Some thoughts on teaching and learning

Written to satisfy some requirement

April 2026

Contents

1 Assessing understanding

The success of modern machine learning has challenged the seemingly obvious assumption that there is a well-defined thing called "understanding." In the 1960s and 1970s, the first wave of machine intelligence was "expert systems," which relied on hard-coded domain knowledge to solve problems in organic chemistry and medicine [2, 4]. To the extent that the human expert programmers understood their fields, so did the expert systems, as their decision-making process relied on knowledge in the form of explicit rules for decision-making. Yet, in 1997, IBM's chess program "Deep Blue" beat a chess grandmaster for the first time, with no reliance on hard-coded rules [1]. Instead of using domain knowledge, Deep Blue was trained on data from a large quantity of chess matches, where different board positions were scored based on their likelihood of success. In the 30 years since this historic defeat, the scientific community has seen over and over again that computer programs trained on data often far surpass those relying on hard-coded rules. From image classification to speech recognition and natural language processing, data-driven machine learning models based on neural networks have proven that expertise and domain knowledge are not required to automate complex scientific tasks. How do these models work? Do they genuinely *understand* the tasks they are trained to solve, or, as the now tired refrain goes, do they simply "fit data"? Is there even a difference?

The philosopher John Searle suggests that it is not necessary to understand in order to satisfactorily solve a problem. In his 1980 thought experiment known as the "Chinese Room," he imagines an individual in a sealed room trained to process incoming messages written in Chinese. This individual does not speak or understand Chinese, but has a complex set of instructions for how to write up responses to the incoming messages and send them back to the outside world [3]. Assuming that the instructions are rich enough to produce satisfactory responses—in other words, assuming that the responses pass the Turing test—the recipient has every reason to think that the author *understood* the initial message. At the level of inputs and outputs, whatever is inside the Chinese Room appears to have understanding. Yet, when we peer inside the room, we see a rote mechanical task being carried out, with none of the attendant mentation we intuitively associate with understanding.

Bringing this example back to machine learning, the Chinese Room is like a model that "fits the data" without recourse to the mechanisms underlying the input-output relationship. In the setting of engineering mechanics, this would be like a data-driven model trained to predict the maximum displacement of a structure as a function of the applied forces, without encoding Newton's laws. We know that the force-displacement relation could be predicted by writing down and solving differential equations by applying Newton's laws (conservation of momentum) to the structure. At first glance, this appears to be a very different procedure than building a data-driven machine learning model—in the first case, we use laws of physics to deduce a model of the system, in the second case, we inductively build a model by fitting data. Tempting though it is, this intuition is incorrect. The physics-based model—the one that is apparently interpretable, mechanistic, understandable, transparent, causal...whatever other term of endearment we want to heap on—is ultimately data-driven as well. These models rely on empirical parameters, like the stiffness of a material or the viscosity of a fluid, that must be calibrated from data. Yet, when we estimate these parameters from data, we think we have really learned something about a system, rather than having merely performed a fit. A differential equation model parameterized by material properties of this sort seems to "understand" the system under study, whereas a neural network building an input-output map does not. If all models are data-driven, why do some encode understanding and others do not?

Understanding understanding is a challenging task. From what I have seen, fluff abounds in the interpretable machine learning literature on the topic of understandable (or interpretable) models. Most answers seem to be that a model is understandable when it has some property, and usually that property is something synonymous with understandability. Answers of this sort are not very enlightening. When one is in the habit of fitting models to data, the difference between understandable models and black-box "just-fitting" type models is simple: restrictions on the model class. Models which can be said to understand the phenomena under study place serious restrictions on the sort of input-output behavior that can be represented. If you accidentally used the wrong data set, the model would not be able to fit it. This is not the case with the just-fitting type of model. These are so flexible that any input-output behavior can be represented. There is much more to be said about all of this—the bottom line is that real learning is possible only in the presence of constraint. Searle suggests that solving a problem is not synonymous with understanding, that understanding is a largely invisible cognitive process, rather than something measured at the level of inputs and outputs. All scientific models are data-driven, but only some encode mechanism, thus providing deeper understanding of the system of interest. The difference between mechanistic and black-box models is simply constraints on the model class—mechanistic models are inflexible compared to black-box models based on neural networks. Because of their inflexibility, when a trained mechanistic model matches the data, we think we have really learned something.

