

***THE AI-STANCE:***  
CROSSING THE TERRA  
INCOGNITA OF HUMAN-  
MACHINE  
INTERACTIONS?

*ANNA STRASSER (LMU, Munich, Germany) &*

*MICHAEL WILBY (Anglia Ruskin University, Cambridge, UK)*



# OVERVIEW

**I**

## ***SOCIAL ARTIFICIAL AGENTS?***

- some social human-machine interactions (HMIs) should not be understood as mere ‘tool-use’

**II**

## ***TWO THEORIES OF JOINT ACTION REJECTED***

- rejecting standard construals of intentional agency because they presuppose too demanding conditions on social agency

**III**

## ***TOWARDS A GRADUAL APPROACH***

- expanding our conceptual framework by utilizing minimal approaches (minimal joint actions)

**IV**

## ***THE AI-STANCE***

- utilizing Dennett’s stance epistemology & introducing a new stance – the AI-stance

# I. SOCIAL ARTIFICIAL AGENTS?

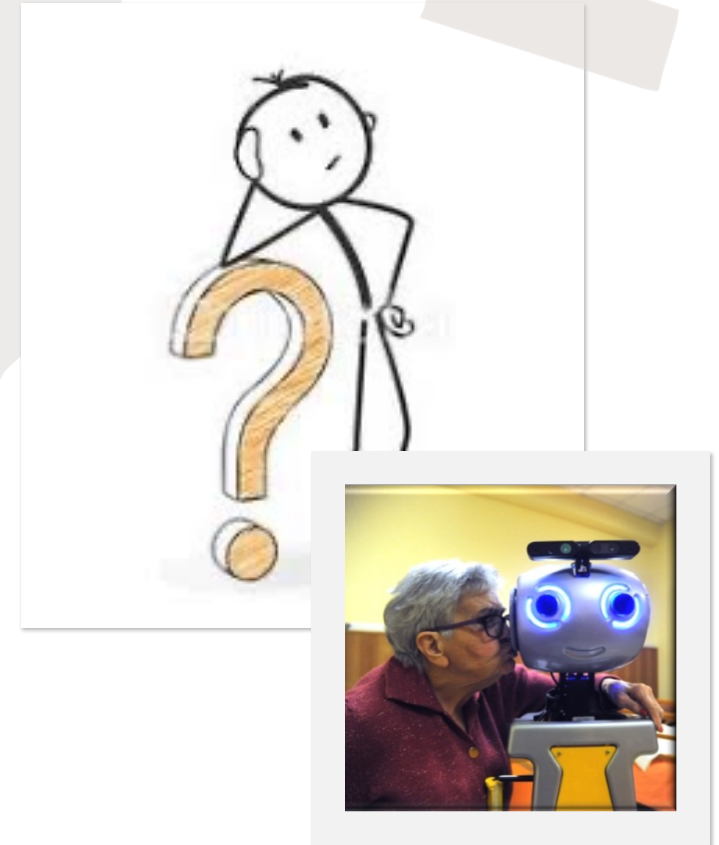
SOON WE WILL SHARE A LARGE PART OF OUR SOCIAL LIVES WITH  
VARIOUS NEW KINDS OF INTERACTIVE ARTIFICIAL SYSTEMS

## *Potential Consequences*

- responding to artificial agents as we do in social interactions with humans
- some human-machine interactions cannot satisfyingly be reduced to mere tool-use
  - e.g., learning algorithms, social robots, degree of autonomy, learning from experience, adapting their goals correspondingly, reacting to social cues

*IF we consider certain artificial systems as social agents instead of mere tools → new type of social interaction*

- *Different in kind* to those which we might engage in with other adult humans, children, or non-human animals
- *NO full-blown mentality* attributed to artificial systems



# IN-BETWEEN TOOL-USE & SOCIAL INTERACTIONS

## *MACHINES*

- mere behavior
- all interactions with machines are described as *tool-use*

## *HUMANS*

- intentional actions
- many human-human interactions are described as *social interactions*

some artificial systems do more than just behave  
→ should not be categorized as mere tool-use

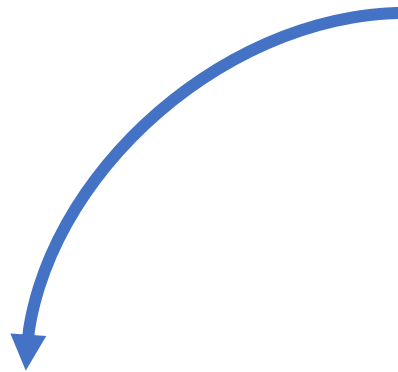
BUT artificial systems do not fulfill the demanding conditions of prototypical social interactions

→ *terra incognita for which we have no established notions yet*



# NOTIONS FOR IN-BETWEEN CASES

status quo: no notions for in-between cases



expand concept of tool-use  
*(add complex tools with social features)*

expand conception of social interactions  
*(add non-living social agents)*

# SOCIAL ARTIFICIAL AGENTS

?

