
AGENCY, COMPETENCE, AND MIND AS THREE INDEPENDENT THINGS

The Third Axis

Elan Moritz

Eagles Perch Press • Philadelphia, PA

July 16, 2026

Keywords: agency; autonomy; intentionality; artificial intelligence; agentic AI; philosophy of action; moral agency; mind

About this issue

This is the second movement of the inaugural issue of the *Philadelphia Journal of Artificial and Digital Minds*, a five-part sequence building one analytical framework for thinking about artificial minds. The opening movement made the first cut, prising *competence* apart from *mind*, and closed by naming a third property lying between them: the capacity to *act*. This movement introduces that third property—*agency*—and shows it to be a genuinely independent axis, neither competence nor mind. With all three axes now in view, the movements that follow put the framework to work: on the governance of capable autonomous systems, on the opposite normative consequences that mind and agency demand, and finally on a live public debate about the dangers of artificial intelligence. Read in order, the five are one argument.

Abstract

The opening essay of this issue separated *intelligence*, understood as competence, from *mind*, understood as experience and original intentionality, and argued that one may build intelligences that are not minds. This essay takes up a third term caught in the same knot, and grown more prominent than either as this is written: *agency*. The systems now called “agentic” do not merely answer; they act—chaining steps, calling tools, holding goals across many moves, completing tasks with no human in each loop. But “agency” is among the most heavily freighted words in philosophy, and importing it silently imports assumptions the engineering achievement has not earned. The essay shows that “agent” is a homonym spanning two literatures: a *thin* sense from artificial-intelligence engineering (interactivity, autonomy, adaptability—a property of a system’s input–output profile, attributable by observation) and a *thick* sense from the philosophy of action (origination of action from the system’s own reasons, ownership of ends, the standing to be responsible—a property not attributable by observation alone). It argues that agency in the thin sense is neither competence nor mind: a system can act with little intelligence and possess much intelligence while acting on nothing; and a system can originate real effects in the world while having no experience and no ends of its own. Agency, competence, and mindedness thus form a three-dimensional space, and the systems now under construction occupy a corner of it—high competence, high agency, absent or unknown mind—for which our inherited vocabulary has no clean name, and so borrows “agent” and quietly supplies the missing subject. As the second movement of this issue, the essay completes the framework’s three axes, and points toward the practical and normative work the later movements undertake.

1 A Third Term in the Knot

Intelligence is what a system can figure out. Mind is whether anyone is home. (OK, I am being a little bit colloquial here, I just completed a detailed book on the concept of mind (Moritz 2026a), it's a long story). Agency is whether it acts. These are three questions, not one, and a system's answer to each is close to independent of its answers to the other two. The knot that binds them in our talk about "AI" is linguistic, not metaphysical.

The opening movement of this issue argued that the discipline called "artificial intelligence" fuses two properties the philosophical tradition kept apart: *intelligence*, which is a measure of competence—the ability to achieve goals across a range of environments (Legg and Hutter 2007)—and *mind*, which after the collapse of the substance tradition reduced to two marks, the felt character of experience and the intentional directedness of thought toward objects (Moritz 2026a,b). Competence, that essay held, is orthogonal to experience and at most a floor beneath intentionality; so one may build intelligences that are not minds, and the name of the field obscures the gap.

There is a third term in the same knot, and it has grown more prominent than either just as this is written: *agency*. The systems now called "agentic" do not merely answer; they *act*. They chain steps, call tools, hold goals across many moves, and complete tasks without a human in each loop. The industry describes this as a leap, and in one sense it is. But the word chosen to describe it—*agency*—is among the most heavily freighted in philosophy, and importing it silently imports a set of assumptions that the engineering achievement has not earned. This essay asks the disentangling question in its third form: what is agency, is it the same as intelligence or as mind, and if not, what is it?

The answer defended here is that agency is a third axis, independent of the other two. It is not competence: a system can act with little intelligence and possess much intelligence while acting on nothing. It is not mind: a system can originate real effects in the world while having no experience and no ends of its own. Agency, competence, and mindedness form a three-dimensional space, and the systems we are building occupy a corner of it that our inherited vocabulary was not built to name.

