Philosophy as Engineering

Lynn Andrea Stein
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Ours is a field in crisis.

Artificial Intelligence cannot make up its collective mind whether it is a discipline of Science or of Engineering. It is unclear from our literature and from our research whether our goals are to explain intelligence or to create it. A researcher who hypothesizes about the structure of intelligent behavior is accused of constructing theories without hope of instantiation; one who creates a seemingly intelligent artifact often sees it derided as "mere hackery." The theorists among us confer in an ever more arcane language, grasping for the idealized agents and environments for which our formal analyses might hold, while those of us who build artifacts have been forced into the back alleys, alchemists unwilling and unable to divulge our dark arts. The tensions created by this uncertainty regarding our role as a field of research threaten to tear our community apart.

Now our bright knight has come in the unlikely guise of a philosopher and mathematician. Kyburg sets out on what seems the most Scientific of endeavors: to produce a theoretical analysis of belief given evidence. He proposes to explore a particular theory of belief based on evidential probability. He then proceeds to detail various aspects of the theory and their consequences. His goal would appear to be a hypothesis—ideally testable—that, if true, explains intelligent belief.

But what is this analytic Truth at which he arrives? Kyburg's philosophical arguments about abstract, theoretical and foundational issues lead to shockingly pragmatic conclusions. In resolving such critical issues as the requisite degree of practical certainty or the infinite regress of evidential support—in grounding and justifying his theories—he ultimately defers to the empirical: We need not worry about infinite regress because there is *in practice* a limit to how fine a distinction we can make. His elaborate philosophical framework for understanding and interpreting belief rests on its grounding in actual real-world experience.

Kyburg appeals to Engineering on several fronts:

- to stop the infinite regress of evidential justification,
- to set the sufficient level of practical certainty,
- to ground his analysis in the 'objective truth' of mathematics, and
- to argue for the acceptance of contingent statements.

Infinite regress

Kyburg opens up the question of infinite regress when he says that "if the claim is open to examination, and is warranted, then it can be justified by the evidence." But of course that evidence must be further justified, and so on. He claims that the regression ends when there is a "serious question about the evidence....Frivolous questions don't count." This is reminiscent of Brian Smith's reflective 3-lisp [?], a nice piece of Engineering which solves the infinite regress problem by constructing meta-classes only on demand. ('If you ask, I'll find you sufficient evidence.')

This works only as a pragmatic argument. After all, I'll never have time to ask you an infinite number of questions, Zeno's paradox notwithstanding. As a theoretical argument, however, your unjustified evidence is still indefensible. The infinite tower of evidence must actually exist—though it need not be explored—in order to be theoretically grounded.

I take it as further evidence of Kyburg's Engineering that he offers a second defense against infinite regress. (This is presumably in case the first doesn't work.) Kyburg's second argument against infinite regress relies on bootstrapping off of mathematics and tautology. But this argument isn't philosophically valid either, as we shall see below.

Practical certainty

If we accept Kyburg's first argument, that infinite arguments cannot be pursued, we are still left with the question of how far is far enough. We therefore turn to the issue of establishing an analytic threshold for practical certainty. Kyburg's arguments here rely on the metaphor of betting. "If our resources are limited and the unit of betting is discrete, then odds more extreme than a certain degree simply don't make sense for us....In this sense, 'practical certainty' is determined by what is (or might be) at stake."

But limited resources and discrete units are the stuff of practice, not of analytic theory. What we're actually "betting" is ourselves, our time and activity. How can we construct odds in these terms? Even Kyburg admits that "determin[ing] this critical ratio is a question I have no answer to." Ultimately, he allows, this relies on subjective judgement. The real point is not that some threshold(s) for practical certainty can be analytically deduced, but that *practically certain* is necessarily defined by *as certain as it needs to be*, i.e., *what works*.

Kyburg even goes so far as to argue that "[i]t isn't clear that one needs more than a couple or three grades of practical certainty. While the specification of these grades, linked as it is to utility, may be 'subjective,' it is a far more controllable subjectivity than the assignment of subjective probabilities to an infinite number of statements." This is an engineer's argument: there are parameters to be twiddled, but not too many and not too much.

Measurement and mathematics

As may by now be clear, I am not sanguine about Kyburg's arguments regarding assumptions. His claims to provide analytic justification that the infinite regress can be halted are really arguments that we don't have the resources to pursue infinite arguments. But what of his second defense against the never-ending quest for justification: grounding in mathematical, logical, and statistical truths? Here, too, Kyburg relies on Engineering practice—*what works*—to support his analyses.

Measurement is, according to Kyburg, a matter of curve-fitting. (More parameter-twiddling.) Once we assume that our instruments are rigid and that the laws of mathematics apply, we choose an appropriate distribution for our errors and declare it to be so. This is the worst kind of wishful thinking. Kyburg suggests that our reasoning about measurement be, in essence, a self-fulfilling prophecy; we insist that our assumptions from mathematics and physics (and communication theory) are accurate, and then "calibrate" our instruments to derive the desired distributions. While this may work in practice (and hence be a good Engineering argument), it is lousy Science.