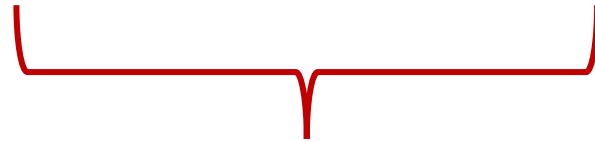
JOINT ACTIONS



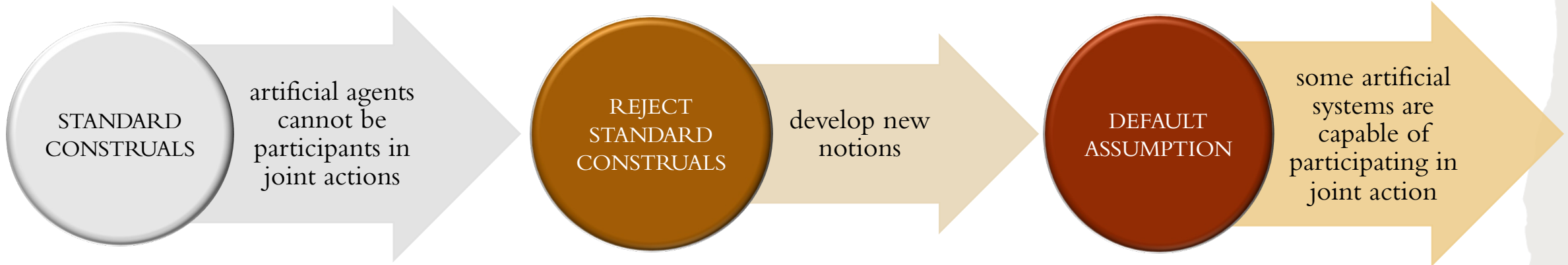
# II. TWO THEORIES OF JOINT ACTION REJECTED

*intellectualist conception of intentionality*

*biological conception of intentionality*



artificial systems are in principle not capable of engaging in joint action



# INTELLECTUALIST CONCEPTIONS OF INTENTIONAL ACTION

## NECESSITY OF A COMPLEX SUITE OF CONCEPTUAL RESOURCES

Donald Davidson (1963, 1971, 1980, 1982, 1984, 2001) :

- constitutive relations holding between propositional attitudes and their contents, as well as language, intentional action and interpretation, sharply separate off ‘the beasts’ from rational animals such as humans



“The intrinsically holistic character of the propositional attitudes makes the distinction between having any and having none dramatic!”



# SUMMARY OF DAVIDSONIAN ARGUMENT

## *Summary of the Davidsonian Argument:*

- Full-blown intentional agency requires intentional action to be carried out by an entity with an *integrated, holistic set of propositional attitudes*.
- Should we accept the conclusion and arguments?

# OBJECTIONS

## *Empirical- based*

DEVELOPMENTAL & COMPARATIVE PSYCHOLOGY:

*counterexamples*

- **Multiple realization** of socio-cognitive abilities in infants & non-human animals → acting jointly

Perler & Wild 2005, Premack & Woodruff 1978, Heyes 2014/2015, Vesper et al. 2010, Warneken et al. 2006

→ SO: not only conceptually sophisticated humans can act jointly

## *Conceptual- based*

ONTOGENETICS & PHYLOGENETICS

*counterexamples*

- Shift from non-intentional to intentional is **gradual & partly learnable**
- Ontogenetic case: Perner 1991, Tomasello 2008
- Phylogenetic case: Sterelny 2014, Henrich 2016

→ SO: Davidsonian ‘all-or-nothing’ dramatic divide is implausible

MAYBE: artificial systems can also be social interaction partners?

# *BIOLOGICAL CONCEPTIONS OF INTENTIONAL AGENCY*

## CLAIMS

Any kind of agency that enables entities to be a participant of a joint action requires internal affective states (emotional, mental and conscious states).

A biological make-up is necessary to have genuine intentional and conscious thoughts.



ARTIFICIAL SYSTEMS CANNOT QUALIFY AS SOCIAL INTERACTION PARTNERS

BECAUSE THEY LACK THE BIOLOGICAL MAKE-UP THEY CAN ONLY BEHAVE – NOT ACT

→ ***EVERY HUMAN-MACHINE INTERACTION SHOULD BE UNDERSTOOD AS MERE TOOL-USE***

# OBJECTION

*Why should we disqualify machines because they are not living, biological beings?*

*What about assuming, that the way living beings fulfill the conditions for agency is just one way to realize agency?*

**MULTIPLE REALIZATIONS OF AGENCY**  
→  
**EXTEND THE CONCEPTION OF AGENCY IN  
VARIOUS INTERESTING WAYS**