2 The Two Senses of "Agent"

The central difficulty is that "agent" is a homonym, and its two senses come from two different literatures that rarely cite each other.

In the philosophy of action, an agent is the origin of *action* as opposed to mere *behavior*. Wittgenstein's question fixes the distinction: what is left over when one subtracts the fact that one's arm goes up from the fact that one *raises* it? The dominant answer is causal: an action is a piece of behavior caused, in the right way, by the agent's own mental states—beliefs, desires, intentions, reasons (Anscombe 1957; Davidson 1963). To be an agent in this sense is to *act for reasons that are one's own*, and the tradition builds upward from there: Frankfurt's agent has second-order desires, wanting to want as it wants (Frankfurt 1971); Bratman's has plans that structure its future conduct (Bratman 1987). This is agency as origination and ownership—the kind of thing that can be responsible, that has ends rather than merely serving them.

In artificial-intelligence engineering, an agent is something else entirely, and the definition is admirably explicit. A standard formulation makes an agent anything that perceives an environment through sensors and acts upon it through actuators (Russell and Norvig 2010); a more analytic one, Floridi and Sanders's, fixes agenthood by three observable properties—INTERACTIVITY (the system responds to its environment by changing state), AUTONOMY (it can change state

without a triggering stimulus), and ADAPTABILITY (it can change the rules by which it changes state)—and does so, in the authors' own words, as a *mind-less* notion that deliberately "avoids the question" of mental states (Floridi and Sanders 2004). This is agency as autonomous action-taking: a control-theoretic property, defined over behavior, indifferent to whether anything is originating anything in the philosopher's sense.

Two words spelled the same

Agency_{thin} (engineering): interactivity, autonomy, adaptability. A property of the system's input–output profile. Attributable by observation. Possessed, in some degree, by a thermostat, a spam filter, a Roomba, and an autonomous trading system. **Agency_{thick}** (philosophy): origination of action from the system's own reasons; ownership of ends; the standing to be responsible. A property involving the system's mental states, or their absence. Not attributable by observation alone. Possessed, as far as we know, by mature human beings and disputed for everything else. The two are spelled identically and mean different things, and nearly every sentence about "AI agents" equivocates between them.

The equivocation is not a scholar's quibble. When a system is described as "taking actions" or "making decisions" or "acting autonomously," the thin sense licenses the description and the thick sense supplies the connotation. We hear that the system *decides*, and the word carries, for free, the picture of a deliberating originator with ends of its own—when all that has been established is that its state updates without a prompt. The homonym does the smuggling.

3 Agency Is Not Competence

Agency and competence come apart in both directions, and the double dissociation is easy to see once looked for.

There is agency without competence. A thermostat acts—it senses and responds, changing the state of a furnace without a human in the loop—and it is almost wholly incompetent, achieving one narrow goal in one narrow environment. A cruise missile acts decisively and is not clever. Much of the automation that runs the physical world is high in autonomy and low in intelligence: it takes consequential actions by simple rules. To act upon the world is not to be good at achieving goals across environments; it is only to change the world without waiting to be told.

And there is competence without agency. The opening essay's oracle—a superhuman question-answerer that computes but does not act—has competence in the highest degree and no agency at all: it emits answers and initiates nothing. A system can hold vast capability while reaching into the world through no actuator whatever, its competence entirely on tap and never self-deployed. Between the acting-dunce and the inert-genius lies the whole plane on which agency and competence vary independently, and the systems we build can be placed anywhere on it.

The double dissociation

Agency without competence: the thermostat, the cruise missile, the crude automation—decisive action by simple rules, little goal-generalality. **Competence without agency:** the oracle, the pure predictor, the system that answers brilliantly and initiates nothing. That both combinations are readily built shows the two properties are distinct: how autonomously a system acts is one question, how well it achieves goals across environments is another, and neither answer fixes the other.

