Defining a Test for Epistolution

(the real Turing test)

Maybe there is a law of nature that states that it is impossible to possess knowledge before it has been created. This might require defining knowledge as the salience of information (possessible only by epistevolvers.) The consequence of this law should be that it is impossible to create an epistevolver that runs a fitness function. If an epistevolver ran a fitness function then that function would contain the knowledge that the epistevolver would need to discover simply to meet the definition of epistevolver, so there could be no genuine epistolution occurring. On the other hand, if there were an algorithm that specified steps to take with regard to an umwelt that were entirely open-ended and universal and would have different consequences in every environment, then that algorithm might be a candidate for the mechanism of epistolution.

On the other hand, an epistevolver can clearly be persuaded to tackle problems in an objectively accessible problem-space, because humans can do so. But the tricky thing is that this in itself is no proof of epistolution, because algorithms of all sorts can do this without any epistolution whatsoever. In a certain sense all of our existing mathematics does this, yet no existing math creates a semantics or a subjectivity. I believe this is a huge distraction that has fooled most of the AI engineers in our era into assuming that they have solved this problem when they have not even begun to address it. If we can somehow clearly exclude from the test any distracting “intelligent problem-solving” then the test can be a more accurate assessment of actual epistolution. I’m afraid this makes the project much harder but I think it is a necessary step.

I’ve been trying to define the coding problem in terms of what it would take to convince me that a submission to an “epistolution prize” has won the prize and genuinely solved the epistolution mechanism. My first draft of the criteria are below. Please refute, alter, or criticize these at will.

Minimal goals:

does not contain any fitness functions and
trains only on data imbedded in its umwelt, yet it
investigates salient features of its umwelt
anticipates patterns in its umwelt

seems to spontaneously intuit causal relationships
improves its ability to do these three things with practice (it remembers)
could possibly universally apply to problems in any domain, and
there is no obvious reason why every living organism cannot be an instantiation of it

Maybe-goals:

defends the boundary of the self
nests within and scales into larger, more competent entities
rewrites its configuration in periodic repair (sleep) cycles
re-enacts its external interactions in internal revisions (dreams)
controls or creates codes and/or genetic units
generates functional genotype-phenotype maps
recovers from amputation
recovers from major stochastic decay or disruption

Non-goals:

contains fitness functions in the relevant domain of learning
solves problems in objectively accessible domains
communicates solutions within a human umwelt
tends to survive or reproduce itself
growth or increase in number of components
develops syntactic sophistication in a high-level language