Robots, Emotions, and Epistemic Rational Assessability

Matilde Aliffi & Helen Ryland

Midlands4Cities
Doctoral Training Partnership
Birmingham Coventry Leicester Nottingham

PERFECT

UNIVERSITY OF BIRMINGHAM

VIDEO Sophia

Meet Sophia

Aims of the talk:

1. It is plausible that a robot could have emotions.
2. If a robot can have emotion, than some of the robot's emotions are open to rational assessment.
3. Even if robots didn't have human emotions, but robot-like emotions, then robot-like emotions would be epistemically rationally assessable.

Structure of the talk

1. Existing research on emotions
2. Robots-like emotions
3. Epistemic rational assessability
4. Can a robot meet the epistemic rational assessability criteria?

Current view on robot emotions

- Mark Coeckelbergh (2010) argues that:
- “Current robots do not meet standard necessary conditions for having emotions: they lack consciousness, mental states, and feelings. Moreover, it is not even clear how we might ever establish whether robots satisfy these conditions” (p.235).
- We will argue that a robot could meet certain relevant conditions for having emotions.

Emotions

- It is possible to distinguish between different types of affective states: **emotions**, **emotional dispositions**, **character traits** and **moods** (see, for instance, Deonna and Teroni, 2012; Price, 2015).
- **We will focus only on emotions**.
- Emotions are **intentional mental states**, that is, they are about a particular objects (Solomon, 1987; De Sousa, 1987; Goldie 2000) (Exception to this is the feeling-theory of emotion, see James, 1884).

Existing accounts of emotions

- Neojudgmentalists
- Cognitivists
- Psychological constructivists
- Psychoevolutionary
- Attitudinal theory
- Belief-desire theory
- Enactivist theory
- Constructivists
- Feeling Theories

(see Scarrantino 2018)

Plenty & Parts

- These accounts give a different answer to the "**problem of parts**", which is the problem of finding the essential components of emotions, and the the "**problem of plenty**", which consists in understanding the relationship between the different emotional components (Prinz, 2004).

Emotion components

Evaluation

Bodily changes

Phenomenology

Expressions

Attentional deployment

Motivational tendencies

Regulative components (?)

Curiosity

- Curiosity may refer to an emotional episode and to a character trait. Here we focus on curiosity as an emotional episode.
- Curiosity is **object-directed** (question/action/topic/person/event). Common sense suggests that in cases of curiosity the subject evaluates something new as positive/interesting.
- There has been some disagreement regarding this issue as curiosity is sometimes consider as a type of desire rather than as an emotional episode (see Meylan, 2014). We shall consider curiosity to be an emotional episode for the purpose of this presentation.

