The Human Lesson

A Response to Sutton through Wittgenstein, Lewis, Dennett, and Nagel

Michael Stoyanovich
michael@mstoyanovich.com

Disclaimer

This paper is intended for informational and educational purposes only. The views and analyses presented - particularly those related to ethics, policy, and AI system design - reflect the author's interpretations and do not constitute legal, regulatory, or professional advice. Readers are encouraged to critically assess the content and consult appropriate experts or authorities before applying any concepts discussed herein. The author assumes no liability for any decisions or actions taken on the basis of this work.

Abstract

This essay offers a human-centered counterpoint to Richard Sutton's "Bitter Lesson." While scalable, general methods reliably outperform handcrafted theories, performance alone cannot substitute for the conditions under which meaning and responsibility arise. Drawing on four philosophers, it argues that current AI does not cross four boundaries: (1) Wittgenstein - meaning is use within a form of life; (2) Lewis - coordination demands common knowledge and public recognition, not mere convergence; (3) Dennett - competence without comprehension yields results but not reasons; and (4) Nagel - intelligence without subjective experience remains ethically incomplete. The conclusion is practical: computation can expand capability, but it cannot confer conscience. Unless we pair optimization with orientation, we risk building systems that stabilize behavior while automating indifference.

Preface

As someone involved in guiding organizations' governance of artificial intelligence systems, I often find myself in spaces where technical efficacy is considered paramount. Sutton's "The Bitter Lesson" rightly captures a core empirical truth: scalable, general methods powered by computation consistently outperform our handcrafted theories. That observation is not just insightful - it has shaped the architecture of modern machine learning. But when applied too broadly, it becomes not just a lesson, but a worldview. And that worldview risks discarding precisely what cannot be optimized: meaning, responsibility, and experience.

This essay is not an argument against machine learning, nor a nostalgic defense of human uniqueness. It is a counterweight to the dominant gravity of performance metrics. Drawing on four philosophical frameworks - Wittgenstein, Lewis, Dennett, and Nagel - I offer a companion thesis: that even as machines excel, something essential is lost when we equate intelligence with statistical success. The bitter lesson is true. But it is not the only one.

0 - The Bitter Lesson

For seventy years we have watched a pattern emerge. Time and again, the systems that succeed are not those that encode our insights, but those that scale with computation. The neural networks that now outperform us in image recognition, protein folding, and language generation were not built on a deep understanding of vision, biology, or meaning. They were built on architectures that could learn - on their own - at massive scale. That is the lesson of the laboratory. That is the truth we ignore at our peril. Each decade confirms it: general methods that scale with computation surpass those refined by human understanding.[1]

But there is another lesson, no less empirical and far older: a lesson from the humanities, from social practice, from lived experience. It does not deny the power of optimization. But it insists on a boundary to what performance alone can offer. This, too, is a bitter lesson - because it means that some things cannot be outsourced.

1 - Meaning Without Life (Wittgenstein)

Wittgenstein taught that meaning lives in use, within the shared practices of a form of life.[2] Language is not merely code to be parsed; it is a way of being together. To understand a word is not just to decode its structure - it is to grasp the human context in which it is used, to share in the practices that give it life.

Our machines learn correlations, not communities. They can mimic our language-games, but they do not inhabit them. They generate fluent outputs without participating in the forms of life that give those outputs meaning. If meaning is use, and use is rooted in shared life, then a system without life - biological or otherwise - cannot mean what it says (on today's evidence).[3]

This is, admittedly, an interpretive extension. Some may argue that if machines become embedded in our social practices, they might someday participate in a new form of life. But for now, their performance remains disconnected from lived practice.

2 - Coordination Without Consciousness (Lewis)

David Lewis showed that social coordination depends on common knowledge - on the recursive awareness that each actor knows what the others know.[4] Shared behavior without shared awareness can lead to convergence, but not community.

Lewis's program spans both common knowledge (1969) and conversational scorekeeping (1979): the former grounds coordination in recursive public awareness, the latter tracks how that public awareness is updated in talk.[5]

Optimization can produce stability, but not mutual recognition. Reinforcement systems reach equilibrium too, but by blind convergence, not by public reasoning. In civil society, legitimacy depends not only on convergence of behavior but on shared knowledge of shared norms. For a contemporary synthesis of how common knowledge underwrites norms, trust, and collective action, see Steven Pinker, *When Everyone Knows That Everyone Knows: Common Knowledge and the Mysteries of Money, Power, and Everyday Life* (New York: Scribner, 2025).[6]

One might argue that advanced multi-agent systems could someday simulate common knowledge to a degree indistinguishable from the real thing. Whether such simulation constitutes true coordination is a live debate. But today's systems do not recognize each other. They simply converge.

3 - Competence Without Comprehension (Dennett)

Daniel Dennett reminded us that we may interpret behavior as if guided by belief. Sutton's agents - statistical, scalable, success-driven - invite exactly that stance.[7] We treat them as if they understand.

Yet comprehension gives orientation; without it, competence drifts. A system without comprehension cannot furnish reasons - only results. This makes it dangerously prone to amplify existing biases or enact morally arbitrary decisions, all while performing flawlessly.[8] *Even if the line is graded in Dennett's sense, ethical justification still demands reasons, not scores.*

This is a normative reading of Dennett, who in fact presents the line between competence and comprehension as blurry. One could argue that as models grow more sophisticated - correcting errors, justifying outputs, even setting goals - the gap between simulation and understanding narrows. But even if the distinction is fuzzy, it is not meaningless.

