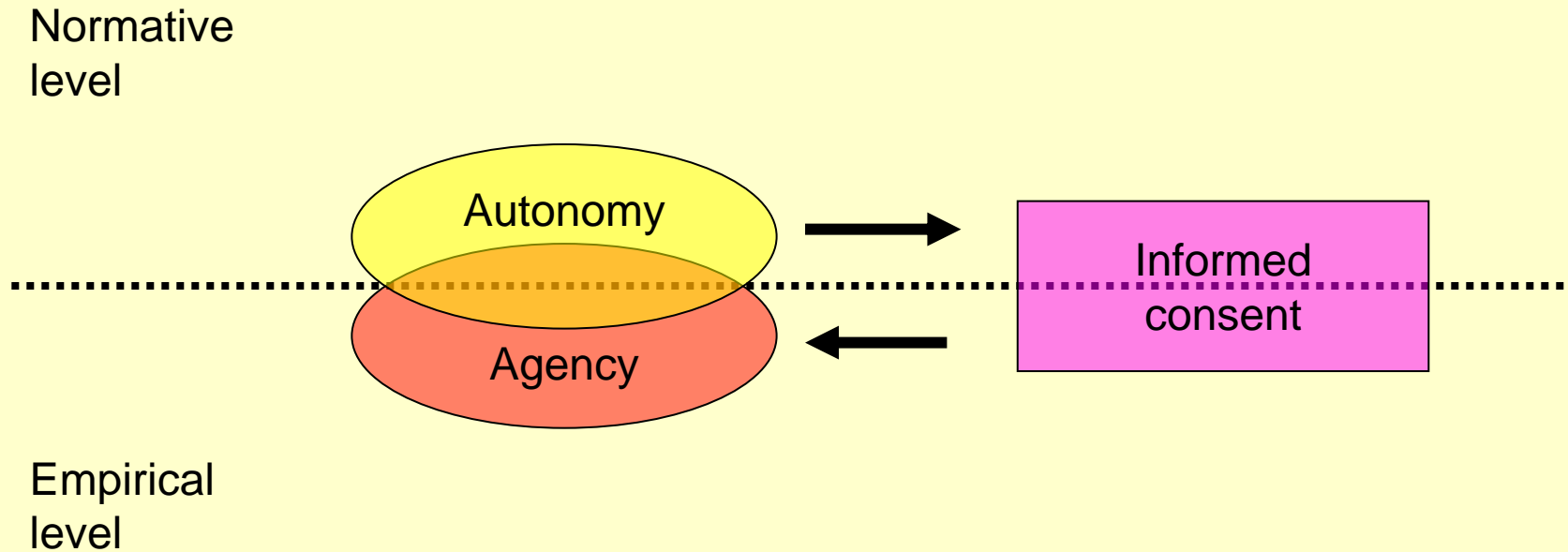
→ *Autonomy as a prerequisite of informed consent.*

→ *Informed consent as an expression of agency.*

Input 1

Movement disorders

Autonomy → Informed consent

Agency ← DBS

Psychiatric disorders

Autonomy / Agency ⇢ Informed consent
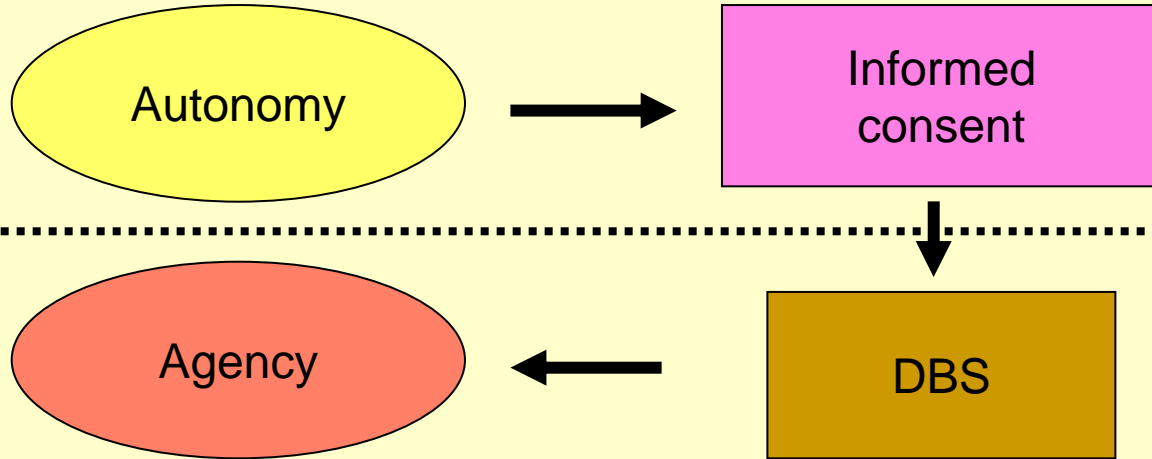
Agency ⇠ DBS
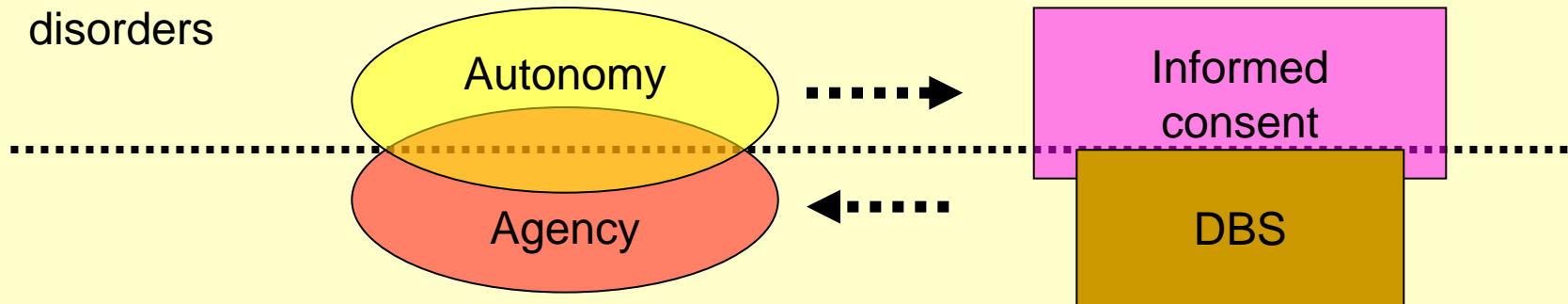
Input 1

Can DBS improve autonomy and agency?

The optimistic philosopher:

Autonomy: Yes!
- allows for 2nd order volitions (e.g. avoiding depression, intrusive thoughts) and thus augments of decision-making capacity
- allows for lucid (instead of disordered) states of mind, thus allows for ex post justification of some interventions (informed consent ex post).

Agency: Yes!
- Movement disorders: Tremor is gone
- Psychiatric disorders: Allows for active life

Input 1

Can DBS improve autonomy and agency?

The skeptical philosopher: skeptical of the clear separation between the empirical and normative levels in psychiatric disorders.

Before intervention
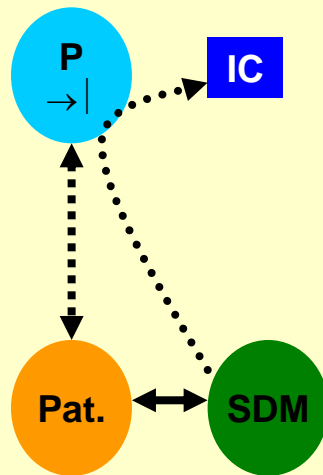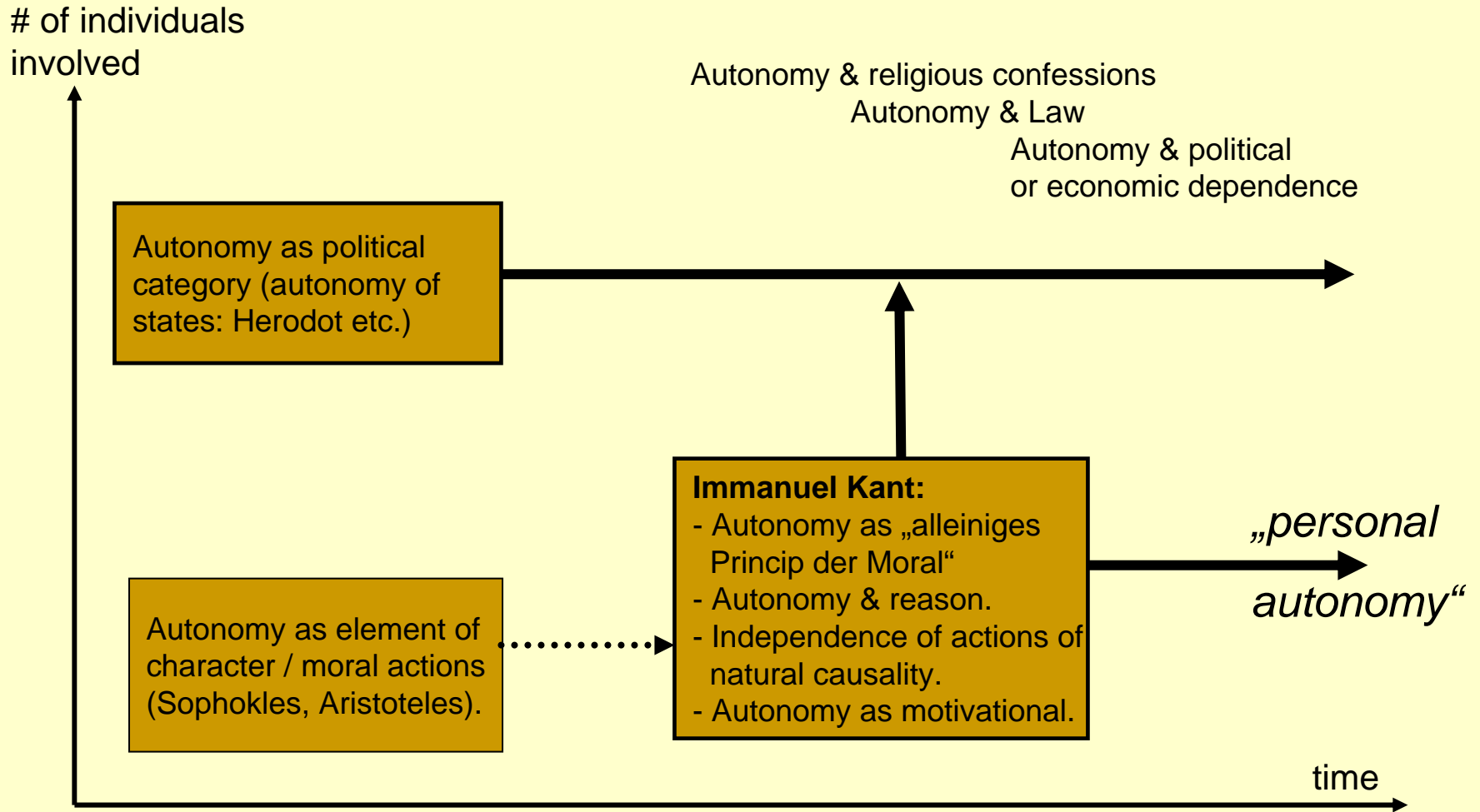
After intervention
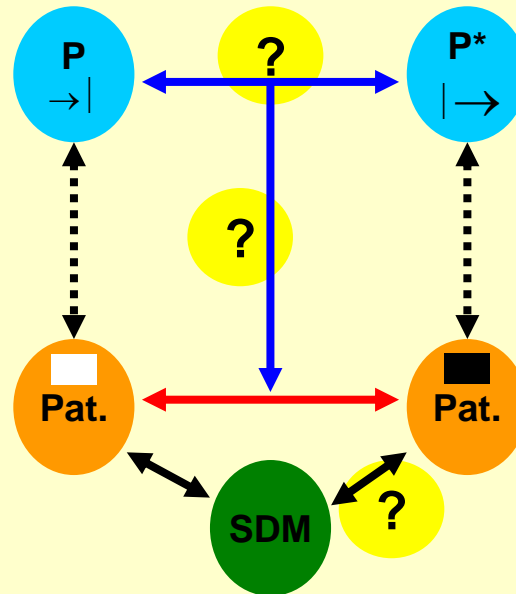
Varieties of Autonomy