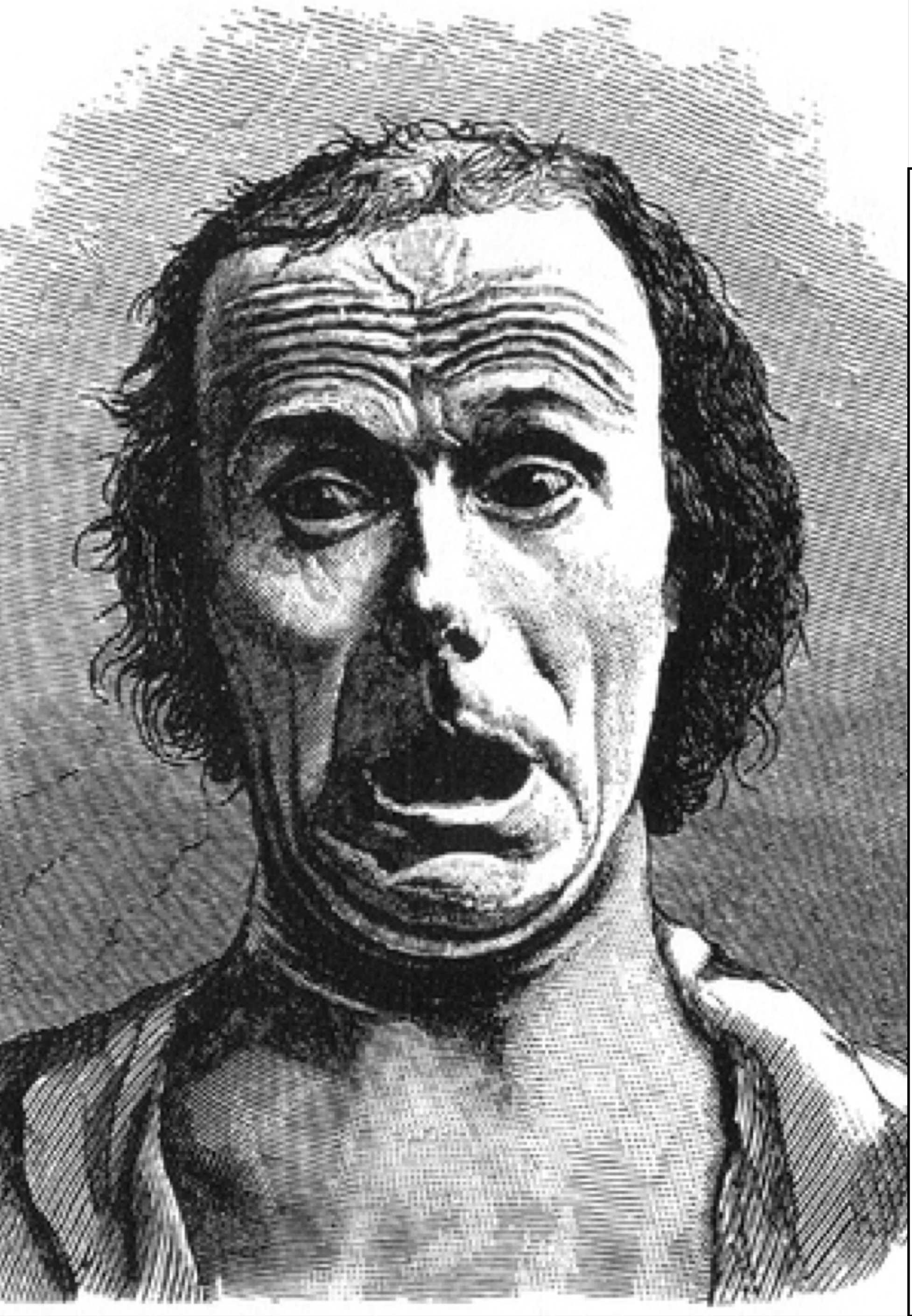
Robot-like curiosity

- There is a current HFSP (Human Frontier Science Program) project examining whether curiosity, as defined above, could be modelled and applied to machines and robotic systems.

- These projects suggest that robot-like curiosity is operationalized “as a mechanism that selects which action to experiment or which (sub) goals to pursue, based on various information-theoretic measures of their ‘***interestingness***’” (Oudeyer, 2018, p.34)

