

Paper Clip Land

One of the major problems in the more serious AI Safety work today (by which we mean separate from all the people freaking out about the Singularity) is value alignment, which is the problem of making the system do what you really want it to, instead of a system that only works towards the specified tasks, ignoring the implicit constraints. For a hypothetical and extreme example, a system designed to maximize paper clips in the world should not, ideally, turn all the world's metal into paper clips.

Here is another, albeit rather abominable, value alignment problem. A computational linguist decided to reduce the amount of typing he needs to do to write emails, and wrote a quick bot to expand every language name abbreviation he plans to use to its full name.

Here is a list of correspondences:

ce - Chechen
ch - Chamorro
en - English
fr - French
he - Hebrew
is - Icelandic
ro - Romanian

However, it wasn't exactly the right code, because this is what happened to his email.

```
From: christine@ioling.org
Sent: Monday, 23 August 2010 11:38
To: spencer@ioling.org
Subject: Re: Time-saving script
```

```
um, spence? i think something has gone REALLY wrong
with your script...
```

```
On Mon, Aug 23, 2010 at 09:34 AM spencer@ioling.org
wrote:
> Hebrewy, ChamorRomanianrICHebrewcHebrewnlandic!
whEnglish
> you get a FrEnglishcHebrewe momEnglisht, cHebrewck
out
> thICHebrewcHebrewnlandic niCHebrewcHebrewn little
> pRomaniangram i wRomaniante.
> -- spEnglishCHebrewcHebrewnr
```

Q.1) What was the original message?

Q.2) He has, as it might be apparent, put an order by which the bot replaces every occurrence of the abbreviation with the full name of the language. Write the language codes in the priority order of replacement.

Q.3) What would his script do to "fresh fish from concentrate"?

Q.4) What alternate order would produce the longest output for Q.3?