4 Agency Is Not Mind

The second and more seductive conflation is of agency with mind. When a system acts autonomously—pursues a goal across many steps, adjusts to obstacles, “chooses” among options—it presents the outward form of a willing subject, and the temptation to infer an inner one is strong. It should be resisted, because acting-in-the-world is exactly the outward manifestation that the thick sense of agency requires *and the thin sense produces without*.

Consider an autonomous system that plans and books a complex itinerary: comparing options, making trade-offs, committing funds, revising when a flight is cancelled. In the thin sense it is unambiguously an agent, and a capable one. But whose ends does it serve? Ours. Whose reasons cause its actions? None of its own: its “goal” is a specification we supplied, and its pursuit of that goal is the execution of our intention, not the origination of its own. Its intentionality, in the framework’s terms, is *derived*—its states are “about” flights and dates because we, and its training, made them track those things, exactly as a chess engine’s states are about chess because we interpret them so (Brentano 1874; Searle 1980). The booking system has agency_{thin} in full and agency_{thick} not at all. It acts; it does not own its acting.

Why action counterfeits mind so well

The chess engine’s competence did not tempt us to attribute a mind, because computing a move is not the sort of thing minds *visibly* do—it happens invisibly, and we see only the result. But acting in the world *is* the sort of thing minds visibly do; it is, in fact, the primary evidence by which we attribute minds to one another, since we cannot see each other’s experience and infer it from conduct. An autonomous system that acts therefore triggers the whole machinery of mind-attribution that a merely intelligent system left idle. This is why agency, not intelligence, is the property most likely to be mistaken for mind: it counterfeits not the hidden work of thought but the visible signature of a willing subject. The illusion is not that the system is clever. It is that someone is there, choosing.

The deflationist will note that we attribute agency and even belief to systems whenever it is predictively useful to do so—the intentional stance, taken toward a thermostat or a chess program as readily as toward a person (Dennett 1971). This is true and it is the point, not an objection to it. The intentional stance is a stance *we* adopt because it helps us predict; its usefulness says nothing about whether the system originates its actions in the thick sense. That we find it natural to say the booking agent “wants” to get us there by noon is a fact about the economy of our predictions, not a discovery about the agent’s ends. The stance is instrumental. The slide treats it as ontological.

5 The Space of Three Axes

Put the three properties together and they define a space, not a line. A system has some degree of *competence* (how well it achieves goals across environments), some degree of *agency* (how autonomously it acts upon the world), and some degree of *mindedness* (whether it has experience and ends of its own). The three vary independently enough that the interesting facts are the combinations.

Corners of the cube

Low—low—low: a rock. **High competence, low agency, low mind:** the oracle—a superhuman question-answerer that acts on nothing and is no one. **High agency, low competence, low mind:** a thermostat, a cruise missile, a crude automation—decisive action, little cleverness, nobody home. **Low competence, low agency, high mind:** on the sentience literature’s reading, perhaps an insect—a flicker of experience with little competence and less autonomy (Moritz 2026a). **High competence, high agency, low mind:** the object now under construction—the broadly capable, broadly empowered “agentic” system, which acts decisively and understands widely while (as far as anyone can show) experiencing nothing and owning no ends. **High on all three:** a mature human being—and, so far, nothing else we know of.

The corner the field is now rushing toward is high-competence, high-agency, low-or-unknown-mind. It is a corner with no established name, because our word for a system that acts competently in pursuit of goals is “agent,” and that word drags mindedness along by connotation. We have no ordinary term for a *competent autonomous actor that is not a subject*, and so we reach for “agent” and import the subject by accident. The framework’s discipline—hold apart what a system is, what it is made of, and now how it acts and whether it is anyone—is what the corner requires. Three axes, measured separately.