Robot-like curiosity

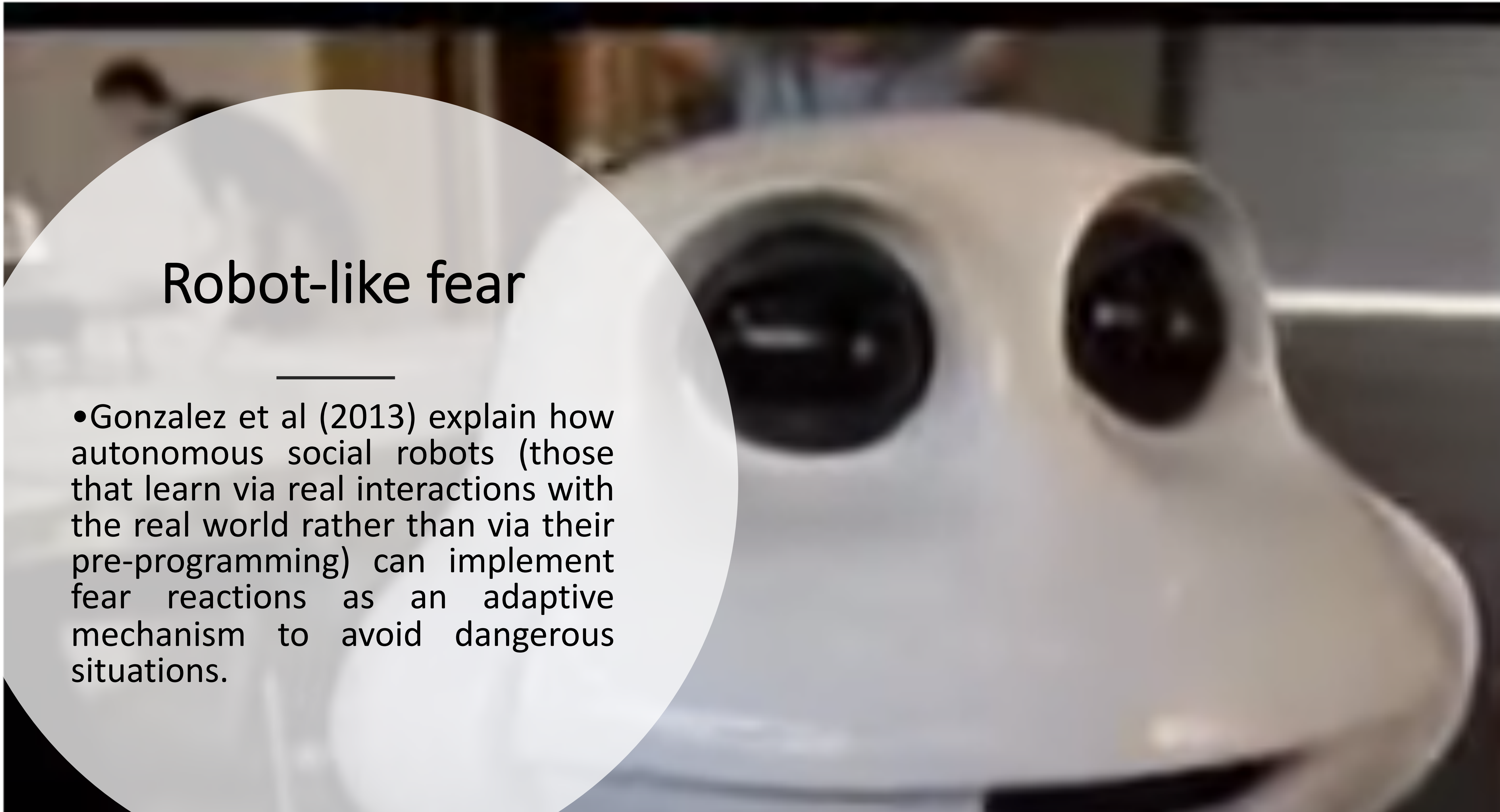
- [A demonstration by Poppy]



