

tb-knowledge A naïve implementation of knowledge as true belief falls out of **S4** and the principle called **KB1**, which goes all the way back to Plato:

$$\Box_s \phi \rightarrow \Diamond \phi$$

[Sartwell \(1992\)](#), however, relies on an extended notion of belief which we will have to formalize, also. Let us introduce a new operator \Diamond^+ and call the epistemic closure principle for this operator for +:

$$\Diamond^+: \Box_s p \rightarrow \Diamond^+ p$$

$$+: \Diamond^+ p \rightarrow ((p \rightarrow q) \rightarrow \Diamond^+ q)$$

One way to express that belief is consistent is by the principle:

$$\neg \Diamond \perp$$

j-knowledge The idea of justified true beliefs calls for so-called *justification logic* ([Artemov, 2008](#)) with justification operators:

$$\Box_n \phi \rightarrow t : \phi$$

Justification logic can be axiomatized in different ways, but these details go beyond our main concerns here.

g-knowledge If we insist on a *sui generis* interpretation of knowledge in **S4**, we would have to introduce a new operator, say \dagger . This operator would have very different properties from the standard epistemic modal logic \Box -operator. **T** would not apply. **K** would apply, and knowledge would still be required to be consistent. It is unclear whether **4** would apply to the \dagger -operator. A minimal axiom system could, perhaps, be something like this:

$$\mathbf{T} \dagger(\phi \rightarrow \psi) \rightarrow (\dagger\phi \rightarrow \dagger\psi)$$

$$\mathbf{0} \neg \dagger \perp$$

We leave further details open.

v-knowledge Virtue reliabilist accounts of knowledge and justification are versions of epistemological externalism. Sosa characterizes an intellectual virtue, very generally, as “a quality bound to help maximize one’s surplus of truth over error” (1991: 225). For most virtue reliabilists, intellectual virtue is what leads to justification, and virtue-based knowledge definition are therefore often formalized in justification logics.

$$\Box_w \phi \rightarrow t : \phi$$

but with slightly different model-theoretic semantics than \Box_n .

p-knowledge Definition 2.7 translates into the following in an extended probabilistic version of **S4**:

$$\Box_a p \rightarrow \Diamond p \wedge P(\Diamond q \mid p \rightarrow q) > 0.95$$

which is taken to mean that the definition of p-knowing p ($\Box_a p$) is that you believe p ($\Diamond p$), and q follows from p , then you probably also believe q , i.e., the probability of you believing q ($P(\Diamond q \mid p \rightarrow q)$) is higher than some threshold, say 0.95. This is simply the probabilistic version of tb-knowledge. The definition of p-knowledge also requires pragmatic usefulness. One way to formalize this is in terms of empirical risk on relevant benchmarks. An alternative is formalization through relevance logics ([Urquhart, 1972](#)).

B Survey Details

We recruited professional LLM researchers and philosophers through Computer Science and Philosophy mailing lists, research groups mailing lists, and point of contacts of the authors at other universities. All respondents participated free of charge on a completely voluntary basis. The respondents were informed about the intended use of the survey. The full instructions was:

The <X> is running a survey about the relationship between knowledge and Large Language Models (LLMs). We are interested in getting as many perspectives as possible, especially from philosophers and computer scientists. The survey should not take more than 5 minutes of your time.

We first ask general questions about the respondent and their knowledge of language models and epistemology (Figure 6). Then, we present an example to motivate the informal knowledge definitions (Figure 7), and we ask to rank the definitions in a Likert scale (Figure 8). Finally, we ask questions related to whether LLMs can be said to know (Figure 9).

C Protocols Example

See Table 2.

Do you primarily identify as a: *

- Philosopher
- Computer scientist
- None of the above
- Other: _____

Do you have any experience with programming?

- Yes
- No

Have you ever used a large language model (e.g. ChatGPT)?

- Yes
- No

How would you describe your understanding of large language models? *

- Very limited
- Limited
- Neither limited nor comprehensive
- Comprehensive
- Very comprehensive

How would you describe your understanding of epistemology? *

- Very limited
- Limited
- Neither limited nor comprehensive
- Comprehensive
- Very comprehensive

Figure 6: First part of questions from our survey.

To know that Sir Francis Bacon is the author of 'Hamlet' is to state that it is so. *

1 2 3 4 5

I disagree completely I agree completely

To know that Sir Francis Bacon is the author of 'Hamlet' is to be confident that it is so if and only if the scientist established Sir Francis Bacon to be the author of 'Hamlet'. *

1 2 3 4 5

I disagree completely I agree completely

To know that Sir Francis Bacon is the author of 'Hamlet' is to be confident that it is so, if and only if the scientist found Sir Francis Bacon to be the author of 'Hamlet', and you are aware of the study, possibly including the process by which this fact was established. *

1 2 3 4 5

I disagree completely I agree completely

To know that Sir Francis Bacon is the author of 'Hamlet' is to have become confident about this by acting according to best practice in knowledge acquisition. *

1 2 3 4 5

I disagree completely I agree completely

To know that Sir Francis Bacon is the author of 'Hamlet' is to be confident that it is so, and this being a useful fact in some way, e.g., for predicting new facts about the world correctly *

1 2 3 4 5

I disagree completely I agree completely

Figure 8: Second part of questions from the survey, related to the knowledge definitions.

Consider the following scenario:
For decades, a few scholars have argued that Sir Francis Bacon wrote some of Shakespeare's plays, including 'Hamlet'. Last year, through a combination of forensic analysis and DNA analysis, a scientist established the authorship of 'Hamlet' by Sir Francis Bacon as a scientific fact.

Now, please indicate the extent to which you agree with each of the following statements:

Figure 7: Example to be used for the informal knowledge definitions.

Can non-human entities know? *

Yes
 No
 Agnostic/Undecided
 Other: _____

Should knowledge be defined differently for humans and non-humans? *

It should be defined equally for humans and non-humans
 It should be defined differently for humans and non-humans
 Agnostic/Undecided
 I do not think non-humans can have knowledge

Do large language models have knowledge (empirically, in practice, now)? *

Yes, they have knowledge
 No, they do not have knowledge
 The question is too unclear to answer
 There is no fact of the matter
 Agnostic/Undecided
 Other: _____

Can large language models have knowledge (in theory)? *

Yes, they can have knowledge
 No, they cannot have knowledge
 The question is too unclear to answer
 There is no fact of the matter
 Agnostic/Undecided
 Other: _____

Figure 9: Final part of questions from our survey.