BAYES Packet v0.1
“I’m going to update on that”

Tossup 1 and only:

This theorem names a fictional monastic order in a series of short stories by Eliezer Yudkowsky. In the predictive processing model of the human brain, sensory evidence and mental models are integrated using methods that approximate this theorem. The statement of this theorem is part of the tagline of the blog Astral Codex Ten, along with the phrase “all the rest is commentary”. This theorem can be interpreted as a rule instructing the user on how to update their prior probabilities to posterior probabilities. For 10 points, what theorem states that the probability of the hypothesis given the evidence is equal to the probability of the evidence given the hypothesis multiplied by the prior probability of the hypothesis, all divided by the probability of the evidence across all hypotheses?

Answer: Bayes’ Theorem [accept Bayes’ Rule, Bayes’ Law, or other variations]

Bonus 1 and only:

For 10 points each, answer these questions about Artificial General Intelligence.

Part 1:
In 1965, the AI research I. J. Good predicted that an AI that was better than humans at AI design would be able to recursively self-improve to rapidly become even more intelligent, thereby producing what effect?
Answer: intelligence explosion [prompt on “singularity”]

Part 2:
However, a superintelligent AI may not necessarily act in the best interest of humanity. To demonstrate this, Eliezer Yudkowsky and Nick Bostrom invented what thought experiment in which an AI tiles the universe with copies of a simple object? This thought experiment inspired a similarly-named 2017 incremental game made by Frank Lantz.
Answer: paperclip maximizer [prompt on “paperclip”, I guess]

Part 3:
In order to build superintelligent AI that *does* act in the best interest of humanity, what field studied at the Machine Intelligence Research Institute attempts to understand agency and interpret machine learning systems? A popular introduction to this field, Human-Compatible, was written by AI researcher Stuart Russell in 2019.
Answer: AI alignment [accept “AI safety” or “friendly AI” or other common synonyms; prompt on “AI” or “machine ethics”, do not accept “machine learning” or equivalents]