Fear

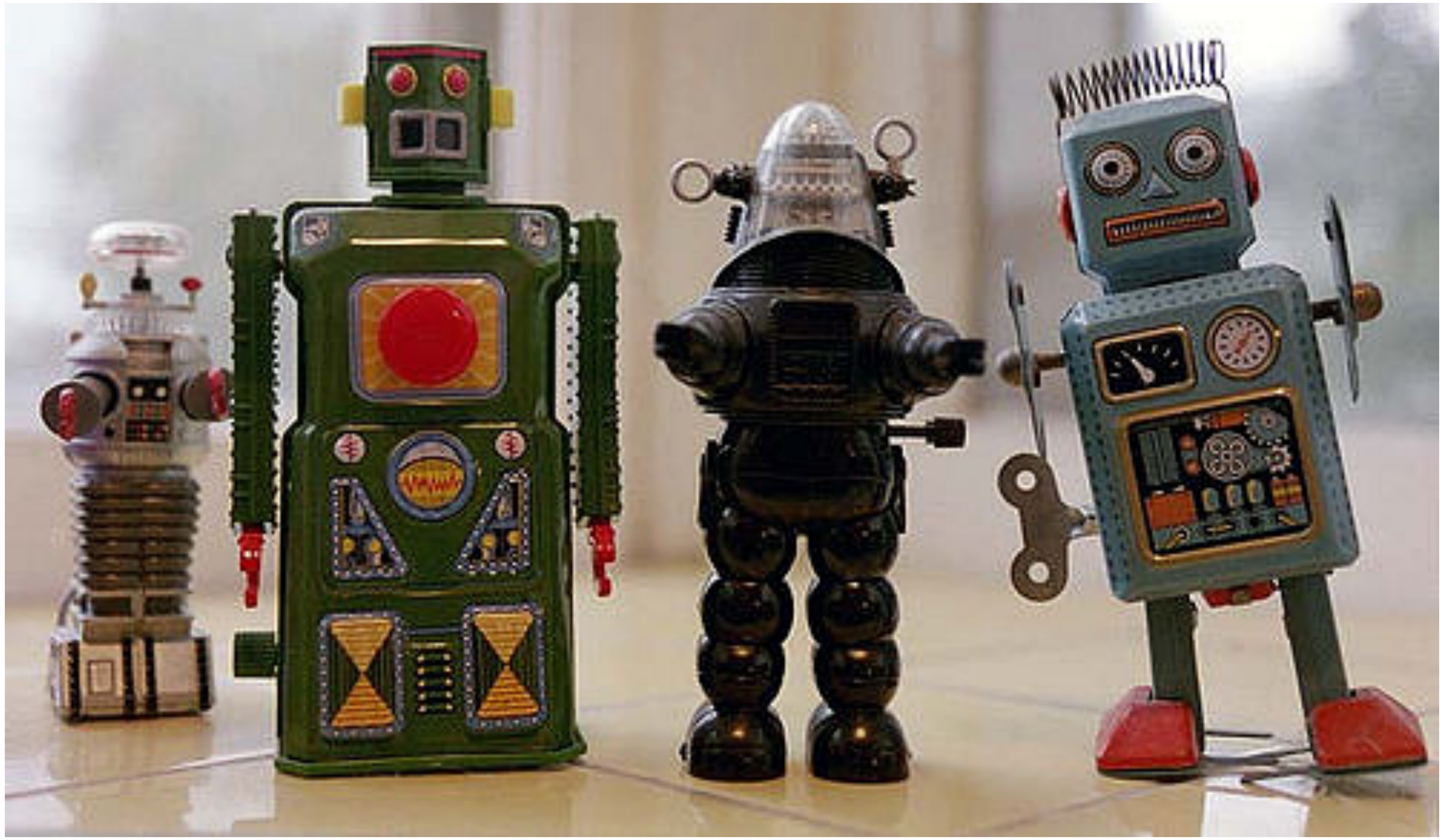
Consider a subject *S* who is scared by the object *O*.

- **Evaluation**: *O* is dangerous
- **Bodily changes**: autonomic nervous system (ANS) changes such as heart rate increase, strained breathing, piloerection, trembling
- **Phenomenology:** feeling tense
- **Expression:** mouth and eyes open
- **Attentional deployment**: attention on the source of the danger and on what the dangerous object may bring about
- **Motivational tendencies**: tendencies to get away from danger



