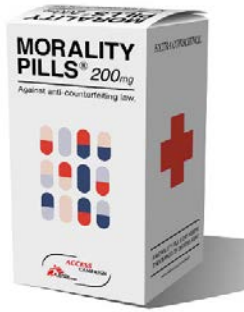
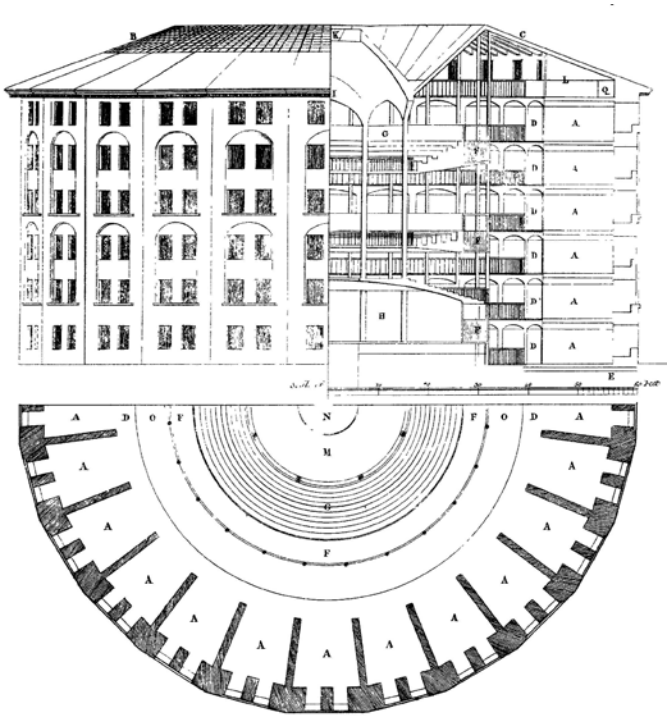
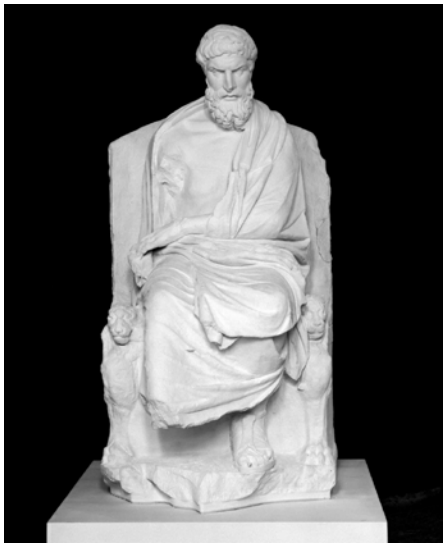
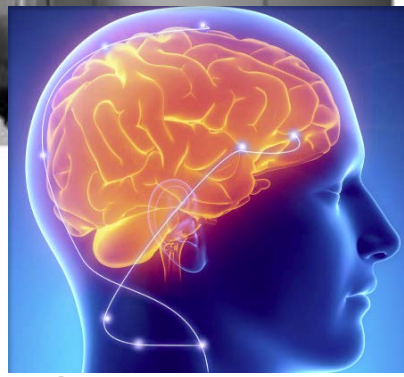
Moral Technologies









Mark Alfano, Delft University of Technology
Markus Christen, University of Zurich

Overview

- What are «moral technologies»?
- Dimensions of moral technologies
  - Social
  - Temporal
  - Disciplinary
- A deeper look into the logical structure of moral technologies
- Outlook: Conference July 11-15, 2016

What are «moral technologies»?

Our working definition:

Moral technologies *are interventions intended to improve moral decision-making in a non-explicit way – i.e. they do not target deliberation itself, but underlying neurological or psychological processes, or operate as technological mediators of human social interaction.*

Preconditions of moral technologies

Based on this working definition, applying moral technologies requires the following preconditions:

- Some consensus regarding the **goal of the intervention** as being morally favorable.
- Some understanding of the **process** in which one intervenes such that there is confidence that the effect one wants to achieve is actually achieved.
- **Tools or intervention techniques** that are feasible and do not produce side-effects that counteract the goal of the intervention.
- The ability to **measure** whether the effect actually has been achieved.

Social dimension of MT

Further specifications: What is the social scale on which MTs operate?

At one extreme of the social scale is the **individual agent**, at the opposite extreme, all of **humanity**.

In between are dyads (e.g., friends, romantic partners, allies, rivals, enemies), small groups (e.g., families, cliques), large groups containing up to the Dunbar number of ~150 (e.g., hunter-gatherer bands), large tribes, classes, races, and nations.

The social scale mainly decides both upon the feasibility and the permissibility of a MT intervention

Temporal dimension of MT

Further specifications: What is the temporal scale on which MTs operate?

At one extreme are **fast**, typically automatic and unconscious, perceptions, cognitions, inferences, decisions, and behaviors.

At the other extreme is the **evolutionary** timescale of speciation and domestication (including human self-domestication).

In between are decisions that take minutes or hours, those that take weeks or months (and are often revisited and modified along the way), and large-scale deliberative institutions, such as corporations and democracies, that endure longer than the lifespan of any particular member involved in them.