What does any of this have to do with assessing understanding in students? The Chinese Room thought experiment suggests that tests or other related assignments (especially multiple choice) are a suboptimal assessment strategy. It is perfectly possible to generate correct answers by learning to mimic a given problem solving process without the elusive cognitive process of understanding. I have seen this in my conversations with fellow scientists over the years—someone can repeat a string of words which may be the correct answer to a given question, without any felt sense of what those words mean. Like the Chinese Room, input-output relationships can be accurately captured without the mentation of understanding—data-driven mimicry is not monopolized by machines. If the goal of education is only to prepare students to solve problems, mimicry may suffice. This is not my taste, and as large language models continue to progress, this version of learning seems more useless everyday. In order to assess understanding, we first must (attempt to) say what it is. I would like a theory of understanding which partially decouples correctness from understanding. I'd bet that an early and undue emphasis on correctness encourages mimicry at the expense of building up the cognitive structures of understanding.

My understanding of understanding has everything to do with a felt sense. Understanding does not have to do with producing correct answers to questions, *it has to do with the subjective feeling that things could not have been otherwise.* I will grant that this definition of understanding is conditioned on the fact that I am an engineer. Because the conservation principles underlying engineering are familiar from first-hand experience of the world (mass does not disappear, forces cause things to move, temperature equilibrates, etc.), it is possible to build intuition of this sort. Whereas it is hard to imagine a world without conservation of mass, momentum, and energy, it is very easy to imagine a world where quantum mechanics and general relativity were different. Not because it is easy to imagine these things, but because their influence on our daily experiences are extremely indirect—we are not sensitive to these laws in the way that we are sensitive to the laws of classical physics. Because we have evolutionarily hard-wired intuition for classical physics, it is possible to build up understanding from this foundation. If one accepts that objects which are pushed accelerate, then the governing partial differential equations of solid mechanics could not have been otherwise. If one accepts that energy is neither created nor destroyed, the governing partial differential equations of heat transfer could not have been otherwise. If one accepts certain intuitively reasonable assumptions about probability, the second law of thermodynamics could not have been otherwise.

I believe that this feeling of necessity is accompanied by vivid mental imagery. This imagery is something that courses and textbooks explicitly do not teach, in spite of the fact that it is arguably foundational to scientists' practice [5]. When I am working with undergraduates, I explicitly tell them about the mental pictures I rely on to understand the material. Assessing understanding then has to do with interrogating a student's mental model of the problem at hand. One corollary of this definition is that it is possible to make progress toward understanding while getting incorrect answers. Suppose that student A memorizes a formula and relies on pattern recognition to apply it to new problems. Student B attempts to build a mental model of the phenomenon of interest in order to have a gut sense of how things work. Almost certainly, student A's approach is a more straightforward way to score high on tests, but in my eyes, it is not the kind of understanding we should pursue.

Earlier, I argued that in the context of scientific models, real learning happens only with constrained model classes. When we learn, as opposed to fit, we only have a handful of parameters to adjust in order to explain what we observe. In my opinion, the analogous process in one's education is pulling back new concepts to already familiar ones. If every new scientific concept was treated as a thing in itself, as opposed to a derivable consequence of more fundamental ideas, one's intellectual life would become very cluttered. Almost the entirety of engineering mechanics can be made sense of with calculus, linear algebra, and conservation laws. This means that new concepts only *appear* new—in reality, they follow from a small set of underlying principles. In my learning process, I only have a few parameters to adjust. When I encounter something new, I don't just pile it on to an existing set of disorganized concepts. I try to file it away in the appropriate cabinet. Keeping things organized this way is an attempt to minimize the number of ideas that one has to take as given. In my view, this is analogous to calibrating a meaningful model as opposed to "just fitting."

In summary, understanding is an internal cognitive process which relies on vivid mental imagery and an organized network of concepts. Understanding is more than simply knowing the right answer, it is a feeling that no other answer would be possible. Understanding requires that one distinguishes between derivable and axiomatic concepts, and organizes knowledge accordingly. How can this be assessed? I suspect that conversation is the best assessment strategy, and writing is the second best. Most problem sets and tests probably do not accomplish this. To assess understanding, one needs to peer inside the Chinese Room. It is very easy to do this when talking with someone. You can ask questions to interrogate the robustness of someone's mental models. You can ask for clarification and further explanation. You can inquire as to how different ideas are connected. All of this requires a real-time, evolving interaction which would be hard to simulate with a static medium like text. But, writing assignments are another way to peer through the keyhole of the Chinese Room. I am not sure why science courses neglect writing as a tool for learning and reflection. As a course instructor, I would find ways to integrate writing with the scientific curriculum. Perhaps through journals, or problem set questions that involve reflections on the material. In my experience, deficiencies in understanding are most glaring in conversation and writing, therefore these media are the best strategies to assess it.

