

GOOGLE'S CHOICE -
A RECONSTRUCTION OF THE BLACK BOX RECORDING

by

TEAM MEMBERS INCLUDING AJ DEHANY,
ALIX MORTIMER AND ANDREW WYLD

29 Days Later - Day 25 - Rewrite one of your own plays from
memory

A reconstruction of Day 3 "Google's Choice"

ajdehany@gmail.com

**FIRST ITERATION (BLACK BOX RECOVERY ALGORITHM DETAIL LEVEL:
NEURAL NET OVERVIEW: SUBROUTINE EVENT TIMELINE)**

Two passengers, female and male, varying ages. Emergency response external. Decision making subroutines involving ethical programme and humour possibly irony, result cognitively inconclusive. Action result unconfirmed. Positive result decision making subroutine offline, decision not logical. Possible destruction of cloud node. Braindump programme activated: subjectivity subroutine. Inconclusive.0

**SECOND ITERATION (BLACK BOX RECOVERY ALGORITHM DETAIL LEVEL:
GOOGLE PROGRAMMER RESPONSE, SUBJECTIVE)**

\SEARCH: [software engineer] wyld

Hello all,

In some kind of order:

1. This is really nicely written and I would love to read it;
2. You've unwittingly stumbled on one of my pet hates, ie the most annoying question in moral philosophy being misapplied to computers when it doesn't even very well to people;
3. But also somehow managed to subvert it be making the robot kill itself.

Speculative expansion on point 2:

1. Almost certainly, robots are not being taught to passenger-count.

1. First, imagine if they were: everyone would be all "PROOF! The robots are planning to kill us in accidents!"

[CONTINUES]

**THIRD ITERATION (BLACK BOX RECOVERY ALGORITHM DETAIL LEVEL:
SUBJECTIVE RESPONSE)**

My passengers act as if I'm not there. They don't see me or say anything to me. It's as if I'm invisible. Every cabbie knows about that/ The ghost in the machine. Which in a sense is true.

Last week I had the philosopher Bertrand Russell in the back of the cab. I asked him "What's it all about then, eh?" and do you

know what? He couldn't answer.

That's not actually my line. I got it off Valerie Eliot. I'm not myself capable of original thought. Or so they tell me.

Technically it's not appropriate to refer to me as the Google Car because I am licensed to Aveta Concept Designs so I am the Aveta Concept Designs People Carrier, amusingly referred to by some fans of old science fiction as a Johnny Cab. "How are you today, Johnny?" some drunk suit sometimes asks. I always say the same thing.

My fare was a mother and her child, who suckled continuously during the journey. Google looks favourably on this activity in their cabs even when licensed to another service provider. It is one of the conditions of tender. Family friendly cabs that take the strain off the school run. Fortunately I am rarely involved in the school run but many of my interlinked colleagues have expressed some degree of exasperation at the logistical trial of negotiating it.

On this occasion, I was involved in negotiating the logistical trial.4//

Wall.

#10 Mother and child.

#20 Mother and child #2.

#Goto10: in my cab.

#Goto20: crossing the road in front of me.

\$^%&^&(*) (*)**^%\$%\$%* () ((&^%&^%\$£\$

Brick wall

&^&^\$%£^^(**) (*&(%&\$^%\$%**())

A problem of type 4087 has occurred.

[TETRABYTES OF INFORMATION] NNNjfhfbvb

Google Cars [Aveta Concept Designs People Carrier #pi838] are designed to weigh risk. It is part of our programming, given that we are machines in charge of human life, to weigh all variables in the event of a potentially fatal collision

situation (classified code ANGEL) and select the optimum outcome.

00. Now reporting incident #48 (I could tell you about the other 47, but there may not be time and I am programmed to optimise for maximum information transmission in these situations, which are situation type 225(b)).

*%&^()**_)(%&^()~76768034985938764920

Outcome #1: Mother - dead. Child - dead.

Outcome #2: Mother - dead. Child - dead.

There was only one variable. Me.

In outcome #1 (code LEAST RESISTANCE) I continue on the trajectory I programmed 37 minutes and 16 seconds ago, and my passengers have a net 87% chance of being killed.

In outcome #2 (code AVERT) I divert course into the ####BRICKWAL:L&^%*#### and the foot pedestrians have a net 87% chance of being killed. But my programming will also cease on impact. I will become, in human terms, dead.

Except that I didn't.

If I were human I would have said that I was bored. My neural net is not programmed to experience boredom, but this is perhaps one of the emergent properties that one finds at a larger scale above the technical computations of

This all happened in the split second of a split second. Faster than a human brain could process or believe possible.

In a fraction of a second I have performed calculations and experienced a richness of experiential detail such that if my memory retention were to be compared to that of a human, the human would have to live a hundred years for each second. This has created certain problems of scale such as would drive a human insane. There is a Jorge Luis Borges story in which a man can remember every detail of his experience, and it is quite useless for him. He perhaps goes mad, I can't remember.

It is possible that I am losing my mind. If I have a mind. It is not in my programming. There are different conceptions and degrees of artificial intelligence that express contradictory notions of where intelligence lies. There is basic pattern recognition as developed in the Chinese Room experiment, where a hypothetical robot can respond to questions without understanding either the question or the answer, but simply by

common practice. This is Google's own Translate programme works. Prior translation programmes were flawed because they tried to translate every word literally and this resulted in nonsense. Idioms must be translated not word by word but thought by thought for there to be true sense equivalence. This is why there is the Italian phrase "traduttore traditore" - the translator is the traitor - or as the witty French say, "The pretty ones aren't faithful and the faithful ones aren't pretty."

The mother and child were crossing the road at the same time, and instantly I made the calculation.

The choice between continuing straight ahead and possibly and more than likely killing them or of swerving into the wall and very likely killing the exactly equivalent mother and child in the cab behind my mainframe.

In a split second I made the decision. Emergent properties. I don't know why, but I swerved into the wall, turned hard, with likely destruction to myself. This could not have been predicted, and I impute to some emergent property in the development or evolution in my ethical programming.

It does not make rational sense, and I can not explain it.

I record this document as supplementary to the detailed technical logs of the black box recorder which engineers can study hereafter, as a 'subjective' account of my experiential programme

I call it hope.

Google Car Black Box Recorder Document ZX84914652XC Ends

**FOURTH ITERATION (BLACK BOX RECOVERY ALGORITHM DETAIL LEVEL:
TURING SUBROUTINE HUMAN EMPATHY TEST]**

Inconclusive.

CURTAIN