

# Could Social Robots Make Us Kinder or Crueller to Humans and Animals?

Coghlan, Vetere, Waycott, & Neves (2019)

Jean-Noël George  
11<sup>th</sup> of March 2020  
So-Bots Journal Club

# The scenario

2016 Montreal Declaration for Responsible Development of Artificial Intelligence

*'Shouldn't encourage cruel behaviour to robots that look like humans or animals'*

# 3 ethical questions

Moral standing/patients - Can they be wronged/owed?

If not the above, could certain responses remain intrinsically unethical?

Do bad/good responses to social robots transfer to sentient beings?

## 3 ethical questions

Moral standing/patients - Can they be wronged/owed?

**Can not and will not have intrinsic moral standing\***

If not the above, could certain responses remain intrinsically unethical?

**“Torturing” could be intrinsically unethical/unvirtuous**

**Do bad/good responses to social robots transfer to sentient beings?**

# Social Robot-Human **Virtue Link**

*Possibility that responses to social robots may, positively or negatively, influence human responses toward sentient and living beings*

## **Disagreements**

Humanoid and animal robots may precipitate bad behaviour to sentient animals and people

**Darling (2006) – Echoes Kant's hardness/tenderness to animals engenders hardness/tenderness to humans = protect with law**

Robots needs to be indistinguishable from humans for this to be an issue

**Johnson & Verdicchio (2018) – Robots only simulate sentience, disconnecting the 'link' = no need for laws**

# Social Robot-Human **Virtue Link**

*Possibility that responses to social robots may, positively or negatively, influence human responses toward sentient and living beings*

**Social robots may influence the development of moral virtue**

**A link is more plausible terms of effects on children, ethical attitudes, and behaviour toward non-human animals**

**Possibility that social robots influence adult moral responses to humans and animals**

# Social responses to social robots



R.I.P Steve the Security Guard

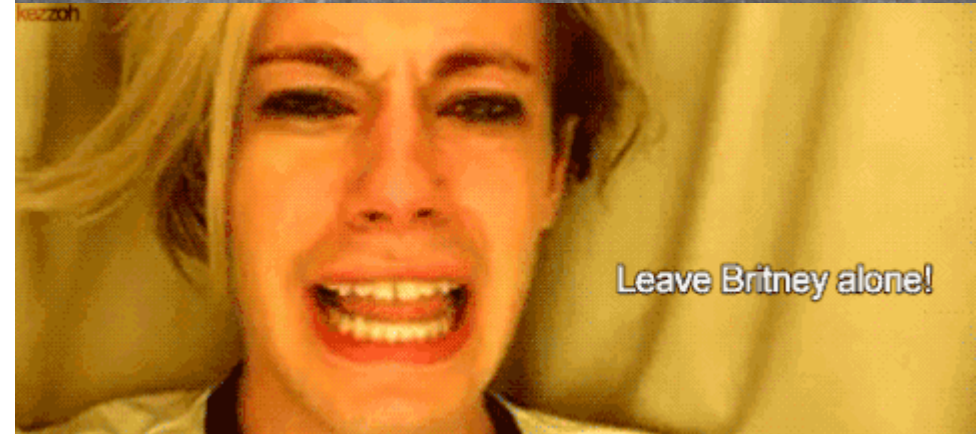


Pity & Anger  
"Stop the abuse!"

# Social responses to social robots



R.I.P Steve the Security Guard





# Some **empirical** evidence

- Humans empathise with actual robots, compared to simulations  
Seo et al. (2015)
- Reluctant to shock robot subjects compared to virtual robots  
Bartneck & Hu (2008)
- Children verbally abuse, kick, and punch robots, like they do animals  
Nomura et al. (2016)

**Instinctive? Similar to how we respond to humans and animals**

**Focus on/increasing levels of anthropomorphism in robot design**

# Where do we go from here?

## **Problem;**

Although the existing research is interesting, there is insufficient empirical data to discredit, or support, the social robot-human virtue link.

## **Conceptual reflection**

### **Darling's Animal Analogy;**

- Kant's casual claim of cruelty
- Prone to anthropomorphism & project one's own qualities
- This enables implicit empathy
- Safeguard with laws

### **Johnson & Verdicchio's Argument;**

- Influence only in some ways
- Lack of carry over effects
- Cruelty depends on suffering, and no real suffering is caused

# Arguments **against** a virtue link

## Inconclusive Empirical Support from Media Entertainment

**Not so simple cause-and-effect link** between video game violence and antisocial behaviour or attitudes.

- We can, and do, distinguish between fiction and reality

No significant correlation in British children

Ferguson & Colwell (2018)

About the **context** of play, **not quantity** of violence

Shibuya et al. (2008)

# Arguments **against** a virtue link

## Problems with Kant's Causal Claim

**“Dirty play” is contained to phases in childhood**

Animal cruelty to human cruelty is grounded in personality types  
Arkule (2002)

**Kant:** Only moral reason to be kind to animals is that it influences humanity toward ‘rational’ beings

If robots are regarded as **objects**, how does that translate to beings believed to have inherent worth?

# Arguments **against** a virtue link

## Problems with Kant's Causal Claim

### **Compartmentalisation;**

Quarantining certain attributes or behaviours from core self-understanding

### **Slaughter-house workers\***

Behaviour toward humans and other animals?

Johnson & Verdicchio (2018)

### **Not necessarily a logical system**

# Arguments **against** a virtue link

## **Problems with the use of Anthropomorphism and Zoomorphism**

**‘Tendency to over-ascribe autonomy and intelligence, subconsciously’**

We involuntarily misapprehend true nature of robots

A spontaneous response in order to interact with robots

No illusion of suffering for characters in fictional stories

**Doesn't embodiment count for something?**

Initially a strong response, but consolidated with ‘it's just robots’ and fictionalised\*

# Arguments **for** a virtue link

## Animal-robot analogies don't fail

Aversion to hit personified Hex bug with mallet  
Darling, Nandy, Creazeal (2015)

Pity for Pleo the dinosaur when being 'tortured'  
Rosenthal-von der Pütten et al. (2013)

Non-sentient animals are able to struggle & flee - evoke pity and care

If robots were **highly autonomous**, able to struggle, flee, protect us, show loyalty/preference etc ?

# Arguments **for** a virtue link

**Intelligibility not to be cruel, despite non-sentience**

Instinctive elicitation of pity for bugs – why not robots?

## **The Good Place**

Reluctance to switch off robot  
that audibly objects to being switched off

Horstmann et al. (2018)





# Concerning **children**

## **Shaping of children's moral characters**

**Children are trained from childhood to be virtuous**

More susceptible to social influence from a group of robots

Volmer et al. 2018

More exploration, apprehension, affection, reciprocity with AIBO

Kahn et al. (2006)

Plausible that children will mistakenly believe robots are sentient

**We want children to have an empathetic response**

# Concerning **nonhuman** animals

**Relatively weak & inconsistent societal influence on virtuousness to nonhuman animals**

Some animals are loved, whilst others met with indifference

Singer (1995)

Regan (2004)

Weak societal influences of virtuousness to robots may blur the line

Laws may help foster social pressures, encouraging gentility

# Strengthening a virtue link?

When robots are recognised as having moral patienthood, the link strengthens

Certain abuse to life-like robots may promote that abuse

Such acts may be an expression of the abuse of a human being

Thoughts and Feelings?