	Fast, automatic	Pause & think briefly	Weeks of deliberation	Large-scale deliberation	Evolutionary timescale
individual	Mood inductions	Buffer against ego depletion			
dyad			Optimism of partner as self-fulfilling prophecy		
Small group (4-10)	In-group bias				
Large group (~150)				Normative trust as an equivalence relation	
Tribe, class, race, nation	Police brutality				
All humans	Compassion collapse	Slow pseudo-inefficacy			Elimination of psychopaths from gene pool

Disciplinary dimension of MT

Further specifications: Based on which disciplinary background is a MT intervention framed?

- The intervention could happen on the **physiological level** of the individual agent using e.g. pharmacological means (as proposed in the "moral enhancement" debate)
- The intervention could target **psychological** constructs using means like situational cues and the like.
- The intervention could operate on the level of **interacting** with the world or other agents, in particular if this interaction is mediated through **technology** (internet search, Apps etc.).
- The intervention could act on the level of the **social design of institutions** (e.g., changing from opt-in to opt-out).

(MT def'n) **W** uses moral technology **T** grounded in scientific knowledge **S** on **X** to promote **P** in order to benefit **Y** with potential side effects **E** on **Z**.

W = agent
X = target
Y = beneficiary
Z = potential victim

T = technology itself
S = scientific grounding
P = desired state of affairs, event, value or other end
E = undesired side effects

Examples of MT

- Add omega-3 fatty acid supplements into the diet of prisoners to decrease violence.
- Give oxytocin to research participants to increase trust in economic games (you actually also get “immoral” behavior depending on the structure of the experiment)
- Change organ donation from opt-in to opt-out (although studies indicate that the effect of these changes is small)
- Set up systems of actual or apparent monitoring and surveillance to prevent cheating and encourage pro-sociality

There is an obvious relation between MTs and classic marketing research, although the goals are different.

Logical properties of MT

(MT def'n) **W** uses moral technology **T** grounded in scientific knowledge **S** on **X** to promote **P** in order to benefit **Y** with potential side effects **E** on **Z**.

W = X?

X = Y?

Y = Z?

Logical properties of MT

(MT def'n) **W** uses moral technology **T** grounded in scientific knowledge **S** on **X** to promote **P** in order to benefit **Y** with potential side effects **E** on **Z**.

- How, if at all, are X, Y, and Z informed?
- Do they have a chance to provide or revoke consent?
- Different paradigms from informed consent, e.g., authorized concealment & authorized deception?
- Democratic participation in design and implementation?
- Do they agree about the value of P?
- Do they agree about the risk of E?

Logical properties of MT

Targets of MT intervention can be first-order or higher-order:

First-order:

- compassionate behavior
- compassionate motives

Higher-order:

- competence to deliberate about how best to achieve compassionate goals
- competence to identify moral reasons

First-order moral technologies risk destroying people's integrity and authentic moral agency.

Example of a higher-order target of MT

X is morally responsible for doing/omitting to Y only if:

- X knew what she was doing/omitting Y
- X was in control over whether she did/omitted Y

Both of these conditions come with culpability caveats.

Moral technologies can either decrease or enhance responsibility by decreasing or enhancing knowledge and control.

Logical properties of MT

Normative aspects of moral technologies:

- Is deliberative engagement necessary? On the part of W, X, Y, or Z?
- How are risk and uncertainty distributed? What recourse does Z have?
- How well are feedback effects (looping) understood in advance?

Advertising our conference

	Monday Theoretical issues	Tuesday Scientific issues	Wednesday Technological issues	Thursday Ethical issues	Friday Closure
Morning **9:00 – 12:30**	**Julian Savulescu:** Moral Enhancement **Batya Friedman**: Value-sensitive design **David Abrams:** Moral Technologies and the law	**Molly Crockett:** Manipulating the moral brain **Paul Bloom:** Deliberation and moral emotions **Ann Tenbrunsel:** Ethical mechanisms in organizations	**Dirk Helbing:** From social simulations to social technologies **Paul Slovic:** Interventions for genocide prevention **Catholijn Jonker**: Persuasive technologies for influencing users.	**Nicole Vincent:** Enhancing responsibility **Marcia Baron:** The ethics of manipulation **John Sullins:** Building ethics into technological systems	**CSF-Award ceremony.** **Workshop presentations.** **Retrospective lecture by a Rapporteur.** **Outlook (30')**
	Lunch (12.30-14.00)				
Afternoon **14:00 – 17:30**	4 short (20', 5' disc.) oral presentations from younger researchers. Sum-up discussion (60') (with speaker panel) moderated by the topic responsible.	4 short (20', 5' disc.) oral presentations from younger researchers. Sum-up discussion (60') (with speaker panel) moderated by the topic responsible.	4 short (20', 5' disc.) oral presentations from younger researchers. Panel discussion (60') with invited persons from practical fields where MT may become relevant topics.	4 short (20', 5' disc.) oral presentations from younger researchers. Workshop (in groups) to identify research topics & funding possibilities.	Departure of participants.


See you on July 11-15 2016
on the Monte Verità
Ticino, Switzerland!