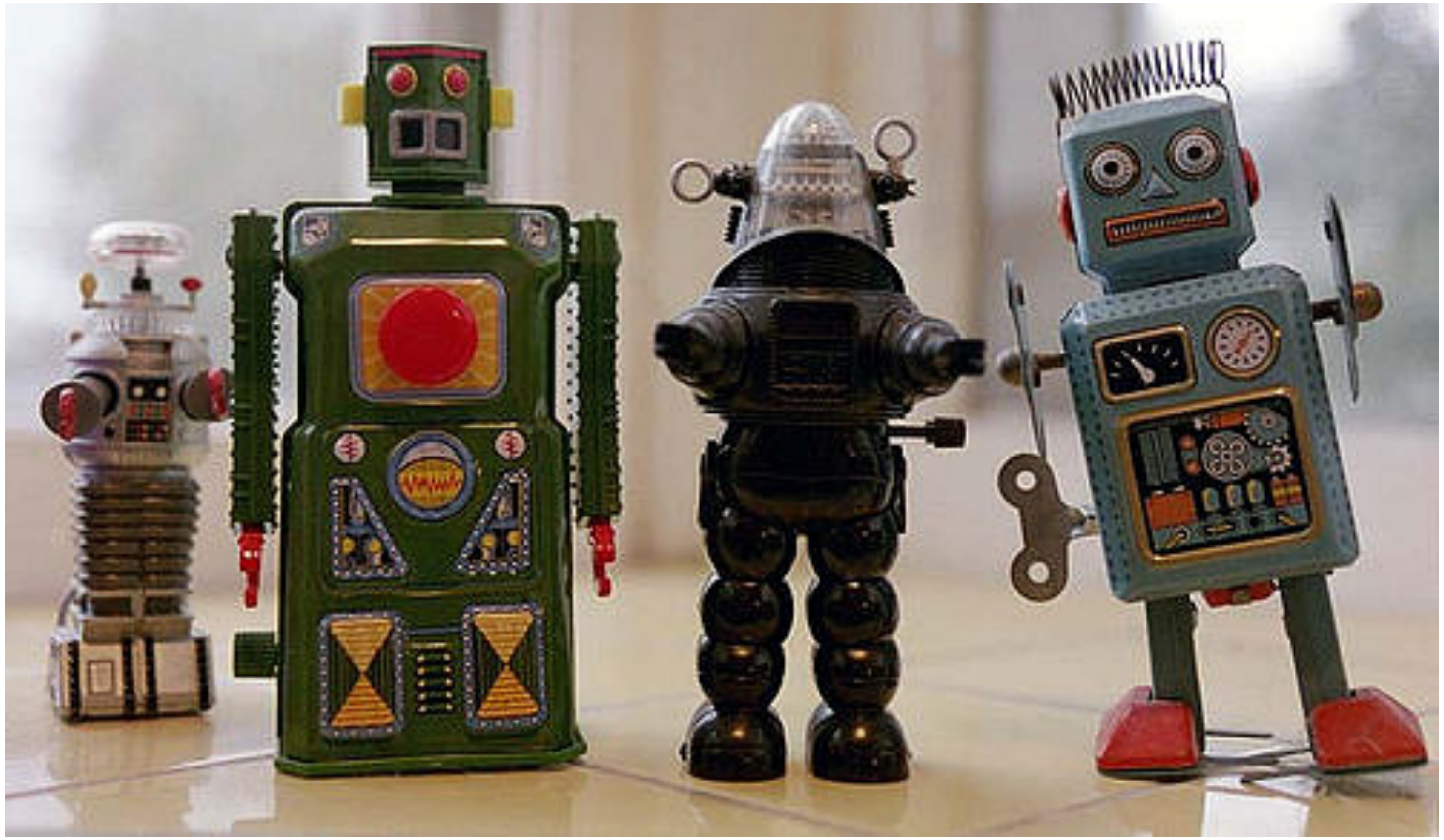
Robot-like fear

- Gonzalez et al (2013) explain how autonomous social robots (those that learn via real interactions with the real world rather than via their pre-programming) can implement fear reactions as an adaptive mechanism to avoid dangerous situations.



Why is rational assessability important for robots?

- One reason why it might be important to look at the rational assessability of robot emotions is to try and understand the relationship between robots and humans.
- One area where our discussion could be particularly useful is in discussions of moral personhood and (human) rights.



Why is rational assessability important for robots?

- On a very general level, most (human) rights accounts argue that entities are persons and rights-holders if they meet certain criteria (such as rationality, consciousness, having felt experiences etc.)
- The discussion that follows will suggest that robots can meet at least some of these criteria. On at least the human rights account that I am developing, this would entail that robots could be persons and rights-holders (at least to some degree).

Why is rational assessability important for robots?