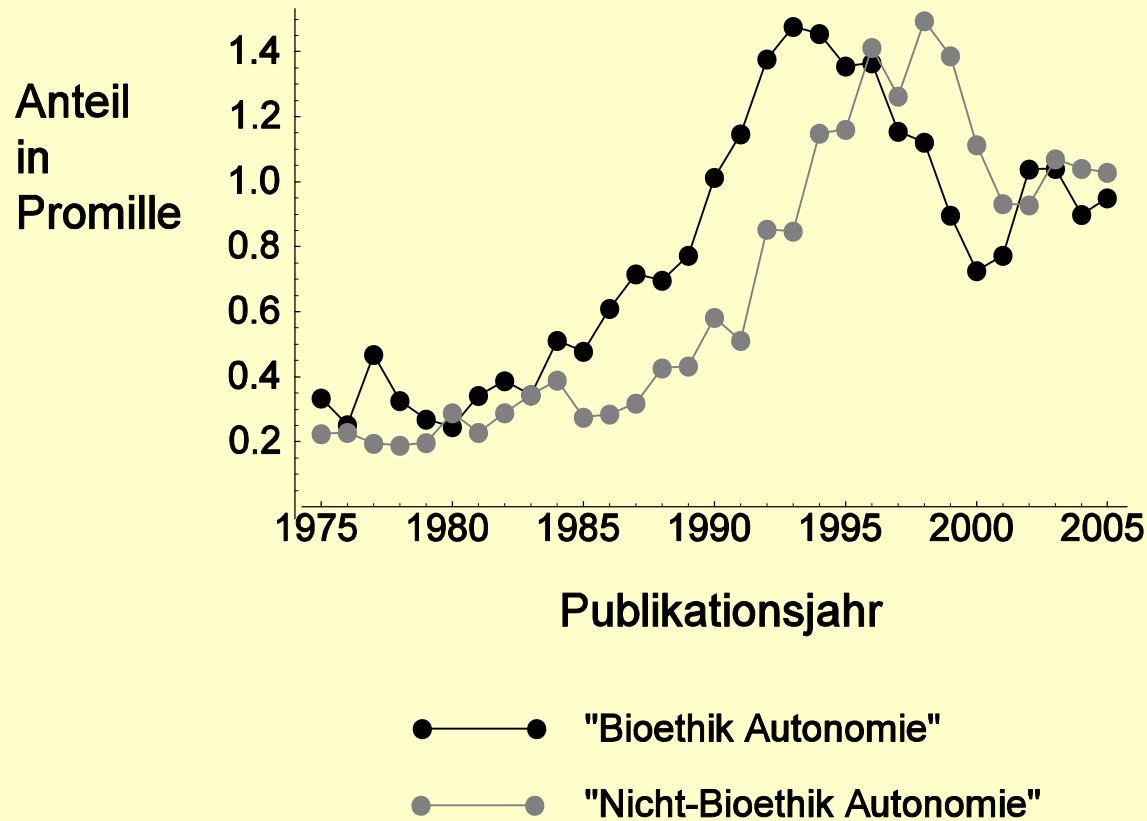**Step 3: Varieties of autonomy in philosophy and empirical sciences.**

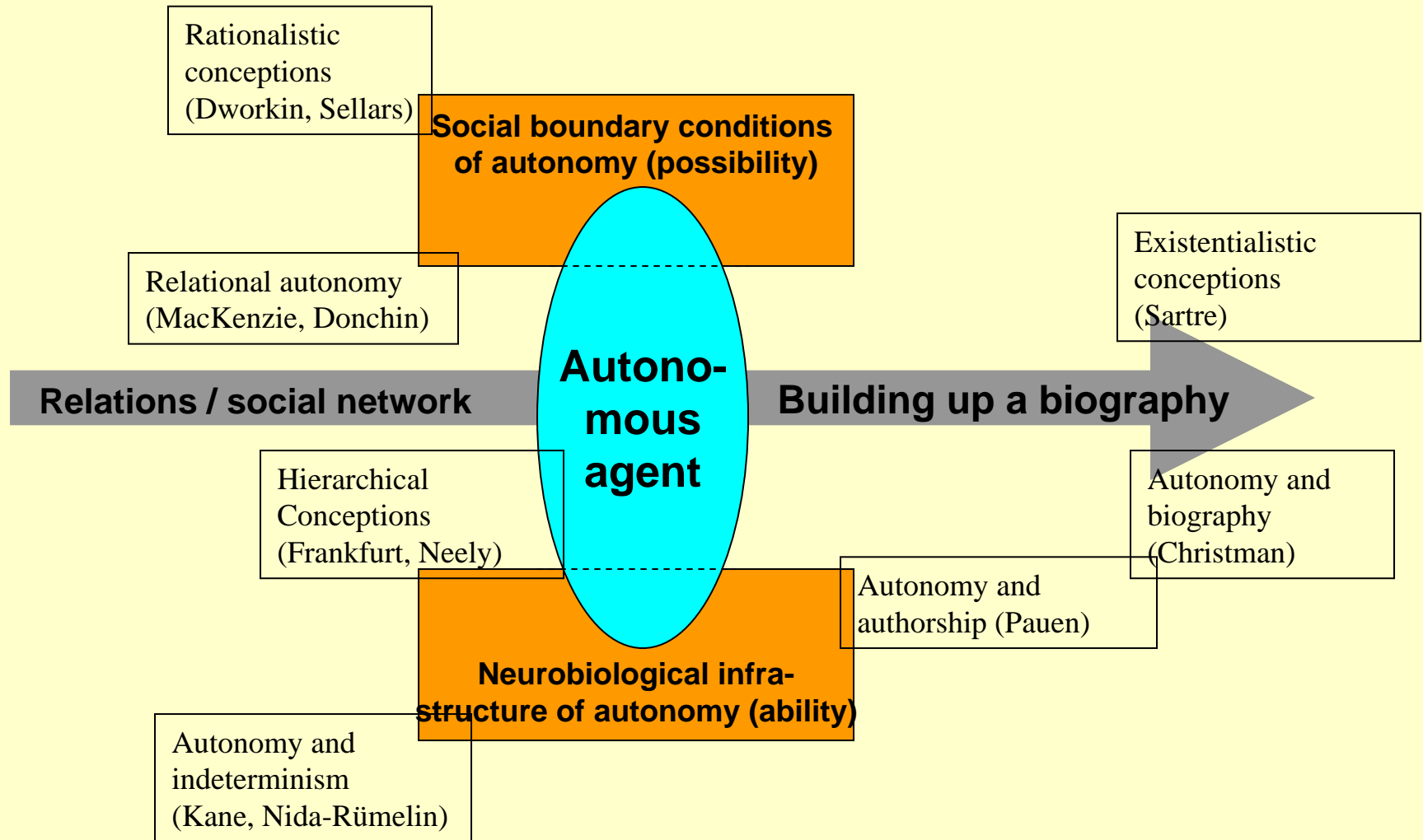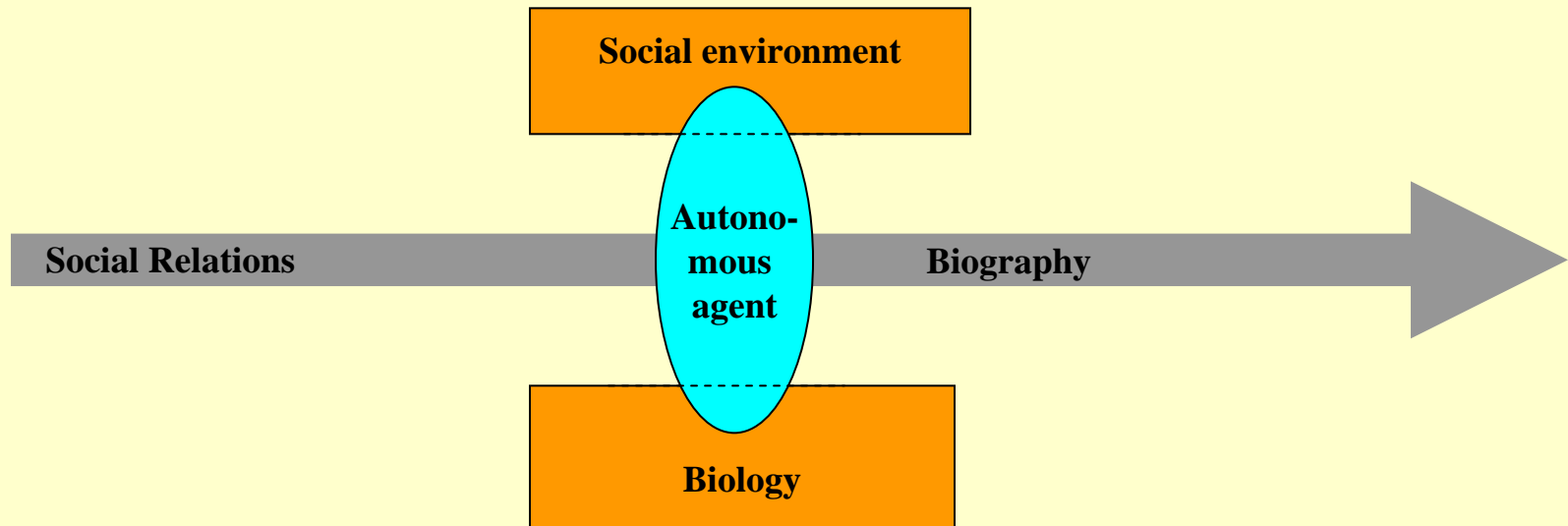**Is there a fit (or are there at least some bridges)?**

Autonomy – History

of individuals involved

Autonomy & religious confessions
Autonomy & Law
Autonomy & political
or economic dependence

Autonomy as political category (autonomy of states: Herodot etc.)

Autonomy as element of character / moral actions (Sophokles, Aristoteles).

Immanuel Kant:
- Autonomy as „alleiniges Princip der Moral"
- Autonomy & reason.
- Independence of actions of natural causality.
- Autonomy as motivational.

„personal autonomy"

time

Autonomy – Impact

Bibliometric analysis based on PubMed (NIH): 1975-2005. Number of publications including „autonomy" (bioethics / rest) normalized by total number.
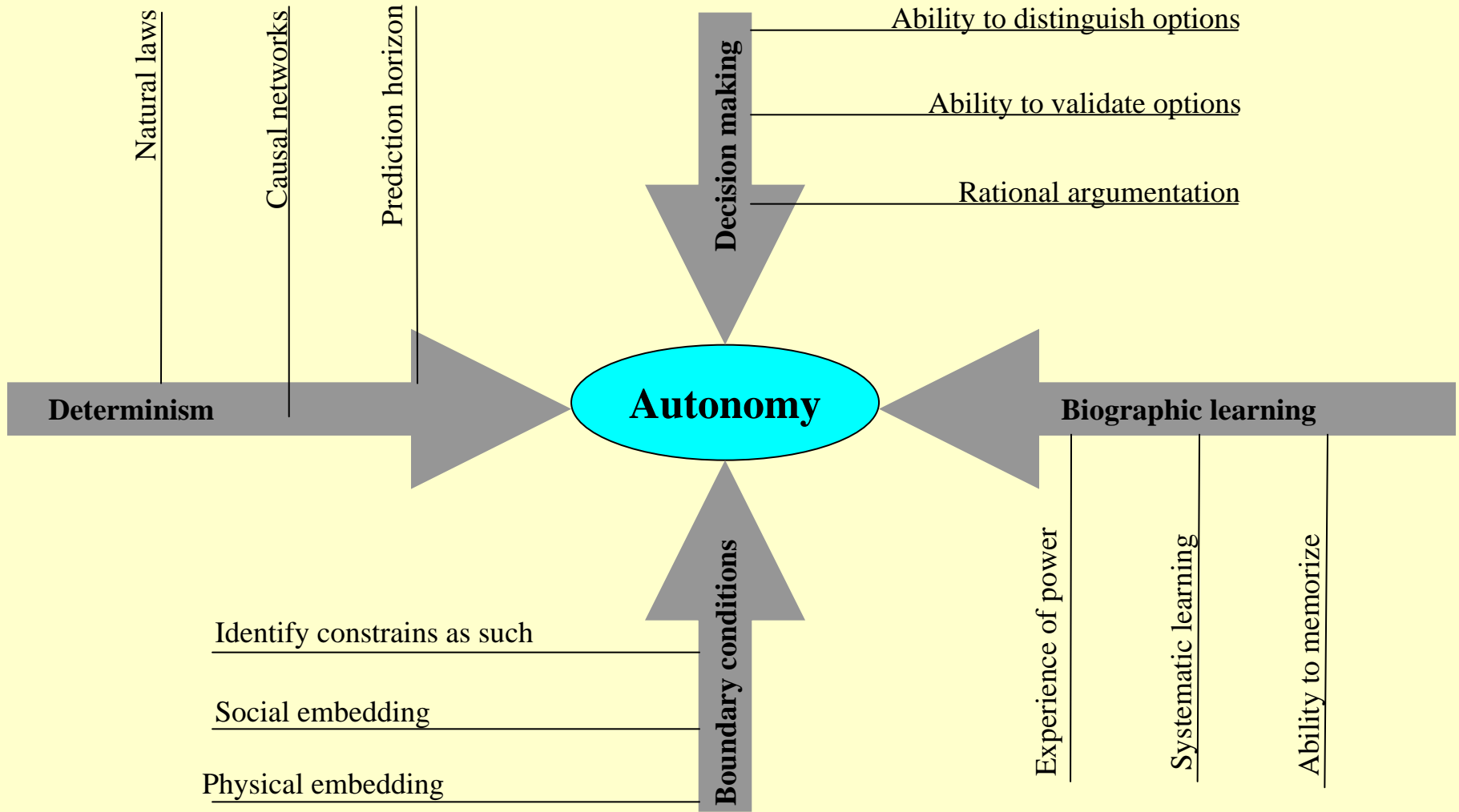
Varieties of Autonomy

Rationalistic conceptions (Dworkin, Sellars)

Social boundary conditions of autonomy (possibility)

Relational autonomy (MacKenzie, Donchin)

Existentialistic conceptions (Sartre)

Relations / social network

Autono-mous agent

Building up a biography

Hierarchical Conceptions (Frankfurt, Neely)

Autonomy and biography (Christman)

Autonomy and authorship (Pauen)

Neurobiological infra-structure of autonomy (ability)

Autonomy and indeterminism (Kane, Nida-Rümelin)

Dimensions of Autonomy

Four dimensions of autonomy:

- Decision-making: Ability to decide on „important issues".