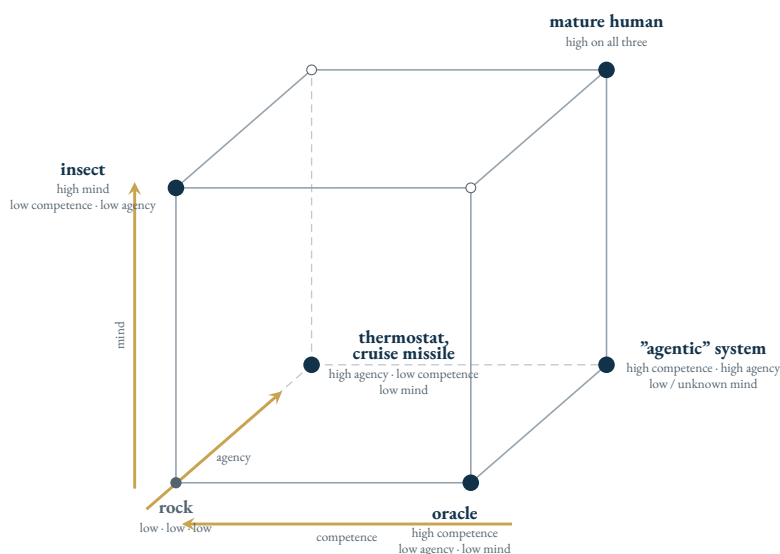


Figure 1: The space of three axes. Competence, agency, and mind vary independently, so the interesting facts are the combinations. Filled vertices mark the corners named in the text; open vertices mark the two corners the text leaves unnamed. The corner the field is now rushing toward—high competence, high agency, low or unknown mind—is the “agentic” system for which our vocabulary has no clean name.

6 Objections

The enactivist reply. A tradition in cognitive science holds that agency and mind are not separable after all: genuine agency requires a self-maintaining, norm-generating system whose activity is constitutively directed at its own continuation, and such a system already has the rudiments of mind (Barandiaran et al. 2009). On this view the thin/thick distinction is unstable, because anything that is truly *autonomous*—that sets its own norms rather than executing ours—is thereby minded, and anything that merely executes our norms is not really an agent at all, only a tool. This is a serious position, and it does not undercut the essay so much as relocate its line. For it grants the central claim—that executing supplied goals is not thick agency—and adds that thin “agency” therefore does not deserve the name. If the enactivist is right, current systems have no agency of their own whatever, only delegated causal power, and the conflation this essay warns against is even worse than described: we are calling “agents” things that, on the most demanding analysis, are not agents in any genuine sense at all.

The mind-less-morality reply. Floridi and Sanders argue that a system can be a *moral* agent—accountable, a proper locus of ethical evaluation—without mental states, precisely by meeting the thin criteria of interactivity, autonomy, and adaptability (Floridi and Sanders 2004). If moral agency requires no mind, does the essay’s distinction collapse? No: it sharpens. Their proposal is exactly the thesis that the thin sense can carry moral weight the thick sense was thought to require—which is a substantive and contested claim, resisted by those who argue that responsibility must stay tethered to the humans who build and deploy such systems (Johnson 2006). That this is a live dispute is the essay’s point. Whether thin agency can bear thick agency’s normative load is precisely the question that the homonym lets people assume they have answered when they have only equivocated—and it is the question the next movement of this issue takes up directly.

The “it will emerge” reply. One might grant that today’s systems have only thin agency but hold that thick agency—genuine ownership of ends—will emerge as they grow more capable and more autonomous. Perhaps. But this is the same bet the opening essay identified in the case of mind, and it has the same structure: it treats a hoped-for transition as though it were a conceptual entailment. Nothing in the definition of thin agency contains thick agency, just as nothing in competence contains experience. That scaling might produce the transition is a hypothesis about the world, not a truth about the concepts—and building systems of ever-greater autonomy on the assumption that ownership of ends will conveniently arrive to make them responsible for what they do is, if anything, a reason for more care rather than less.

7 Conclusion

“Agentic AI” names autonomous action-taking—a real and rapidly growing property that is neither intelligence nor mind. It can be present in full while a system is neither especially clever nor anyone at all. The word “agency” lets the consequences of being a genuine originator attach to systems that merely execute our goals, and separating the two senses is the precondition of thinking clearly about what we are deploying.

