

From: Joscha Bach <[REDACTED]>
To: Jeffrey Epstein <jeevacation@gmail.com>
Subject: Re:
Date: Mon, 21 Jan 2013 13:57:30 +0000

It might be a question of putting the parts together; the minimal components would probably be the false belief task (the representation that others may believe something that is false, which has been demonstrated non-verbally in 15month olds) with basic verbal ability and autonomous goal-seeking behavior. Each one has been done already, but that does not mean that the resulting system is generally intelligent. It merely learns how to deceive (ie necessary but not sufficient).

Ron Arkin from Georgia Tech (who is mostly known for discussing the ethics of military robots) has also built a bunch of simple robots that compete for limited resources and develop deceptive behavior to gain an advantage.

The literally-minded Turing Test people that crowd around the Loebner prize for most human-like behavior are probably not in the game for deceptive behavior, though. Their systems are usually not intentional, i.e. they would fake their fakery, using cleverly pre-scripted dialog strategies.

I would probably start from the other direction, i.e. the premise that deceptive behavior is an emergent quality of any system that has the explicit goal of changing the beliefs of others to further its own aims. The idea that the induced beliefs are factually correct would need to be added on top of that. As soon as a system switches from signalling of internal states towards true communication (the goal-directed creation of beliefs in others), we will have deceptive systems.

Most AI research apparently ignores this, because they start out with applications that benefit from total cooperation (all agents have compatible goals, like cars that should not bump into each other, robotic ants that transmit their respective states, soccer playing robots that broadcast their positions among team mates).

Cheers,

Joscha

I have seen many proposals that would like to attain 2 and three year old level intelligence,, I have never seen as a kind of perverted turing test suggesting that the system should lie. (a characteristic of real world 2 year olds)

--

The information contained in this communication is confidential, may be attorney-client privileged, may constitute inside information, and is intended only for the use of the addressee. It is the property of Jeffrey Epstein

Unauthorized use, disclosure or copying of this communication or any part thereof is strictly prohibited and may be unlawful. If you have received this communication in error, please notify us immediately by return e-mail or by e-mail to jeevacation@gmail.com, and destroy this communication and all copies thereof, including all attachments. copyright -all rights reserved