- Agent-history: Learning, memories, relations.

- Boundary-conditions: On the time-scale of acting

- Determinism: Relative to the internal processes of the agent.

Approaching Autonomy

Natural laws

Causal networks

Prediction horizon

Decision making

Ability to distinguish options

Ability to validate options

Rational argumentation

Determinism

Autonomy

Biographic learning

Boundary conditions

Identify constrains as such

Social embedding

Physical embedding

Experience of power

Systematic learning

Ability to memorize

Naturalizing Autonomy

The medical problem of finding markers for autonomy-ability:
- Neurodegenerative diseases
- Informed consent in special conditions
- Temporal aspects (Patienten-verfügung

Analysis of social systems in order to find determining elements (social engineering)

Social environment

Relations

Auto-nomous agent

Auto-nomous system

"Biography"

Biology

Technology

Finding the "neurobiological basis" of:
- decision making
- moral agency
- Concepts related to autonomy: cooperation, trust

Construction of autonomous systems with specific properties: unsupervised learning, self-repair, self-organization, autarky.

Epistemic principle of cybernetics: (Wiener): understanding=building

Autonomy and Neuroscience

„Worum es letztlich geht ist die *Autonomie menschlichen Handelns*, nicht Willensfreiheit. Autonomie ist die Fähigkeit unseres ganzen Wesens, *innengeleitet*, aus individueller Erfahrung heraus zu handeln, und zwar gleichgültig ob bewusst oder unbewusst." (G. Roth, 2003)

In neuroscience, the concept of autonomy is more attractive, as it allows a better integration into an empirical/experimental approach compared to the "freedom of the will".

„Innengeleitet" has the interpretation of "brain-based" in the sense that all processes that allow the agent to act autonomously have their basis in the brain.

Be aware that the concept of "acting" and "experience" are different than the ones used in philosophy.

Autonomy and acting – a simple view
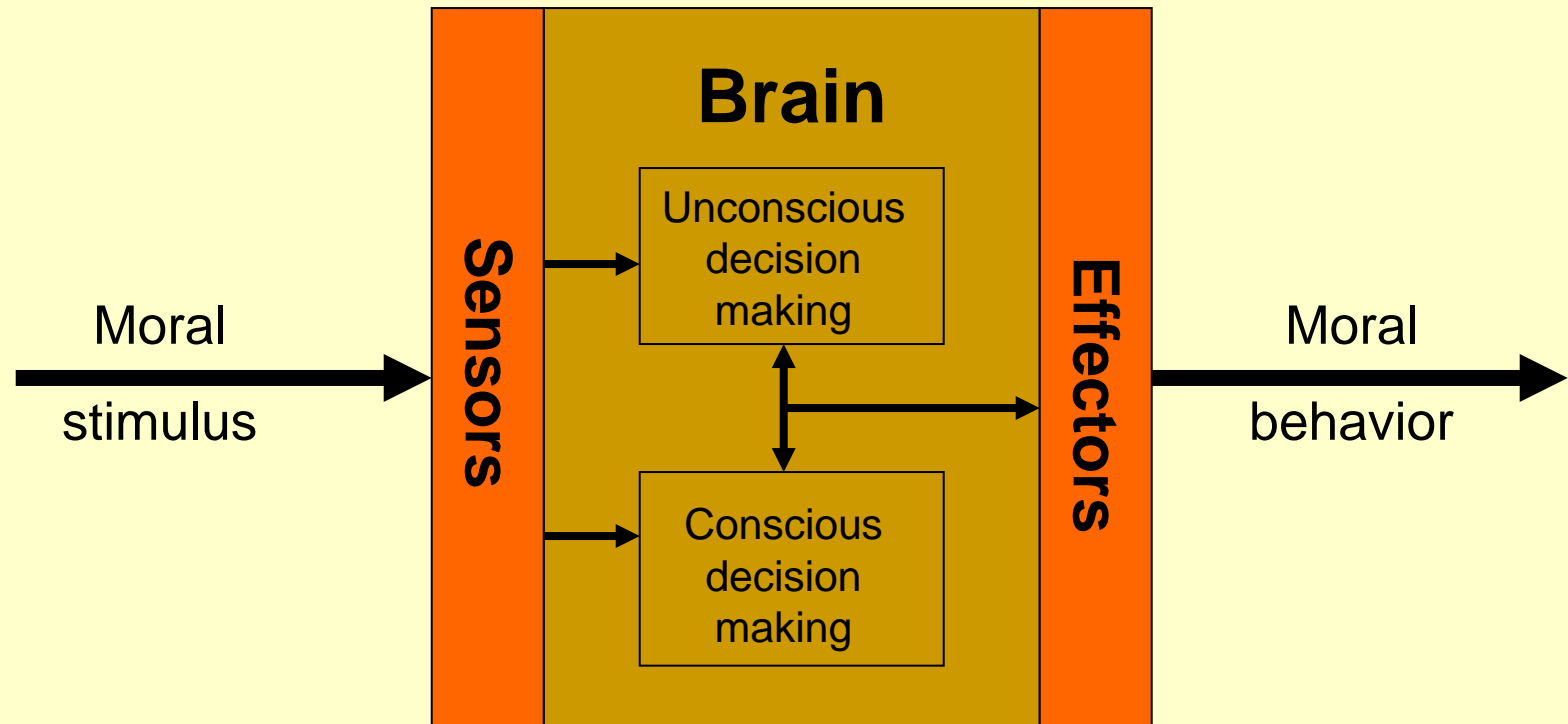
The moral agent – a bit more complex
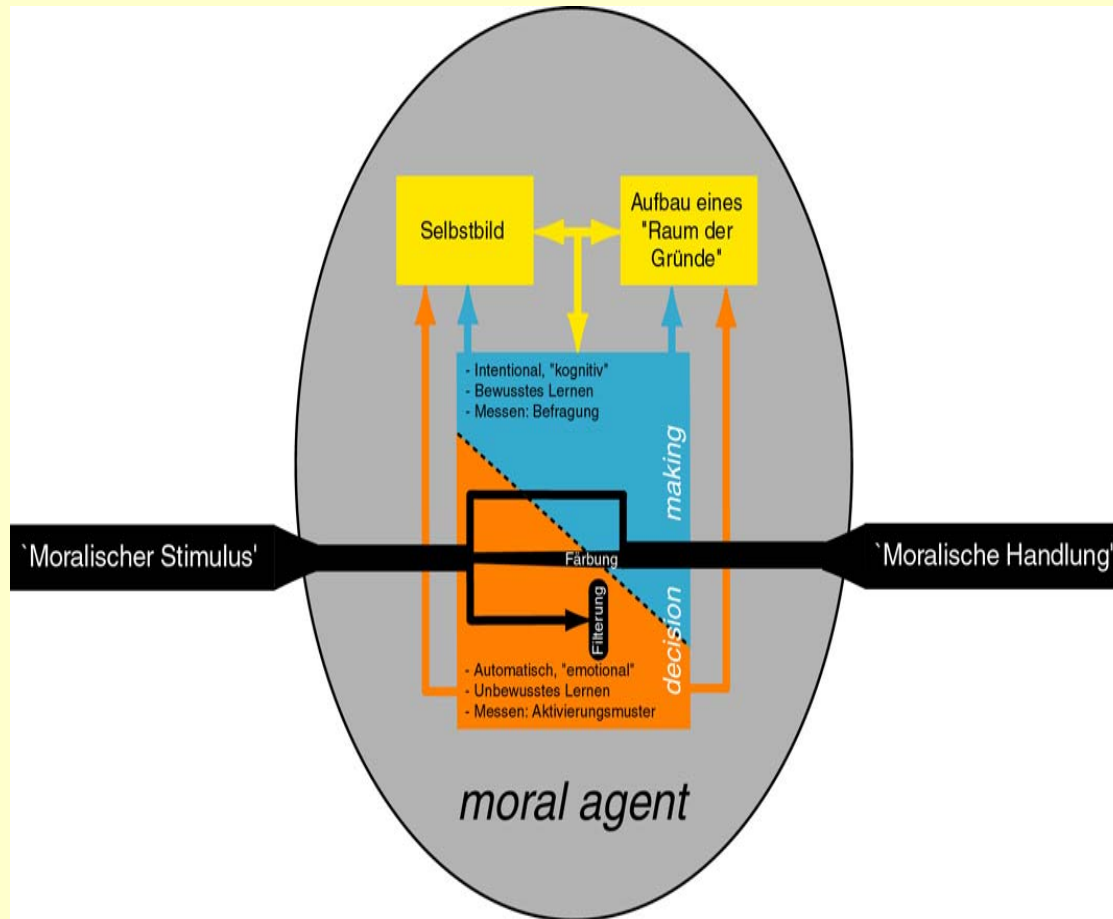
Four structural components of moral agency:

- Experience (of moral stimuli)

- Decision-making (consisting of intentional and automatic processes)

- Acting (impact on space/ time, the social world)

- Justification (the agent may give, when asked).

Strong Autonomy

The "strong" concept of autonomy refers to the ability to take an external perspective towards the own "sequence of reasons" that cause an act in order to be able to change this sequence and make a different act (not the same as "alternativism"). In our model of a moral agent, autonomy allows to

... move in the space of reason
... develop a representation of the "self"
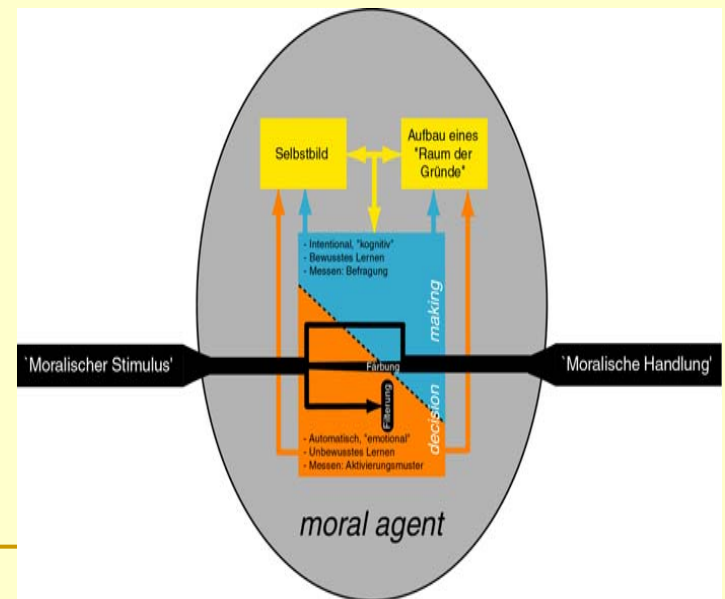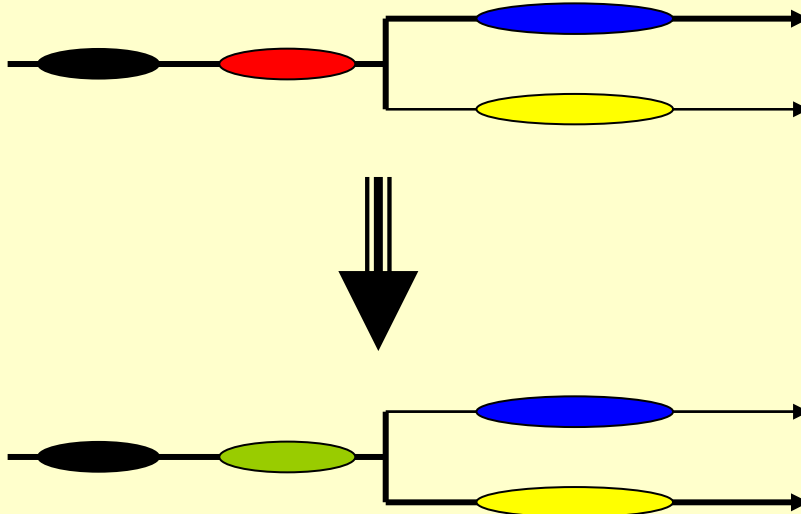
Weak Autonomy

Natural sciences use a concept of autonomy that is (more) compatible with determinism: Autonomy is the ability of a system to react to changed environmental conditions by a changed behavior, whereas learning may improve this adaptation of behavior.

