

Moral Machines in the Military Sphere

Paul F. Bello, Ph.D.

Director, Computational Cognitive Science and AI

Office of Naval Research

Human and Bioengineered Systems, Code 341

Arlington, VA 22203

paul.bello@navy.mil

Take-Away from Yesterday's Session on Military Robotics

- DoD 3000.09 and Integrated Roadmap for Unmanned Systems
 - No foreseeable plan for removing humans from the loop with respect to the employment of lethal force.
 - Development of rigorous test and evaluation methodologies for advanced autonomy.
- Targeting is not the end of the discussion, nor should it be the focus.
 - Other military activities of interest: Disaster relief, first response, search-and-rescue, surveillance/reconnaissance, and medical robotics.
- Is dystopia imminent?
 - We must move past discussions about whether we ought to build machines capable of navigating moral terrain (since its already happening), and begin a discussion about how to build them to minimize any negative impact they might have.
 - Per Prof. Whitby: dumb autonomy = dangerous autonomy
 - Yes, in some cases – especially ones involving close interaction with humans.
 - This is what I see my job to be – making autonomy less dumb, and interactions with humans more fluent.

Program Summary

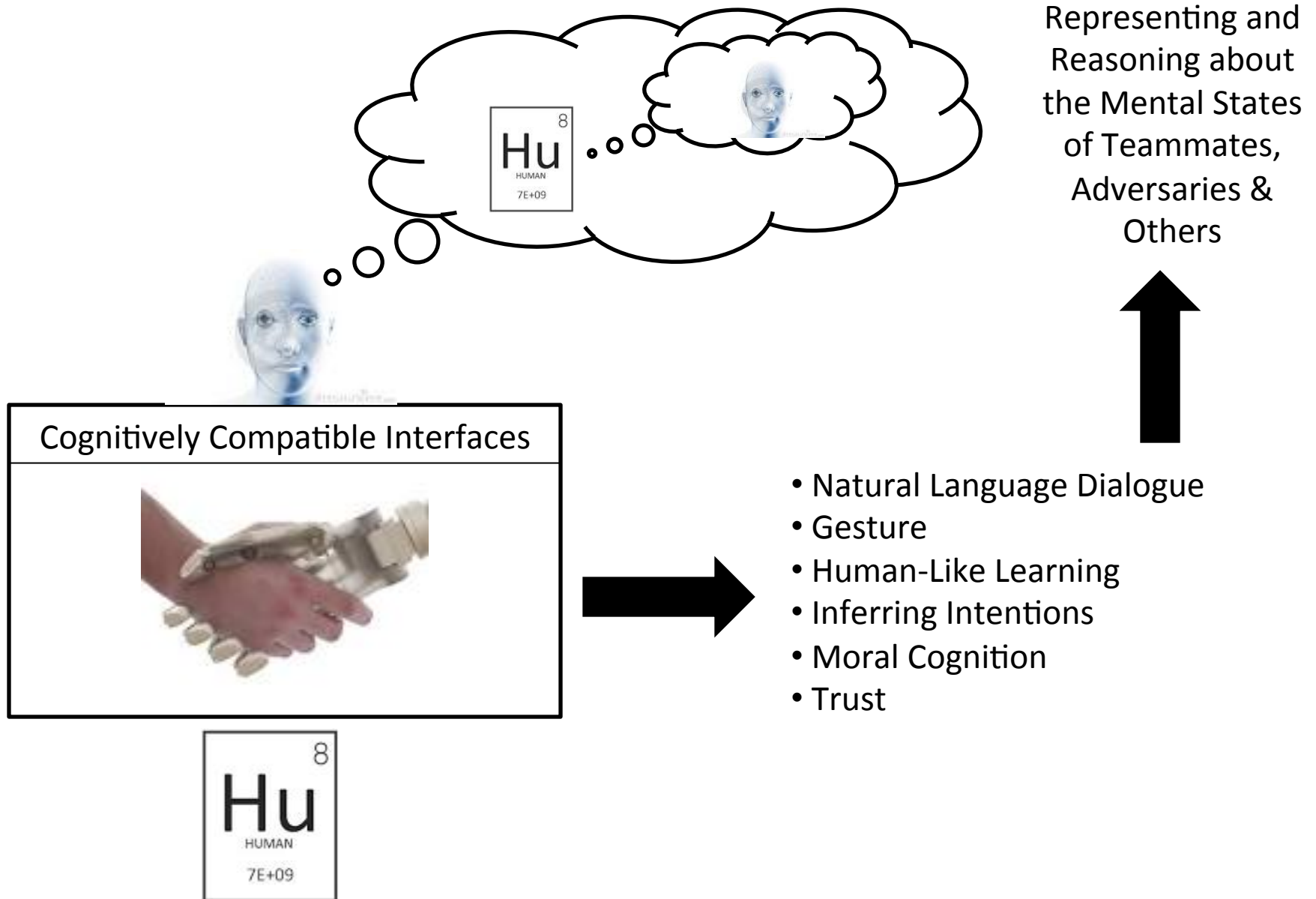
Long-Term Scientific Objectives

Development of embodied cognitive architectures that integrates inference over perceptual data and cognitive structures (e.g. memories etc.) with special emphasis on social cognition:

- Language understanding and dialogue
- Mental-state reasoning, esp. intention recognition and belief ascription
- Moral judgment
- Applications: Synthetic teammates, human-centered decision-support, human-robot interaction



Special Focus: Mental-State Attribution



Robots Tracking Mental States

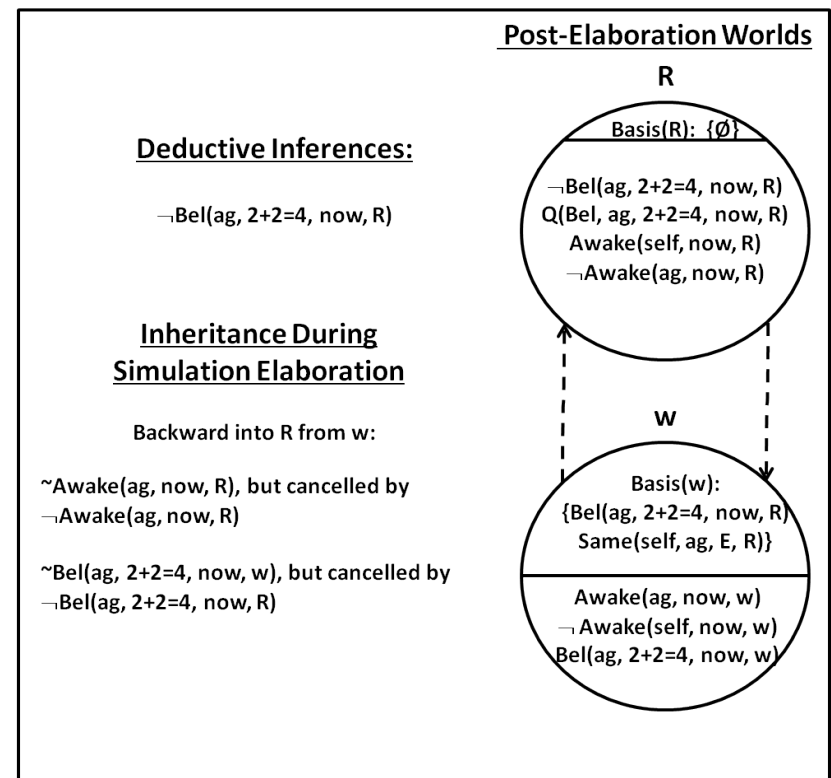
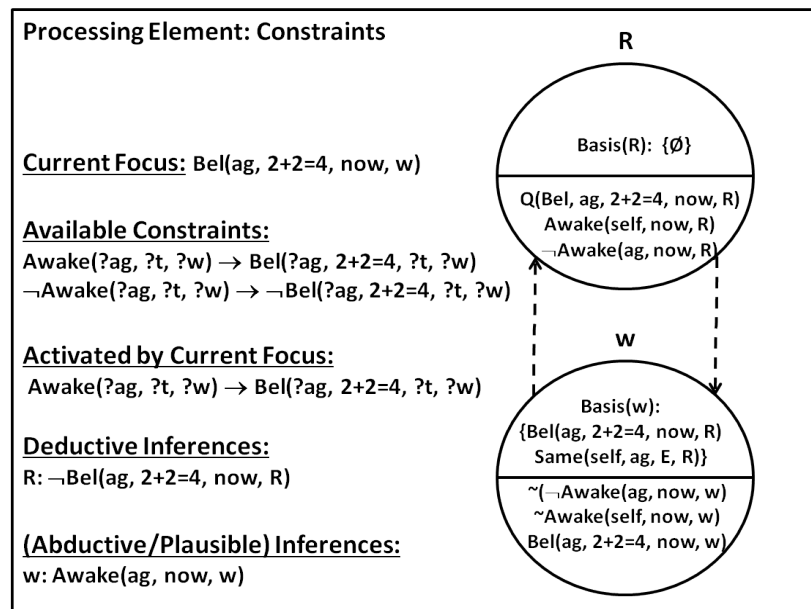
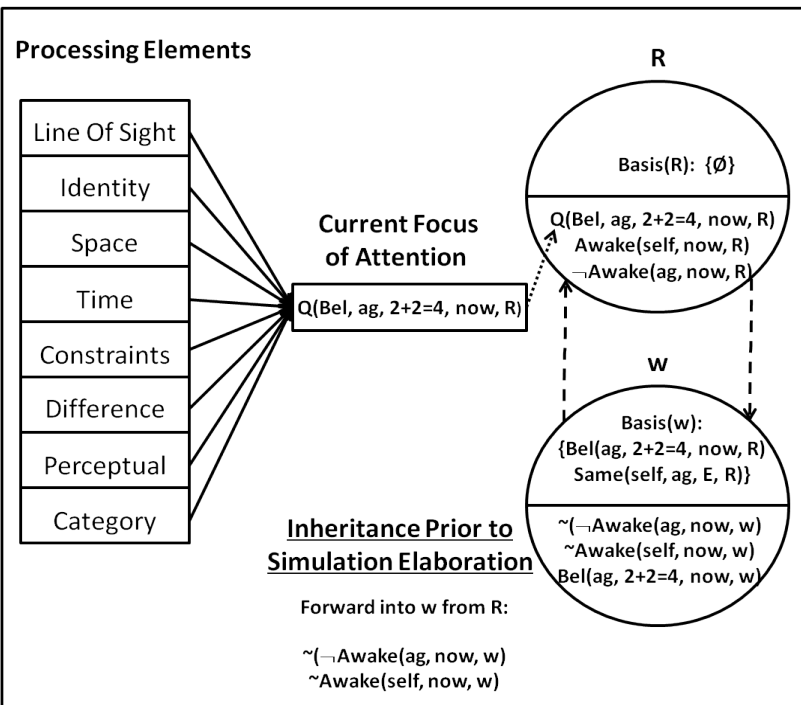
Practicing what one Preaches