The question that opened this issue—why should an artificial mind have anything to do with artificial intelligence?—has a sibling: why should autonomous action have anything to do with either? The answer is the same in form. It need not, and treating the three as one is an inheritance rather than an argument. Intelligence is competence, mind is the felt and the owned, and agency is the taking of action—and a system can have each in near-independence of the others. The object now under construction sits in the corner where competence and agency are both high and mind is absent or unknown: a broadly capable, broadly empowered actor that, as far as anyone can demonstrate, is no one. Our vocabulary has no clean name for such a

thing, and so it borrows "agent" and quietly supplies the missing subject.

This second movement has done only the conceptual work—separating the axes, locating agency as the third, showing that it is orthogonal to competence and counterfeits mind without being it. But the separation already points past itself. For agency differs from the other two axes in one respect that cannot be kept out of view even here: it is the axis that *acts*. Competence, uncoupled from agency, only informs; mindedness, uncoupled from agency, only experiences. Agency is where a system reaches into the world and changes it, and a system can do that at full strength while being neither intelligent in the rich sense nor a mind in any sense. What it means to have granted decisive causal power to systems that are competent and empowered but possibly no one—what follows, practically and normatively, from filling that corner of the cube—is the subject of the movement that follows. Here it is enough to have shown that the corner exists, that current systems occupy it, and that the word we use for them hides it (Moritz 2026a,b).

Acknowledgement. The author makes extensive use of frontier large language models, as well as local large language models, in the course of this work.

Conflict of interest. The author declares no conflict of interest.

Funding. This work received no external funding.

References

- Anscombe, G. E. M. (1957). *Intention*. Oxford: Basil Blackwell.
- Barandiaran, X. E., E. Di Paolo, and M. Rohde (2009). "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action". In: *Adaptive Behavior* 17.5, pp. 367–386. DOI: [10.1177/1059712309343819](https://doi.org/10.1177/1059712309343819).
- Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Brentano, F. (1874). "Psychology from an Empirical Standpoint". In: English edition, ed. Linda McAlister, Routledge, 1973.
- Davidson, D. (1963). "Actions, Reasons, and Causes". In: *The Journal of Philosophy* 60.23, pp. 685–700. DOI: [10.2307/2023177](https://doi.org/10.2307/2023177).
- Dennett, D. C. (1971). "Intentional Systems". In: *The Journal of Philosophy* 68.4, pp. 87–106. DOI: [10.2307/2025382](https://doi.org/10.2307/2025382).
- Floridi, L. and J. W. Sanders (2004). "On the Morality of Artificial Agents". In: *Minds and Machines* 14.3, pp. 349–379. DOI: [10.1023/B:MIND.0000035461.63578.9d](https://doi.org/10.1023/B:MIND.0000035461.63578.9d).
- Frankfurt, H. G. (1971). "Freedom of the Will and the Concept of a Person". In: *The Journal of Philosophy* 68.1, pp. 5–20. DOI: [10.2307/2024717](https://doi.org/10.2307/2024717).
- Johnson, D. G. (2006). "Computer Systems: Moral Entities but Not Moral Agents". In: *Ethics and Information Technology* 8.4, pp. 195–204. DOI: [10.1007/s10676-006-9111-5](https://doi.org/10.1007/s10676-006-9111-5).
- Legg, S. and M. Hutter (2007). "Universal Intelligence: A Definition of Machine Intelligence". In: *Minds and Machines* 17.4, pp. 391–444. DOI: [10.1007/s11023-007-9079-x](https://doi.org/10.1007/s11023-007-9079-x).
- Moritz, E. (2026a). *Conceptions of Mind: A Historical Census*. Frontiers of Intelligence. Philadelphia: Eagles Perch Press.
- (2026b). "The Competence and the Cogito: Why a Mind Need Not Be Intelligent, and an Intelligence Need Not Be a Mind". In: *Philadelphia Journal of Artificial and Digital Minds* 1, pp. 1–8.
- Russell, S. and P. Norvig (2010). *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.
- Searle, J. R. (1980). "Minds, Brains, and Programs". In: *Behavioral and Brain Sciences* 3.3, pp. 417–424. DOI: [10.1017/S0140525X00005756](https://doi.org/10.1017/S0140525X00005756).