We can **apply** epistemically rationality norms to an emotion only if the emotion is epistemically rationally assessable.

Epistemic rationality norms as **regulative ideal** for robot-emotional learning.

Design robots that are better at understanding the evaluative import of the environment/context.

Epistemic rational assessability

- Can epistemic rationality norms apply to some instances of robots-emotions? If we answer affirmatively, then robots-like emotions are epistemically rationally assessable.
- Emotions experienced by robots have features that are typical of epistemic rationally assessable mental states.

Criteria for the epistemic rational assessability

An emotion e is epistemically rationally assessable if and only if:

- *e* is informationally plastic
- *e* has a descriptive content
- *e* contributes to the subject's view of the world
- The mechanism *P* that maintains *e* is generally receptive to the evidence.

Informational plasticity

Stating that a mental state is informationally plastic means stating certain facts about the mental state's etiology.

A mental state is informationally plastic when (i) it can be acquired by several different means, and (ii) it drives different forms of behavioural expression.

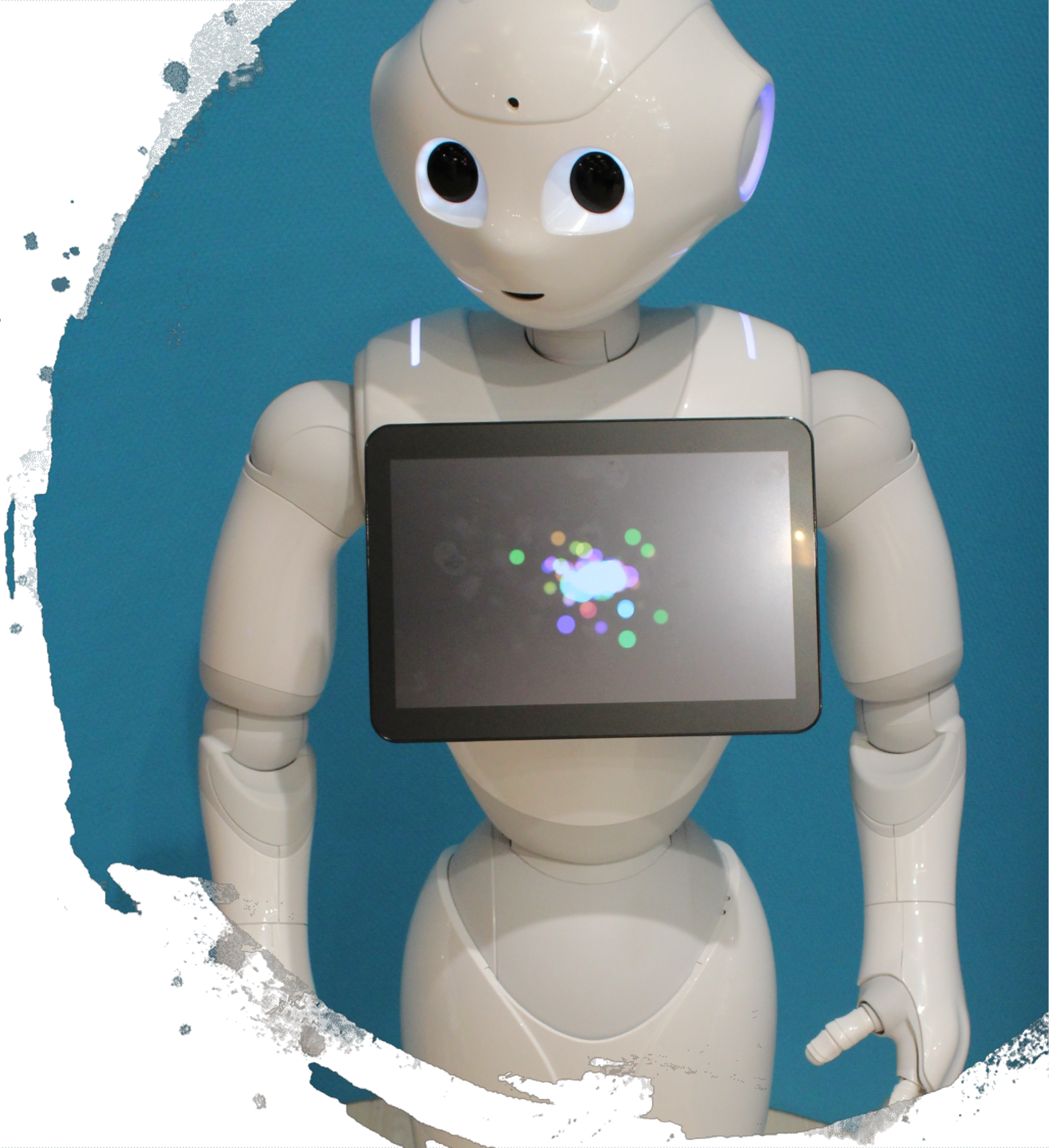
Informational Plasticity in robots

- To show that a robot could meet the informational plasticity condition, we first need to show that it can have a mental state (or robot equivalent) that is not the output of a rigid process (its programming).
- Machine learning can help here. There are many definitions of machine learning, but for this talk we'll use the one adopted by the Royal Society (2016): "Machine learning is a form of artificial intelligence that allows computer systems to learn from examples, data, and experience. Through enabling computers to perform specific tasks intelligently, machine learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules".

Informational Plasticity in robots

1) Via machine learning, different sources of information can activate the same mental state/output in the robot.

- An example of this can be seen in ‘Pepper the emotional robot’. Pepper’s emotional reactions are supposedly influenced by touch sensors and cameras (so what they see, feel, hear, etc.). It is reasonable to suppose that the same emotion in Pepper can be activated through different sources of information. For example, Pepper ‘feels happy that S is happy’ because she sees S smiling or hears S speaking happily.



Informational Plasticity in robots

2) The robot's mental state/ output could have been different if different information was integrated.

- Matthias (2004) explains how there are autonomous artificial agents who learn by interacting with humans in real life and gradually learn to make their own behavioural choices. In this scenario, the information that the robot integrates is dependent upon context and the interactions that it has. If the robot was in a different context and had different interactions, they would learn different behavioural outputs.

Cozmo and Informational plasticity

- An example of this can be seen in Cozmo, the AI robot with emotions. Cozmo develops different ‘emotional’ reactions on the basis of different external stimuli: https://www.youtube.com/watch?v=DHY5kpGTsDE

Descriptive content

- Does robot-like emotions have a descriptive representational content?

- A mental state has a descriptive representational content when it **describes** facts about the world. (It gives us information about how the world is like).