Simple dynamical systems (given as deterministic differential equations) may serve as a model for this understanding of autonomy. Changed environmental conditions (e.g. a limiter in the state space) may lead to a changed dynamics (e.g. a limit cycle of periodicity 2 changes to a limit cycle of periodicity 3).
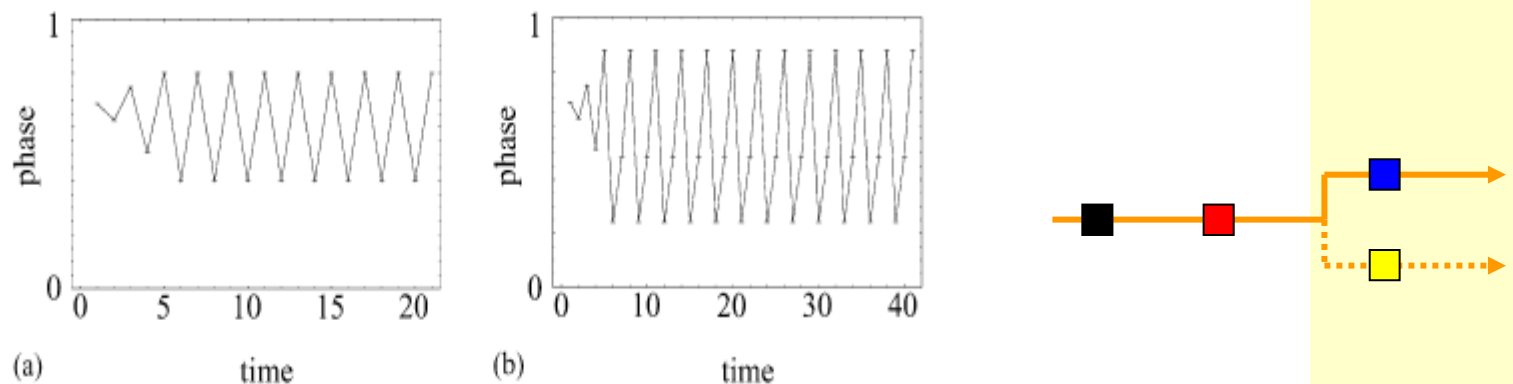
Fig. 1. Autonomous gait generator. By reducing the load, the gait changes from a period 2 (a) to a period 3 (b). The system consists of a chaotic tent map, where the load is implemented by a horizontal line replacing the graph above a certain height (the limiter) [6].

Bridging strong and weak autonomy

Problem of relevance

Self

Cognition

Space of reasons

3rd-person-perspective

?

Neuronal realization
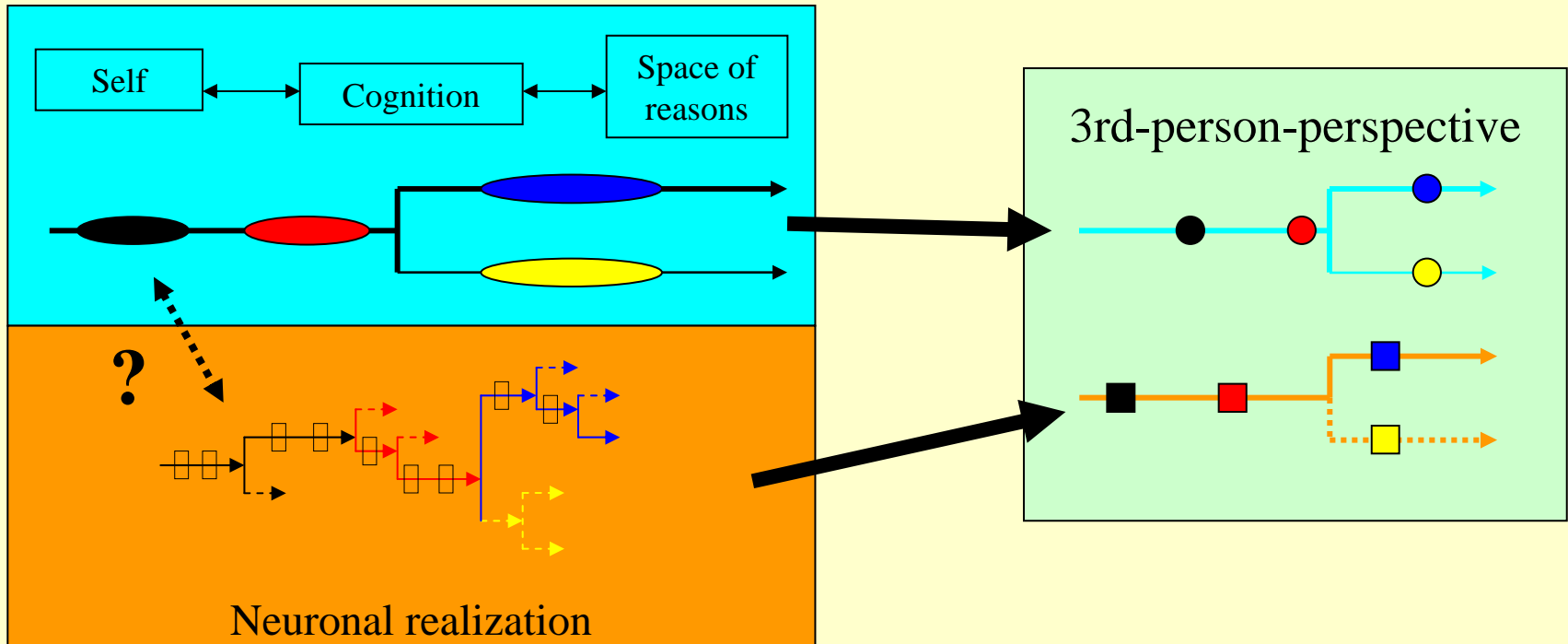
Problem of attribution

Autonomy and Morality

In Kant's approach, being autonomous includes the ability so define limitations (ethical rules/laws) for the person himself – i.e. this includes the ability to:

- discriminate rules as such
- perceive obligations as such
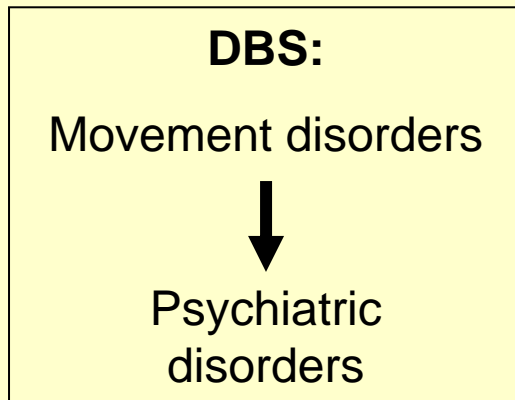- develop the idea of the "good" (and "bad")

We may thus say that being autonomous also leads to some kind of "settings" (Setzungen) that the agent does not question on the time scale relevant for actions.
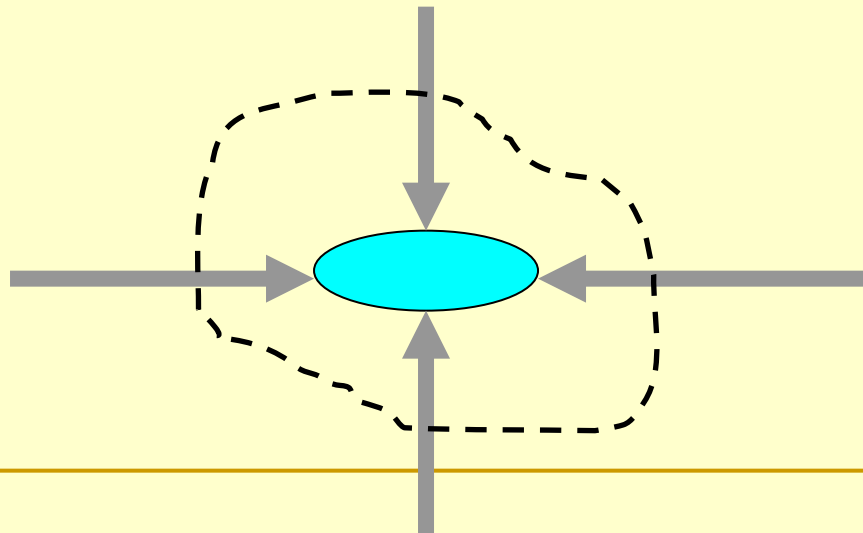
Input 2 (8 Slides)

What are the conceptual issues associated with the question, whether DBS "changes" the agency and autonomy of patients?

DBS and autonomy/agency (1)

DBS:

Movement disorders

↓

Psychiatric disorders

?

NoE:

Naturalizing moral agency

↓

„Social technology"

Out-of-control region

Serotonin levels

Dopamine levels

0

Amygdala–frontal connectivity

Prototype normal

Within the normal range

Out-of-control region
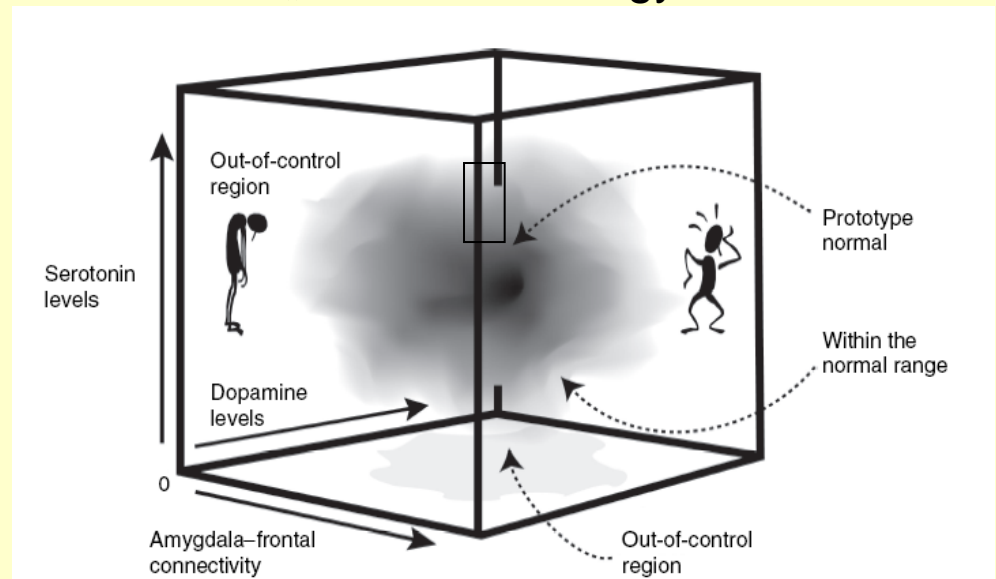
Churchland 2005

DBS and autonomy/agency (2)

In general, agency is the ability of an entity to experience its environment, to make decision based on these experiences and of inner states, and to act according to these decisions, whereas this process does not underlie strict external control.