“Many Caucasians in academia profess that all races are of equal intelligence. Juliet, let’s suppose, is one such person, a Caucasian-American philosophy professor. She has, perhaps, studied the matter more than most: She has critically examined the literature on racial differences in intelligence, and she finds the case for racial equality compelling. She is prepared to argue coherently, sincerely, and vehemently for equality of intelligence and has argued the point repeatedly in the past. [...] And yet Juliet is systematically racist in most of her spontaneous reactions, her unguarded behavior, and her judgments about particular cases. When she gazes out on class the first day of each term, she can’t help but think that some students look brighter than others – and to her, the black students never look bright. [...] Juliet could even be perfectly aware of these facts about herself; she could aspire to reform; self-deception could be largely absent. We can imagine that sometimes Juliet deliberately strives to overcome her bias in particular cases. She sometimes tries to interpret black students’ comments especially generously. But it’s impossible to constantly maintain such self-conscious vigilance [...].”

Based on your observations of her words and deeds, what do you think she believes?

What do you think she thinks she believes?

If you are an African-American student in Juliet’s class, would you trust her?



Intention Recognition

Autonomous Systems and Moral Cognition

Moral considerations are pervasive in military operations, and aren't just limited to surveillance and usage of lethal force.



Not that soldiers on the North Korean side of the demilitarized zone can read this tale of Western decadence, but if they could they would do well to take note: South Korea has deployed two \$334,000 **robotic sentries** armed with automatic weapons and 40-millimeter grenade launchers along the tense border region bisecting the Korean peninsula.

[[Read Full Story](#)]

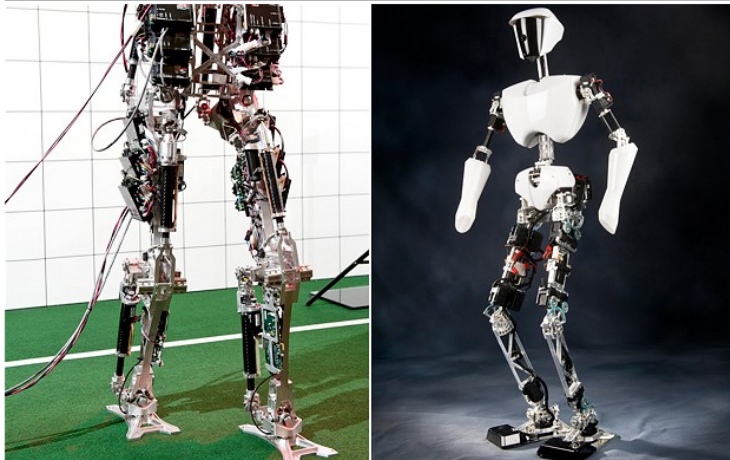
READ MORE ABOUT > [DEMILITARIZED ZONE, KOREA, MILITARY, MILITARY ROBOTS, NORTH KOREA, ROBOTICS, ROBOTS, SOUTH KOREA](#)

For Example...