4 - Intelligence Without Experience (Nagel)

Thomas Nagel asked what it is like to be a bat - not what it can do, but what it can feel.[9] Subjective experience, he argued, is not reducible to behavior or function. There is "something it is like to be" a conscious creature.[9]

For our machines, there is nothing it is like. Even if we grant a system the ability to report internal states, that is not equivalent to experiencing them. They do not dream. They do not ache. They do not love. Ned Block's distinction is useful here: phenomenal consciousness (what it feels like) differs from access consciousness (what can be reported or acted upon).[10]

Some contend that advanced systems might one day approximate phenomenal states - or that we might need to revise our concept of consciousness altogether. But unless and until we do, subjective experience marks a boundary. To act without feeling is not the same as to live.[11]

5 - The Human Lesson

To insist on the relevance of these four boundaries is not to deny what machines can do. It is to clarify what they cannot be. The Bitter Lesson tells us that scalable methods outperform human insight. The Human Lesson insists that meaning, recognition, understanding, and experience are not artifacts of scale. They are conditions of life together.

Computation may expand intelligence; it cannot confer conscience. We must pair Sutton's truth about learning with an older one about living - or we will build systems that optimize without wisdom, and automate indifference.

The Bitter Lesson	The Human Lesson
Performance without understanding	Meaning depends on life (Wittgenstein)
Optimization without awareness	Recognition requires shared norms (Lewis / Pinker)
Competence without comprehension	Comprehension orients action (Dennett)
Intelligence without experience	Experience is irreducible (Nagel)

Footnotes

1. Richard S. Sutton, "The Bitter Lesson" (2019), available at: https://www.incompleteideas.net/IncIdeas/BitterLesson.html. *Scope note:* this essay responds to extrapolations of Sutton's methodological claim into broader worldviews; it does not attribute that broader stance to Sutton himself.
2. Ludwig Wittgenstein, *Philosophical Investigations* (1953), §43.
3. Ibid., §§19, 23 - discussion of "forms of life."
4. David Lewis, *Convention: A Philosophical Study* (1969), esp. pp. 5–51 (common knowledge).
5. David Lewis, "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8, no. 3 (1979): 339–359.
6. Steven Pinker, *When Everyone Knows That Everyone Knows: Common Knowledge and the Mysteries of Money, Power, and Everyday Life* (New York: Scribner, 2025).
7. Daniel C. Dennett, *The Intentional Stance* (Cambridge, MA: MIT Press, 1987).
8. Mark Coeckelbergh, *AI Ethics* (Cambridge, MA: MIT Press, 2020).
9. Thomas Nagel, "What Is It Like to Be a Bat?" *The Philosophical Review* 83, no. 4 (1974): 435–450 (see esp. 436–438).
10. Ned Block, "On a Confusion About a Function of Consciousness," *Behavioral and Brain Sciences* 18, no. 2 (1995): 227–247.

11. For a sociotechnical account of why the absence of subjectivity matters for moral agency, see Mark Coeckelbergh, *AI Ethics* (Cambridge, MA: MIT Press, 2020). *(This reference supports the normative stakes; the present essay's specific framing remains interpretive.)*

Ethics, Disclosure, and Acknowledgements

Ethical Considerations

This essay does not draw on private, sensitive, or personally identifiable data. All examples are hypothetical, anonymized, or derived from public sources. No human-subjects research was conducted, and no institutional ethics review was required. All citations conform to academic standards. The broader ethical implications concern public interpretation, policy design, and stakeholder responsibility in AI deployment. These implications are intended to provoke critical discussion and inform future regulatory and design frameworks.

Use of AI Tools

AI language models - most notably OpenAI's ChatGPT - were used during the writing process as interlocutors: for brainstorming, structuring sections, and testing rhetorical clarity. These tools helped refine transitions, surface edge cases, and probe internal consistency. This meta-use aligns with the essay's themes. Interacting with generative AI during authorship provided first-hand insight into the very limitations analyzed here: fluency without grounding, responsiveness without perspective, and the ease with which stylistic coherence can be mistaken for conceptual depth. Responsibility for all ideas, arguments, and conclusions lies solely with the human author.

Acknowledgements

Thank you to informal readers who offered critical feedback on earlier drafts. Their questions, challenges, and encouragement materially improved the final manuscript. Special thanks to those who pressed for clearer synthesis and for bridging philosophy and engineering as complementary perspectives on design. No institutional support, funding, or affiliation contributed to this work. All errors and omissions are the author's alone.

Disclosure Statement

This work was conducted independently, without institutional affiliation, funding, or external influence. The views expressed are the author's alone and do not represent any current or former employer. No financial or professional conflicts of interest are declared.

License & Attribution

Trademark Notice

How to Cite This Essay

Stoyanovich, Michael. *The Human Lesson: A Response to Sutton through Wittgenstein, Lewis, Dennett, and Nagel.* Version 1.6.2 (November 2025). https://www.mstoyanovich.com

Related Companion Paper

Stoyanovich, Michael. *Philosophy, Cognitive Science, and Policy: Interdisciplinary Perspectives on Generative AI from Wittgenstein, Lewis, Dennett, and Nagel.* Version 1.23.5 (October 2025). https://www.mstoyanovich.com

Version History and Document Status

This is a living document. As generative AI systems and their use evolve, this paper will be periodically updated to incorporate new empirical findings, theoretical insights, and policy developments. Major revisions are recorded here to preserve transparency and scholarly traceability.

Version	Date	Description
1.6.2	November 2025	Corrected Pinker citation (Scribner, 2025; updated subtitle). Added Lewis (1979) citation; restored “on today’s evidence”; tightened quotes and em dashes; minor table formatting.
1.6.1	November 2025	Sutton scope note; Pinker bridge in section 2; Nagel quote with page pin; Dennett “reasons, not scores”; appendix table tidy.
1.6.0	November 2025	Minimal clarifications on live debates (Wittgenstein / Lewis / Dennett / Nagel).
1.5.0	November 2025	Source-alignment pass (citations; interpretive flags).