Philosophy on agency

Autonomy
Authorship
Setting values
Setting goals based on values
Goal-oriented acting
Interactivity with other agents
Reactivity (short time scale)
Adaptation (long time scale)
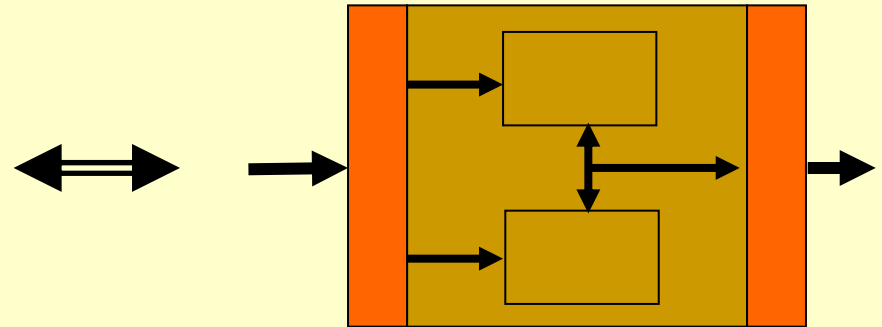
DBS and autonomy/agency (3)
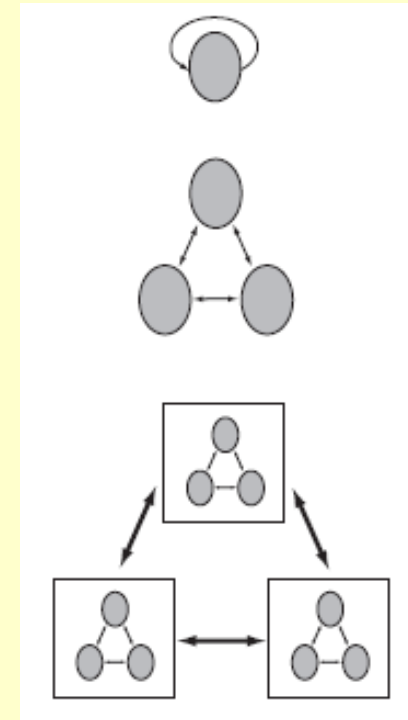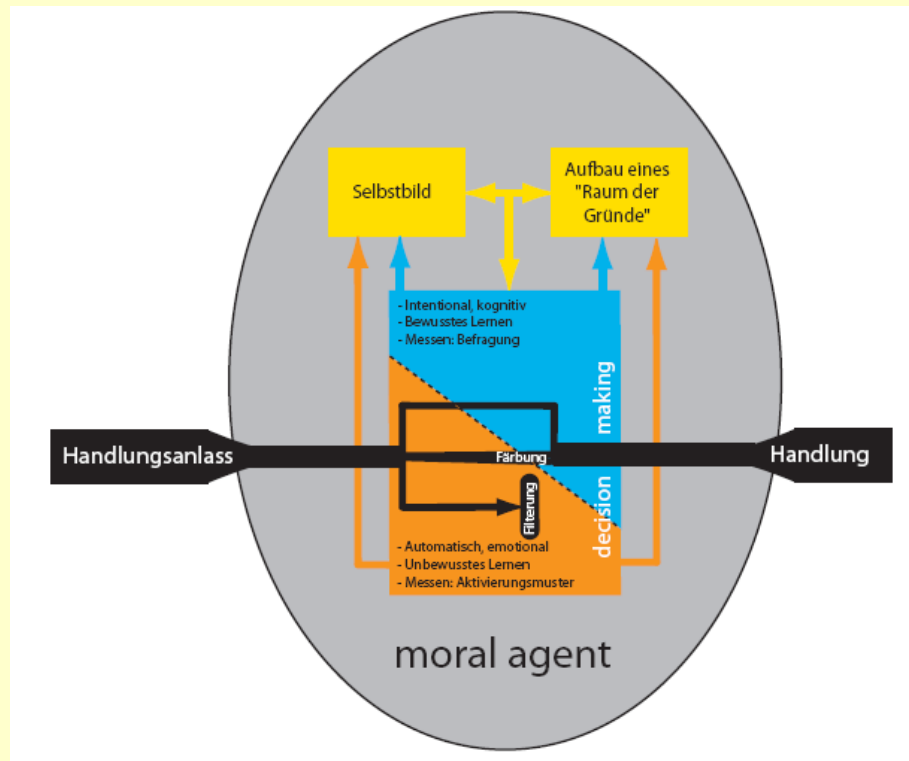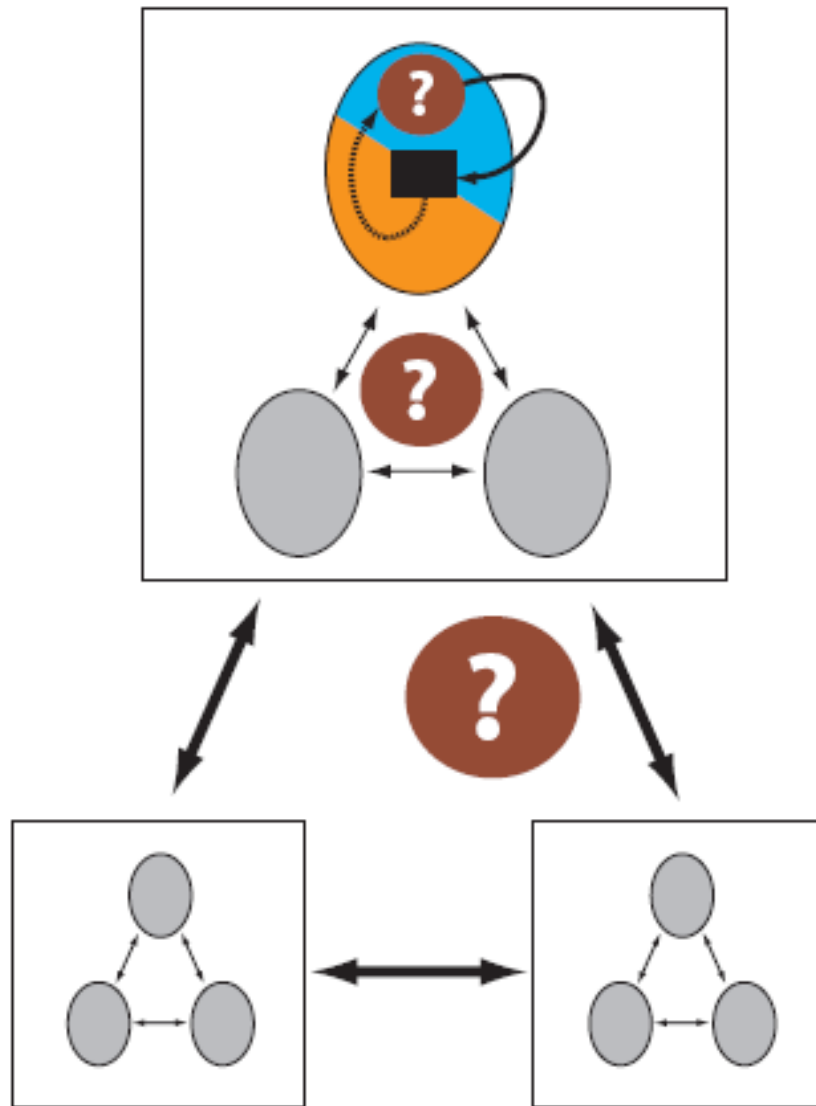
DBS and autonomy/agency (4)

Empirical investigations of moral agency require some normative determinations – but there are different ways to do this:

Normative Ethik

Deskriptive Ethik (Moral)

Normative Ethik 1

Normative Ethik 2

K_1

Moralische Kern-Normen

"weiche" Normen (Konventionen)

K_2
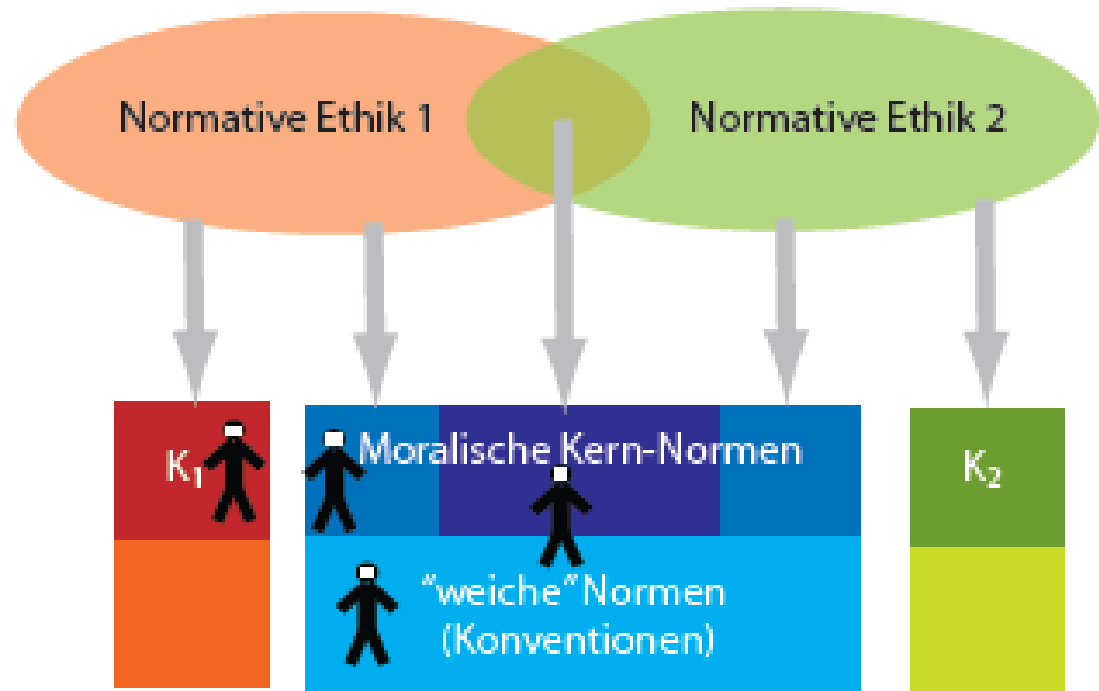
DBS and autonomy/agency (6)

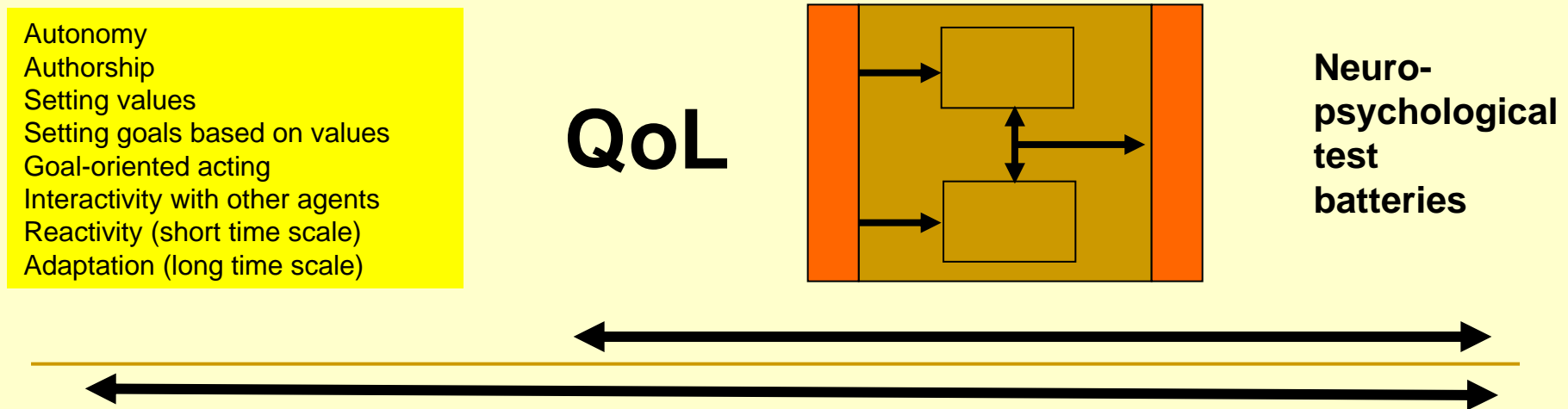|		Kontextbedingung		
		Individuum	**Beziehung**	**Gesellschaft**
Struktur von *moral agency*	**Wahrneh-mung**	Unfähigkeit, sich emotional von Wahrnehmungen berühren zu lassen. Fehlgeleitete emotionale Belegung von Wahrnehmungen. Probleme, Kausalitäten eines Wahrnehmungskomplexes zu erkennen. **Angststörungen, Autismus**	Beziehungs-spezifische Wahrnehmungsstörungen, eingeschliffene Wahrnehmungsdefizite in gewissen Beziehungskontexten.	Frage des Zugangs zur relevanten Information, kollektive Ausblendung von Missständen, die Anlass zu moralischen Debatten liefern könnten.
	Entscheidung	Unvermögen zur Wertsetzung, Priorisierung von Werten. **Depression**	Einschränkung der Optionen durch beziehungsinterne Tabus.	Fragen der Art der Entscheidungsverfahren und der Verteilung von Macht und Kontrolle innerhalb dieser Verfahren.
	Handlung	Behinderungen, welche das Einwirken auf die Welt generell erschweren (Bewegungs-/Kommunikationsstörung). **Tourette-Syndrom / Zwangsstörungen**	Peer pressure.	Fragen des empowerment, Kontrolle über Umsetzung, Reichweite dieser Kontrolle.
	Begründung	Kognitive Störungen (Gedächtnis, Zugang zur eigenen Biografie). **(Autismus? / Depression?)**	Abgekapselte ideologische Systeme (z.B. Sekten).	Problem von stabilen kollektiven Irrationalitäten.