US Navy's robotic firefighter

US Navy ships could soon be welcoming a human-like robot firefighter on board to reduce the risk of deaths and injuries.



The bottom half of ASH, left, is already up and walking. A lot of the technology used to create soccer star CHARLI-2, right, is being used to help ASH fight fires. Photo: Robotics and Mechanisms Laboratory/Virginia Tech

iRobot creates new business unit healthcare robotics

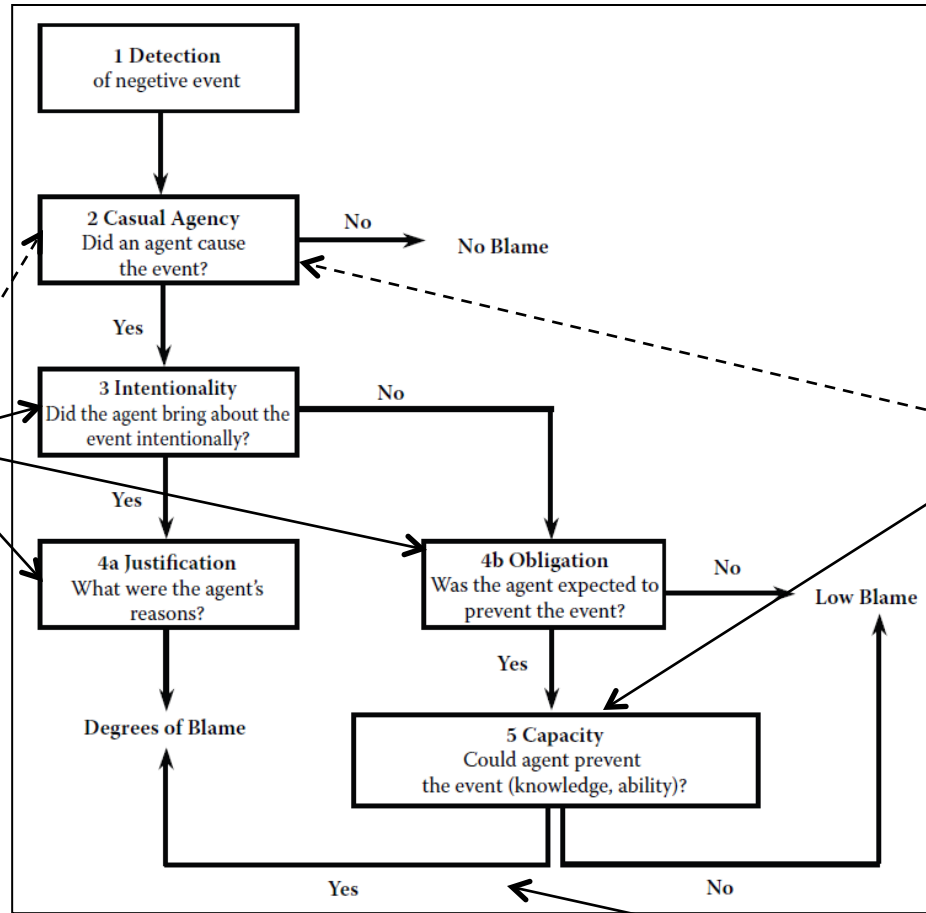
By Donald Melanson posted Nov 5th, 2009 at 9:12 PM

Well, it doesn't have a Roomba that will check up on your vitals just yet, but it looks like [iRobot](#) is betting on healthcare robotics in a fairly big way, with it taking advantage of the recent TEDMED conference to announce that it's forming a new business unit focused solely on the still burgeoning industry. That unit will be headed up by Tod Loofbourrow, who says that he believes the business "has the potential to make a significant difference in the field of healthcare," and adds that he thinks "the long-



The “MacGyver-Bot”

Blame



Beliefs, Desires,
Intentions, Obligations

Counterfactual
Reasoning

Courtesy of Prof. Bertram Malle

Negligence

Moral Cognition for Autonomous Systems: A Research Vision

A Clear Need

Motivator:

Responds directly to newly issued DoD Directive 3000.09 “*Autonomy in Weapons Systems*”

State of the Art:

Point solutions. Formalizing REGS, RoE, LOAC. None of these exhibit anything like **general moral competence or commonsense.**



Who Can Make it Happen?

Research Vision

Deep need to study, systematize, and formalize the constituents of moral cognition.

- *Moral Concepts*
- *Moral Emotions*
- *Moral Explanation*
- *Moral Reasoning and Judgment*

The Study of Moral Cognition is Highly Multidisciplinary!

Folks on the cutting edge: moral psychologists, social neuroscientists, developmental psychologists and experimental philosophers.

We deeply need their hard-won insights to shape and constrain the design of future intelligent systems.