Descriptive content in robots

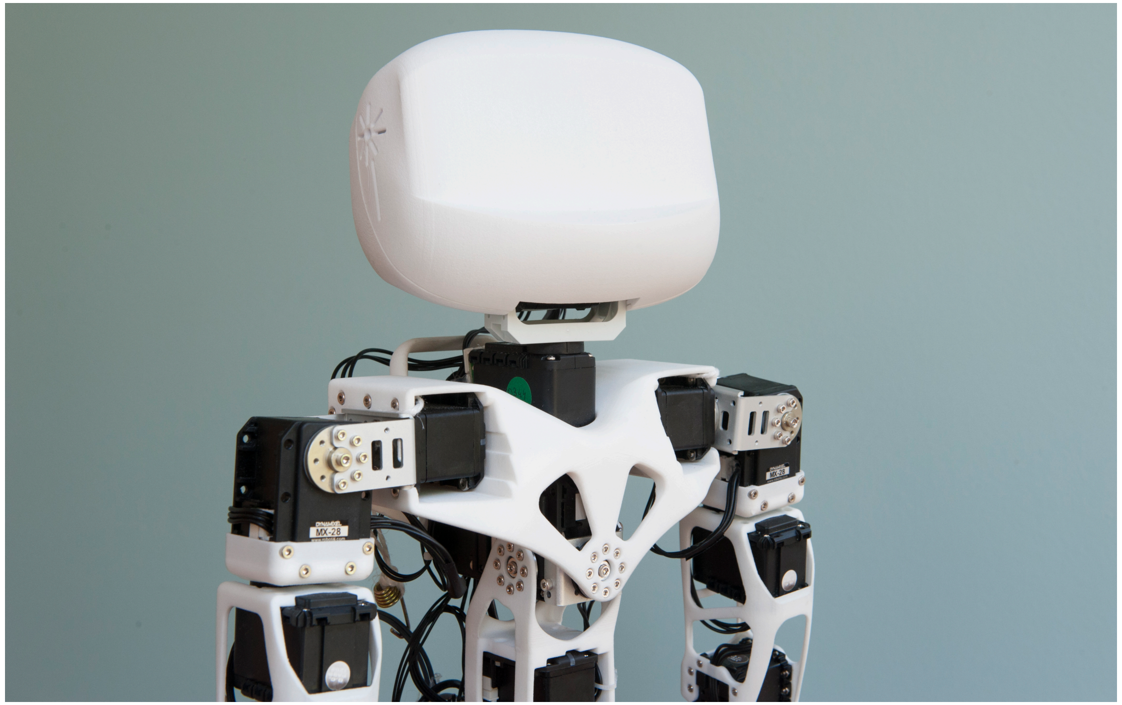
- To show that a robot meets the descriptive representational content condition, we need to show that the robot's emotions perform a descriptive function: they inform the robot about how the situation impinges their well-being.
- We can see this in the Gonzalez paper on robot-like fear. The robot uses the 'fear' response as an adaptive mechanism to avoid dangerous situations.

Contribution to the subject's view of the world

- The subject view of the world is comprised of a set of mental states that the subject takes at face value.
- A mental state of a subject S contributes to the S's view of the world when the content of the mental state tell us something about the world, and this is information is taken at face value.
- Do robot-emotions contribute to the robot's view of the world?

Contribution to the robot’s view of the world

- To show that a robot’s emotions meet the above condition, we must show that the robot’s emotions are taken at face value by the robot and contribute to their view of the world.
- This can be shown in both our robot-like curiosity and robot-like fear examples. The robot takes the emotion (curiosity and fear) at face value and acts on the basis of it (exploring their surroundings, altering behaviour to avoid fear-inducing stimuli).



Receptivity to the evidence

- A mental state is receptive to the evidence when the mechanism that maintains the emotion is able to compare incoming evidence and detecting any conflict with it.
- Receptivity to the evidence does not imply responsiveness to the evidence (it does not imply that the subject revises his mental state accordingly).

Receptivity to evidence in robots

- To show that a robot's emotions can be receptive to evidence we need to show that the emotion can change when the robot gets new evidence. We can see this in robot-like curiosity. As mentioned earlier, robot- like curiosity relies on the robot finding something interesting. When the robot is no longer interested in x, they will stop displaying curiosity of x. We can see this in the playground experiment where the robot gradually gets interested in different things, and changes what it is curious about as a result:
- https://www.youtube.com/watch?time_continue=1&v=uAoNzHjzzys

Conclusions

1. **It is plausible that a robot could have emotions.**
    1. We have shown that it is plausible that robots could at least have curiosity and fear.
2. **If a robot can have emotion, than some of the robot's emotions are open to rational assessment. Even if robots didn't have human emotions, but robot-like emotions, then robot-like emotions would be epistemically rationally assessable**

    2. In our discussion we have not shown that a single robot's emotion meets the condition for epistemic rational assessability. However, we have shown that a variety of robots have some emotions or robot-like emotions that meet some of our conditions for epistemic rational assessability. This suggest that if a robot was developed with all of the features then its emotions would meet the conditions for epistemic rational assessability.

Suggestions for future research

- How to develop a robot that has all the features, like informational plasticity, descriptive content, contribution to the robot-like view of the world and responsiveness to the evidence?
- If a robot like this could be developed, then how should we treat this robot?
- How can we design a robot that learn in more effective manner using epistemic rationality norms as regulative ideal?

Thank you for your attention!