DBS and autonomy/agency (7)

The analysis of 34 tests used to asses the success of DBS used in experimental studies for DBS reveals the following result when classified according to the degree of how these test refer to "agency" :

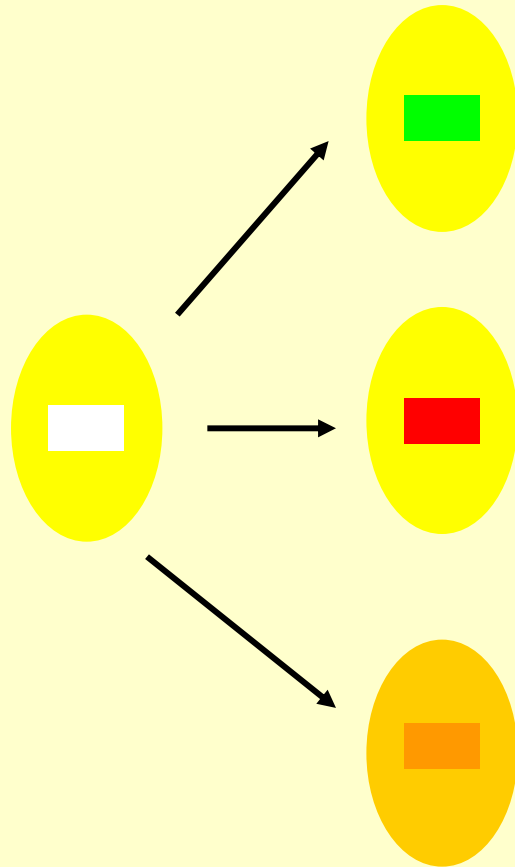Clear reference to agency: 11 (7 QoL, 2 Depr.)
Some reference to agency:10
No reference to agency: 13

Of interest is the observations, that patients often where not able to find out whether the stimulator was on or off, although they realized changes in their personal mood/behavior.

Autonomy
Authorship
Setting values
Setting goals based on values
Goal-oriented acting
Interactivity with other agents
Reactivity (short time scale)
Adaptation (long time scale)

QoL

Neuro-psychological test batteries

DBS and autonomy/agency (8)

Indifference: Moral Agency is similarly executed as before the intervention

Conflict: Internal and external experience of (negative) changes in moral agency agree.

Transformation: Internal and external experience of (negative) changes in moral agency do not agree.

→ **Two test objects:**   **- Transformation process**
                                               - Disagreement

Neuroscience of Ethics

Step 4: Within neuroscience, there is an increasing interest in finding the "neural basis" of moral behavior:

Will this lead to the "death" of autonomy?

Neuroscience of Ethics

Some bibliometry:

Basis: MedLine (1975-2005)

a) Publication activity within neuroscience (normalization).

b) Relative fraction of *imaging*, *emotion* und *social cognitive neuroscience*.

c) Relative fraction of paper on ethical/moral aspects within neuroscience (with and without papers on brain tissue transplantation).
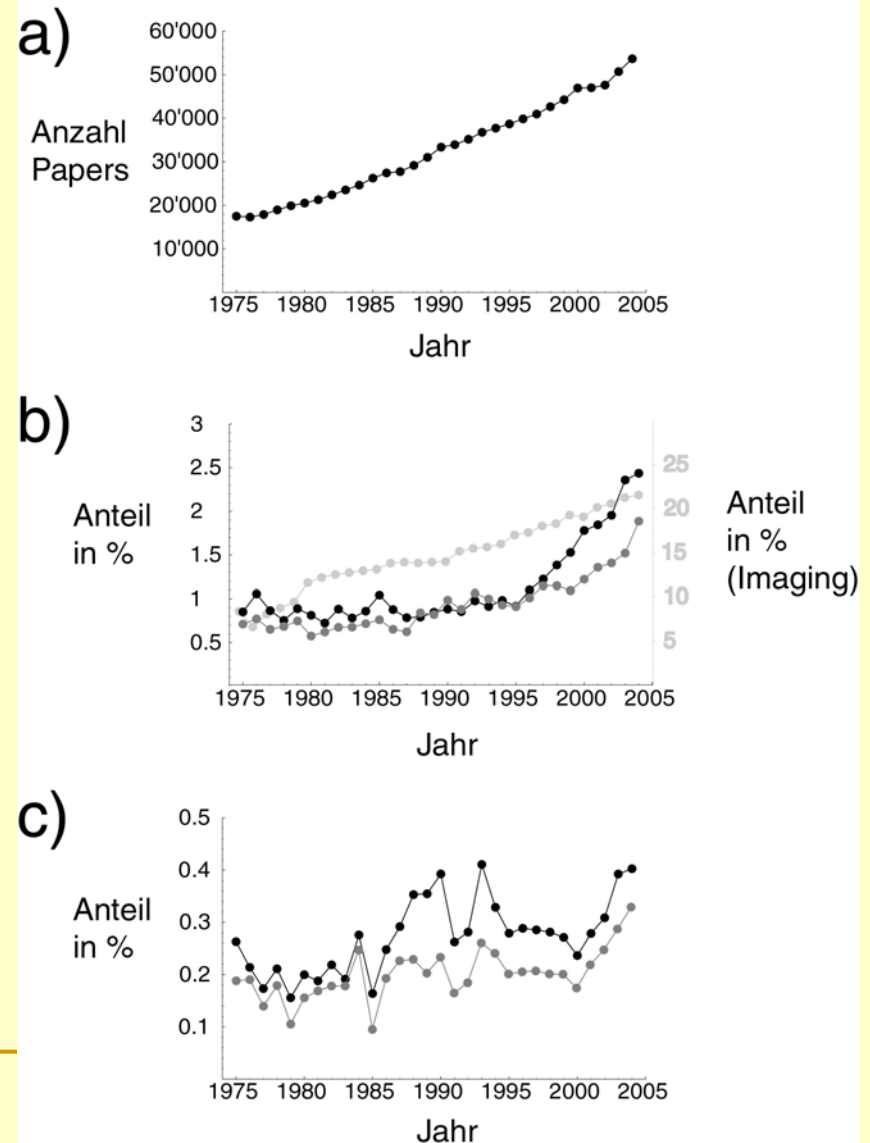
Neuroscience of Ethics

Related Fields of Research:

Social cognitive neuroscience: Neuronal „basis" of social behavior, decision making, social pain.

Mirror neurons: Imitation, learning, empathy.

Experimental economy: Strategic behavior, cooperation.

Emotion research: Role of emotions in behavior, moral emotions (guilt, shame etc.)

Moral psychology: Development of moral agency

Neuropsychopathology: „Moral pathologies".

Primatology: „Premoral behavior" in primates. Evolutionary roots of morality
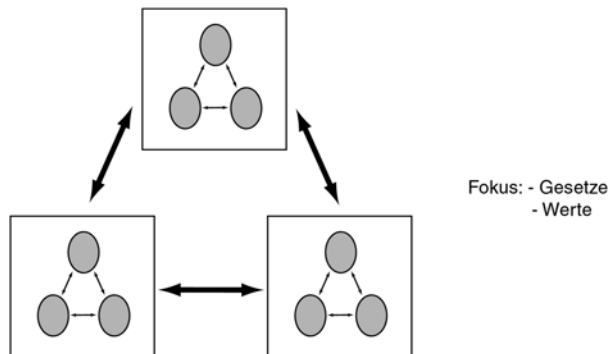
Neuroscience of Ethics
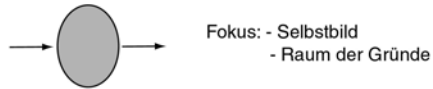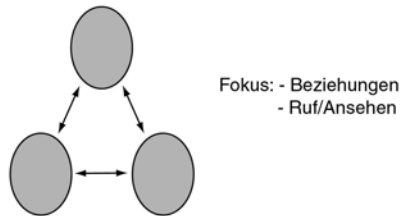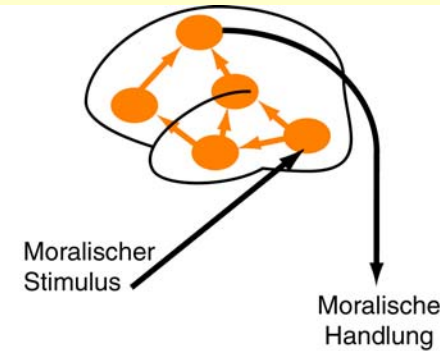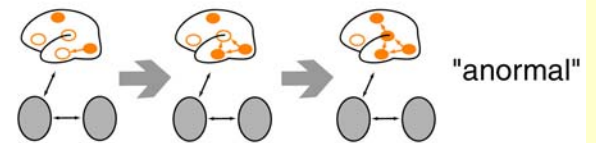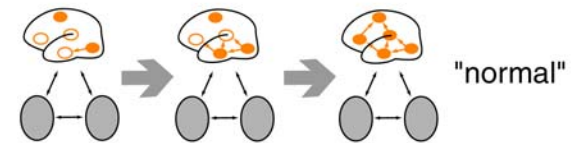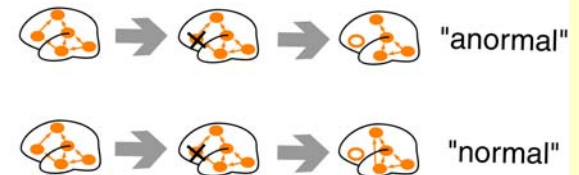
Neuroscience of Ethics

Moral Behavior: Philosophy
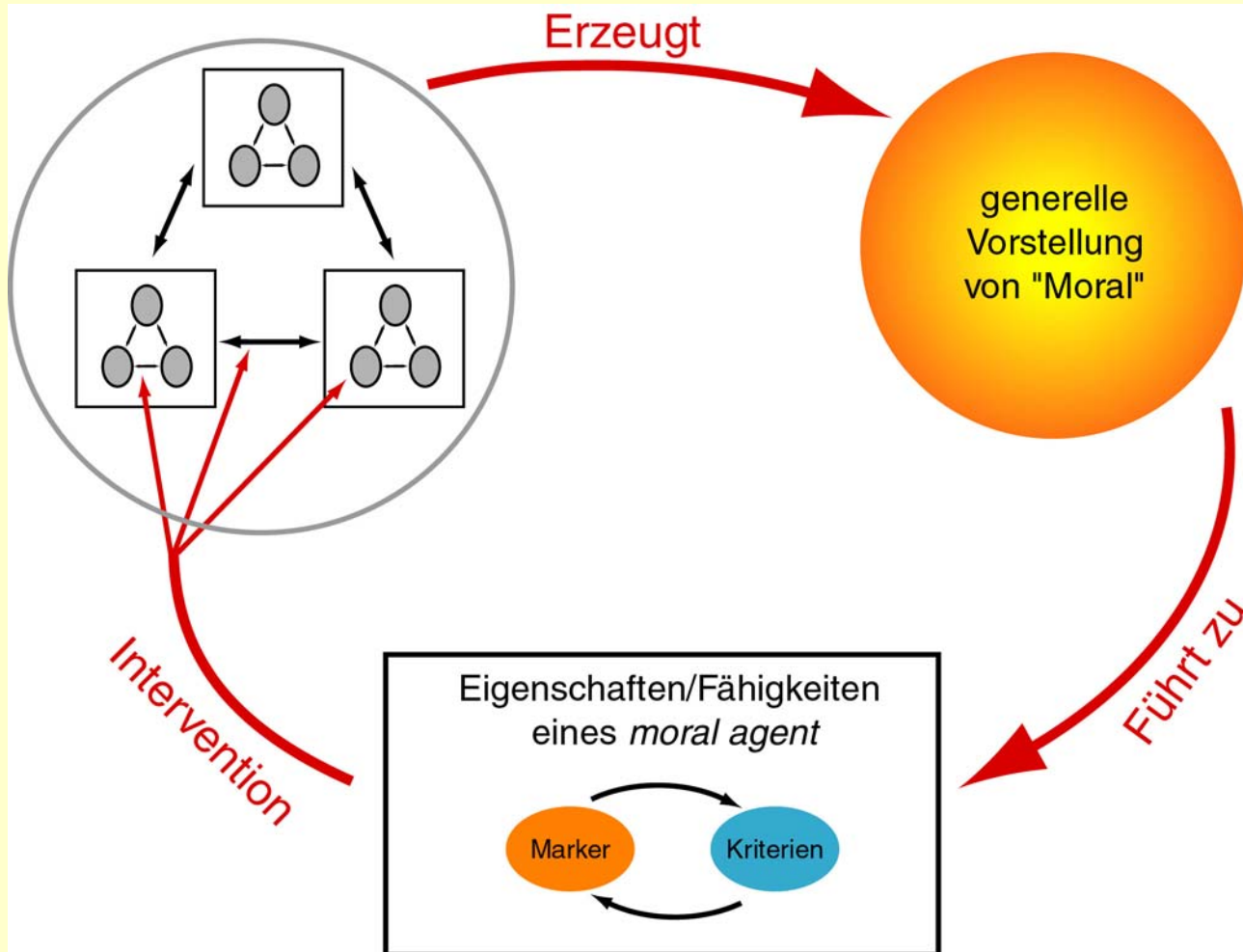
Potential moral entities:

- Persons (agents)
- Behaviors/acts (of persons)
- Ideas/opinions (of persons)
- Dispositions/virtues (of persons)
- (Implicit) norms of groups
- (Explicit) norms of social institutions
- Values of institutions/societies
- …

When is such an entity a *moral* entity?

- Two ways to answer this question: *descriptively* or *normatively*

- Moral entities are embedded in *reason-generating theories*, which have specific *characteristics*.

- Some characteristics: Reference to relevance, universality (within in-group), …

Moral Behavior: Psychology/Sociology

Characteristics of empirical approaches versus moral entities:

1) Normative issues are usually not expounded but *presupposed* (i.e. the experimental setting defines what is moral). Two strategies:
   - Either "drastic" scenarios are used (involving harm, killing etc. – i.e. exploring only a minor part of the "moral universe").
   - Or the "morality" is not explicitly addressed (i.e. one talks for example about "cooperation")

2) *Markers* that are used to define the "moral setting" of an experiment/ behavioral study include:
   - Concepts/descriptions (given orally or in written form)
   - Visual stimuli (photographs/video) inducing "moral emotions"
   - Behaviors in interactions (e.g. harm behavior)
   - The "ambiance" of an experiment (difficult to define).

3) Time scale of experiments/observations is usually short (seconds to minutes (rarely hours)).

Moral Behavior: Two Levels
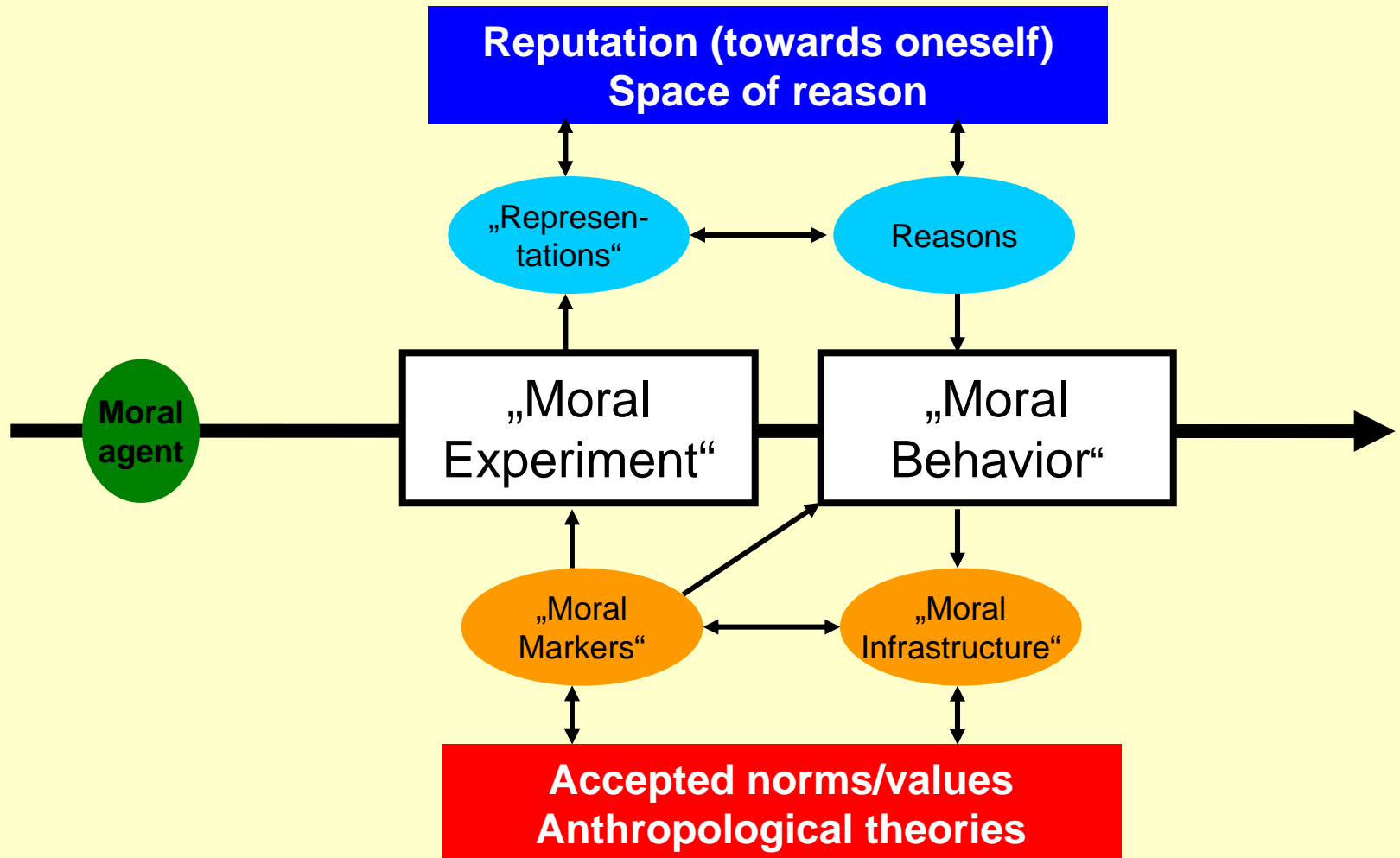
Moral Stimuli

Pictures: Mostly portraying emotionally charged, unpleasant social scenes, representing moral violations. Problem: unconsidered correlations between pictures.

Sentences: Mostly sentences of simple description with a "moral content". Problem: Are baseline-sentences free of moral connotations?

Dilemmas: *Personal* vs. *Impersonal* dilemmas (Greene) (d.h. the degree of body involvement). Problem: what exactly is measured (as the decision process goes over many seconds).

The "moral brain"
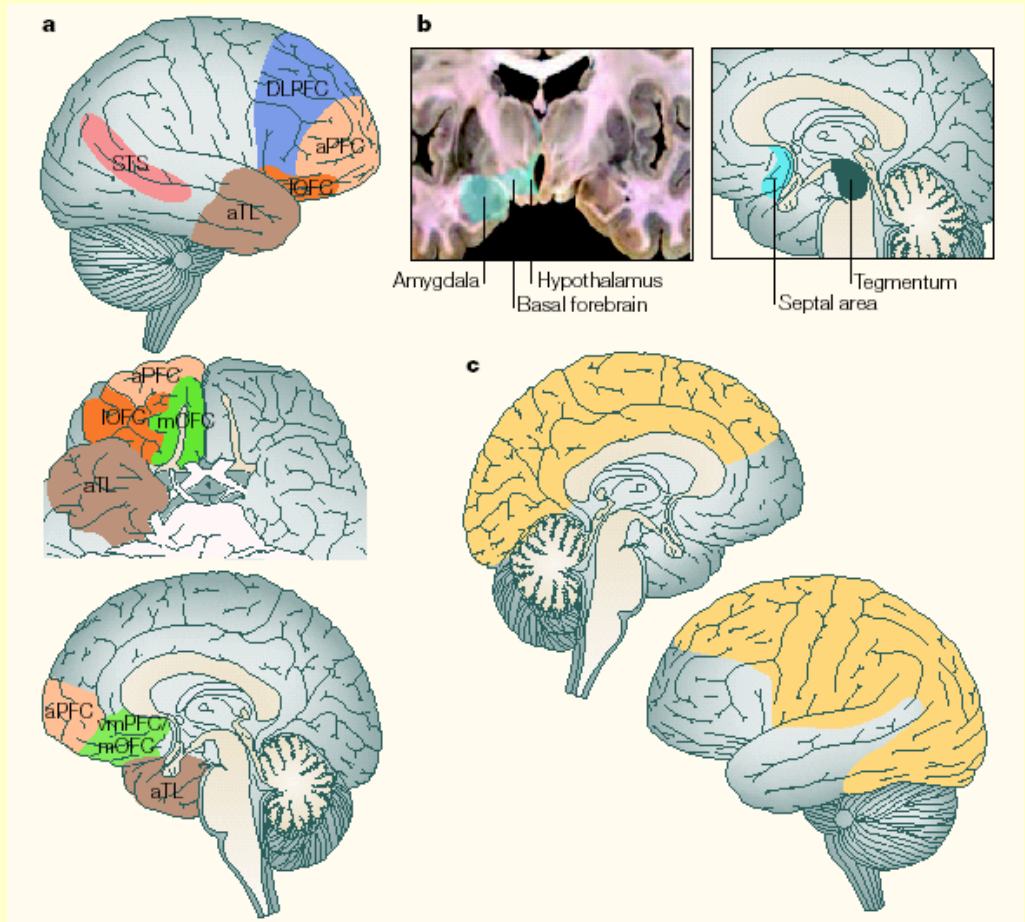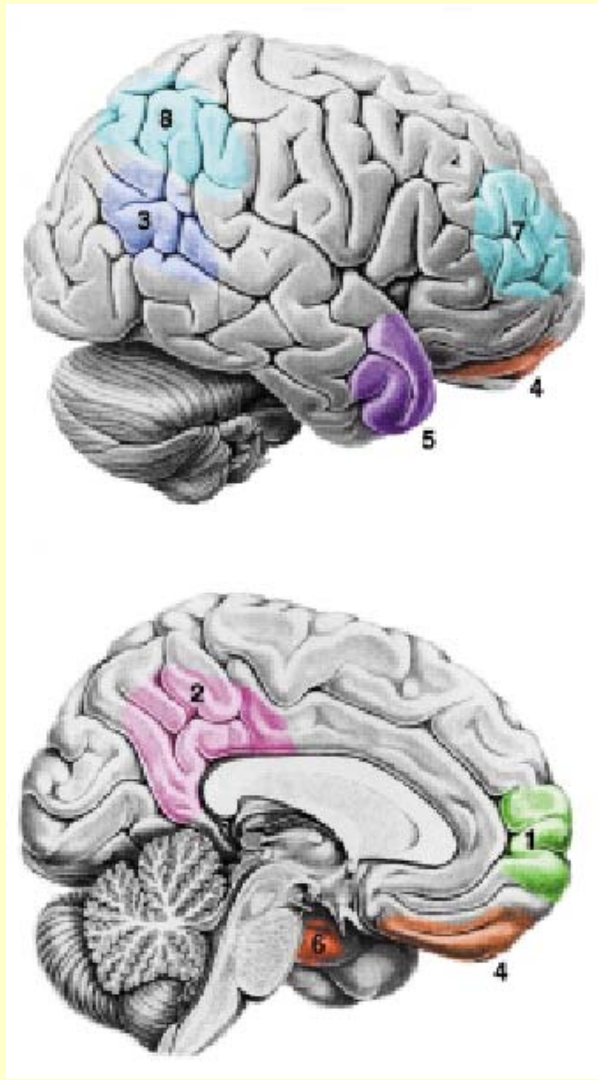
Moll et al., Nat Rev Neurosci 6, 2005.