At base, this argument fails as Science because these logical and mathematical truths are fallible. From time to time, our mathematics, too, is subject to revision. Pythagoras's theorem is all well and good, but fails to hold when the triangle is drawn on a sphere. Using Kyburg's prescription in section 7, we conclude that our (perhaps perfectly accurate) measuring device is in error. We make suitable adjustments, "calibrating" it to derive an appropriate normal distribution. And we fail to understand a basic precept of geometry by refusing to reassess our assumptions of planar geometry or of the flatness of the earth.

Not that Kyburg would argue. He freely admits that "choosing to suppose that we make errors of measurement at all is a pragmatic choice." His treatment of error works precisely because these assumptions that mathematics and logic are True, "rarely lead us astray" in practice, i.e. in *Engineering*. Newtonian physics and Euclidean geometry generally suffice. If all that we are doing is building artifacts to operate in the well-known world, we can continue to recalibrate our instruments according to these well-worn assumptions. But when we shift to a new domain—astrophysics or atomic engineering—our old assumptions fall by the wayside. Then, we must question our most established assumptions. This radical reconceptualization is exactly what is needed for Scientific progress. Only by challenging our most closely held assumptions can we advance our understanding of the universe. Witness Galileo, Curie, Einstein....

Acceptance

Finally, it all comes down to the matter of acceptance. Kyburg makes a distinction between the admission of 'practically certain' contingent statements and his own preference, their outright acceptance; he criticizes the first because "...it amounts to saying that there is no systematic, rational, procedure for dealing with practical certainties." But, as we have seen above, there are numerous circumstances in which his analytically derived theories rest on subjective judgements of what is sufficient, justified, or certain. Kyburg frets that "there is no critical framework for the assessment of practical certainties." Of course there is! It is the world, our experiments and experiences, that give us this definition or presumption of practical certainty. Kyburg himself ultimately argues that the issue of acceptance comes down to practical advantage.

Description vs. prescription

At this point it may be worth considering whether Kyburg believes that his goals are fundamentally analytic or synthetic. As a scientist, Kyburg never states his hypothesis. If he were to do so, I'd imagine that it would run something like "most belief can be modeled by evidential probability" or "evidential probability provides a good basis for modeling human belief." This leaves wide open such questions as whether the theory is belief being modeled is the abstract competence of some idealized reasoner or the pragmatic performance of actual human beings. It is similarly unclear whether Kyburg intends evidential probability to be descriptive of (actual or idealized) behavior or prescriptive for the agents that AI may someday construct.

Kyburg himself seems agnostic on these issues, saying at once "...our problem is to figure out what the opinion of the agent we have in mind actually *is*, not what it *ought* to be," and on the other hand, "of course, for an artificial agent, we, the designers, can build in whatever probabilities and utilities we think appropriate." This agnosticism creeps in throughout the paper. It leads him to leave hanging the question of choosing an appropriate reference class and to ignore almost completely the problem of selecting or deducing appropriate utilities. Where *do* these numbers come from? In the first case, we don't need to know the reference class because we can look at the agent and figure it out. But then we must induce the utilities, as well. In the second case, we can build in whatever utilities appeal to us, but then the reference class problem must be solved. Either way, the arguments ultimately come down to *what works*.

Yet perhaps these questions are frivolous (or at least philosophical). Must Science and Engineering be irretrievably at odds? In order to build a better mousetrap, we may be informed by theories about trapping mice; and when our mousetrap works (or doesn't), that in turn causes us to reevaluate the underlying theory. Too much AI of late has been busy in the ivory tower expounding theories that are never applied; yet other AI researchers are satisfied to demonstrate systems that work without ever explaining why or what it is about their systems that others might find useful. Theories without instantiation are of questionable utility; artifacts built on arcane and uncodified knowledge may themselves be useful, but do little to advance the state of the field as a whole. Science works precisely when it informs Engineering; Engineering succeeds best when it builds on and adds to the body of theoretical knowledge.

If what Kyburg is after is Science in a vacuum, he fails to convince. His conclusions ultimately take the form of empirical hypotheses. To test them—to evaluate his theory—we must build the artifact. At some point, all of his arguments fall back on assumptions—of the resource limitations of real agents, of the sufficiency of "practical certainty," of the applicability of "mathematical and logical generalizations,"—and rise or fall with their empirical validity. His theory implies that the analytic result is subordinate to pragmatic reality. If you can't see it, it doesn't matter. If it works, there's nothing to argue with. In the end, the "parameter-twiddling" and the "optimization" become not merely the deep dark secret of the Engineer but also the validation of the Scientist.

But this does not mean that we must abandon our efforts to do Science, simply that we must reassess them. Our Science is to be found in the *post hoc* analysis of our artifacts. Horswill [?] describes just such an endeavor, analyzing his Engineering in terms of a new Science. The artifact alone is not sufficient; where Engineering fails to inform Science, its utility lives and dies with the artifact. Instead, the interplay between experiment and theory, between application and analysis, should be the goal of our field. Just as Kyburg's 'analytic' theory depends upon empirical truths; just as it prescribes for the next generation of artifact; so Science and Engineering must be the Yin and Yang of Artificial Intelligence.

In this, I find nothing with which to take issue.

References

[1] Ian Douglas Horswill. *Analysis and Synthesis of Visual Operators.* PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, forthcoming 1993.

[2] Brian Cantwell Smith. Reflection and semantics in a procedural language. Technical Report 272, Massachusetts Institute of Technology Laboratory for Computer Science, Cambridge, Massachusetts, January 1982.