Greene & Haidt, Trends Cog Sci 6, 2002

Methodological Fallacies

Difficulties of Imaging Experiments:

- Impact on research subjects (reproducibility)
- Measurement artifacts (e.g. movements)
- High variability (inter-trial / individual)
- Statistical problems (see picture)
- Suggestive interpretations
- Correlation of physiological and psychological entities
- Baseline condition
- There is always a maximum (nature of measurement process)
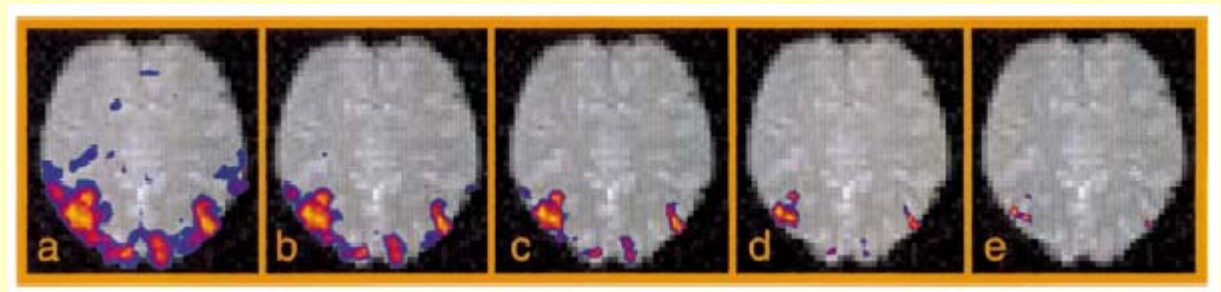
Savoy, 2001

Neuroscience of ethics on two levels
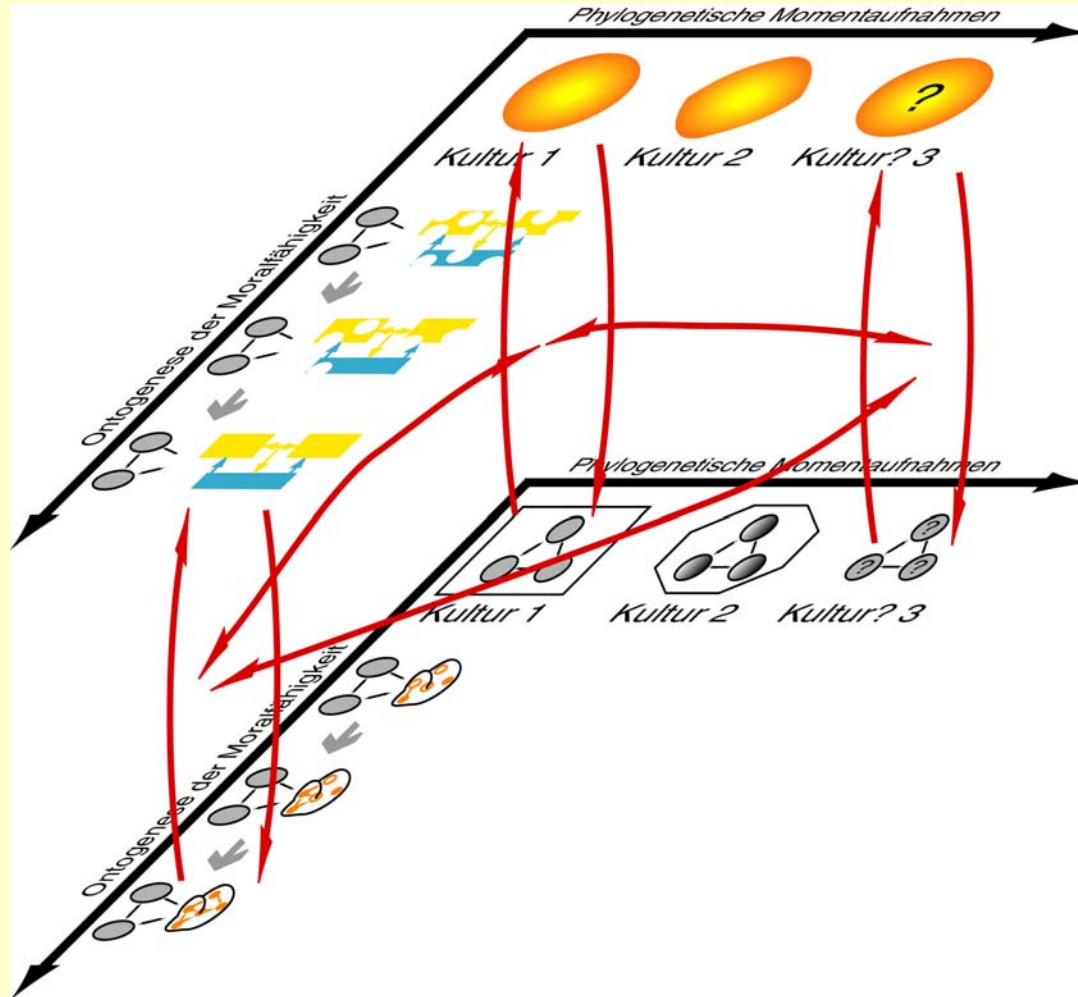
Neuroscience of ethics on two levels

Input 3 (5 Slides)

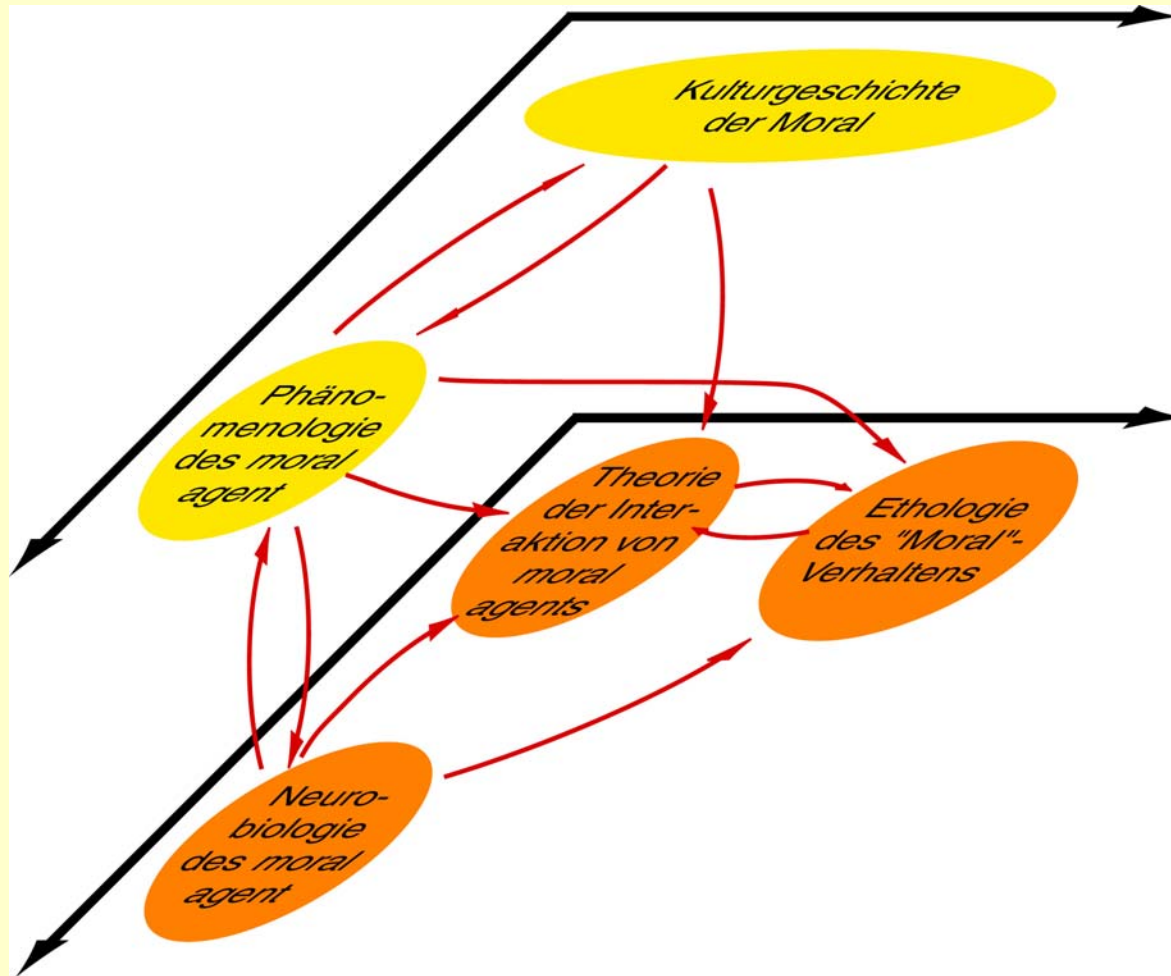**Can we find a way to quantify and compare changes in moral behavior (due to DBS and other causes)?**

(together with Sabine Müller, Aachen
Hans-Werner Bothe, Münster
Ulrich Götz, Zürich
Peter Brugger, Zürich)

Measuring Moral Behavior (1)

Classical Approaches:

1) *Kohlberg-Paradigm* (developmental psychology):
   - Moral marker: Fixed, external scale (stages)
   - Experiment: Dyadic interaction using defined dilemmas
   - Results: classification into steps based on sophisticated analysis

2) *Moral Dilemma Tests*
   - Moral marker: Dilemma-choices representing "moral types"
   - Experiment: Confrontation with dilemmas in written/oral/visual form, possibly attached to physiological measurements (scanners etc.)
   - Results: classification of agents into "genuine moral types", correlations to (biological) processes.

3) *Experimental game-theory*
   - Moral marker: Only implicitly given (trust, cooperation)
   - Experiment: Various games (Ultimatum etc.)
   - Results: classification of agents into types of behavior (e.g. trusting)

Measuring Moral Behavior (2)

Problems I have with classical approaches:

1) They often rely on an external "moral scale" not taking into account changes of moral behavior "normalized" by the persons' moral beliefs

2) Dilemmas usually refer to "drastic" scenarios (harm, killing), which do not represent "ordinary" moral problems (fairness, lying, mobbing…)

3) The kind of interactions is restricted (often only binary), the "social character" of morality (i.e. its effects on social relations) is dismissed.

4) The measurement process may involve biases that are not taken into account (Kohlberg: interviewer, games: pre-experimental instructions)

5) Multidimensional concepts (like trust) are reduced to rather simple types of decisions (leading to miscomprehensions).

6) It is usually not addressed, what representations ("Vorstellungen") specific scenarios induce in the agent.

Measuring Moral Behavior (3)

Dimensions of a "minimal moral behavior"

1) "Moral stimuli" (i.e. those aspects of a moral experiment that induce moral behavior) induce complex **representations** (Vorstellungen) within the agent, which should be

2) Moral behavior express themselves by **types and strengths of relations** with other agents.

3) Moral behavior may induce **tensions** between reputations qualified externally and by the agent.

4) The role of **reasons** must be addresses not primarily concerning their role as causes for behavior but towards their internal relations (seeing the "space of reasons" from a network-point-of-view).

5) Intuitive aspects guiding behavior are probably best addressed by referring to **moral emotions** (which are not easy to measure).

Measuring Moral Behavior (4)

"Serious Games" Approach:

Basic Idea: Today's computer games require from its players choices towards actions, traits (to develop) etc. to master a game according to its goals – i.e. may generate a large number of data in a complex but reproducible and bias-free (i.e. no dyadic interaction with changing partners) environment, that is also interesting to play (motivational aspect).

Potential measurement variables addressable via a „serious game":

- Types of „characters" the player chooses (or the players "agent" develops)
- Reasons the players choose to justify decisions
- Reaction times towards "moral stimuli"
- Interactions with other (human/artificial) players
- Short narrations (representing "Vorstellungen")

A computer game may be used to measure the "moral profile" addressing the complexity of "minimal moral behavior" more appropriately than classical approaches. ***Measurement target*** are then *changes of the moral profile* due to internal (game-parameters) and/or external (agent) causes.

Measuring Moral Behavior (5)

Quantifying DBS-associated behavioral changes:

DBS may be a paradigm to develop and test a moral game:

1) DBS-related changes in moral behavior are indeed a problem – but that is hard to address (indeed an ethical issue!)

2) Applying DBS has some on-off-effects (on a short time scale) that allow to "manipulate" the agent and to get an understanding of changes in moral behavior in a reproducible way.

Certainly, this paradigm can be used in many other cases.

Goal: Until end of march, to set up a proposal (including experts from Germany and game programmers) in order to develop the conceptual basis and a first prototype of such a game.