2 Power of play

The Oxford Dictionary defines play as "[engagement] in activity for enjoyment and recreation rather than a serious or practical purpose." I like this definition, because it makes play out to be an attitude rather than a particular activity. Almost anything can be done with an attitude of playfulness, or an attitude of seriousness. It is interesting to note that animals play, and that they do not need to be taught how to do it. I have heard speculation that play is an evolutionary adaptation to learn vital social and physical skills, but, evolutionary benefits aside, it is undeniable that animals *enjoy* playing. Compared to our animal friends, we humans are often a glum lot. We have a positive genius for making everything dreary and miserable. Sports are a means to the end of college admission, education is a vehicle for career advancement, eating well is a way to optimize our health, travel is a form of social capital. To be clear, I do not at all exempt myself from this—I have managed to turn many things that I enjoy into chores and arenas to compete with other people. What a nasty and useless tendency! Young kids seem to know how to play, but as we get older, we forget at some point. And I don't think it is as simple as deciding one day: "I am going to resurrect that forgotten youthful attitude!" By this time, the habits are deeply ingrained—it takes serious work to extract oneself from these patterns. Apparently, many of us need to be taught how to play as adults.

I hear it already, an objection to bringing an attitude of playfulness into the study of science and engineering: "You mean to say that cancer researchers and flight test engineers and bridge designers ought to be *goofing around*? This is serious business!" I'll admit, my sample size is limited, but in my experience, a sense of seriousness in one's work is more a function of ego size than the work itself. The world is much too absurd to hold together the precarious idea that what one does is serious without the glue of self-deception. It is not the claim that there are serious things in the cross-hairs here, it is the *feeling* of seriousness. I am a rockclimber, and climbing big walls or adventurous routes in the mountains is, objectively, a serious thing. The consequences of a mistake are undeniably injury or death. Does this mean that all joking around, all awareness of the absurdity of what you are doing—in short, all playfulness—is inappropriate? Of course not! In climbing, mastery of the situation provides the freedom to play. When I am out of my depths, I am tense, fearful, reactive. Not only is this unpleasant, but it is not a useful strategy for handling serious situations—calm, confident, and deliberate are generally better traits. But at even higher comfort levels, there is playfulness. When I watch videos of climbers goofing around at the belay on hard routes on El Capitan (for example), I admire their mastery of that environment—the ability to play in increasingly complex and severe situations is one of the rewards of competence.

So some climbers know how to play, but they are exceptional in this regard. The world rarely encourages us to play as adults, so the skill is eventually lost. How can playfulness be incorporated into one's learning? Perhaps more pointedly, how can the skill of play be deliberately fostered in the context of a college education? I believe that an attitude of playfulness can, and should, be fostered in college, but with a caveat. Play should not come at the cost of depth of learning. It is very easy to imagine this happening. One of the emphases in the Power of Play workshop was on embodiment and movement. I am not convinced that this is appropriate for learning about differential equations, for example. Sure, we might line a bunch of students up and send a pressure wave by passing objects of different weights through the line. I'm sure you'd find that the time the object takes to traverse the line depends on its mass, just as the wave speed in a material depends on the density. For a college course, I think this is a bit childish. As future scientists and engineers, it should suffice for college students to use their imagination in certain cases. I do think that this is a useful *thought experiment*, but I do not think that it is necessary to carry out. An instructor could spend an entire class performing this experiment, or simply articulate it and allow the students to imagine, which would take all of five minutes. Part of playing is learning what kind of play is appropriate for the task at hand. Learning deeply in the context of wave propagation is learning how to do the math, playing is called for to the extent that it serves that goal.

What sort of play is appropriate for college engineering courses? I think illustrations of seemingly abstract concepts with physical objects is appropriate. For example, when teaching on the displacement and strain of a continuum body, it is instructive to have a deformable object on hand, perhaps with a coordinate system drawn on it. Discussions of vibration would benefit from an object that rings when struck. Having hands-on real systems grounds discussions, and helps students realize that mathematics is often less abstract than it seems—that systems in the real world genuinely call for this kind of analysis. Another way to foster a playful attitude is to encourage "what if?" questions. Most questions of this sort have to do with challenging assumptions. What if we added higher-order terms to Hooke's law? What if the boundary conditions were time-dependent? What if the fluid changes temperature? I think of this as the intellectual equivalent of taking things apart and putting them back together—a playful attitude toward the built stuff of the world, to be sure. Another playful approach to learning is to make connections. People seem to find connections inherently satisfying. When a comedian makes a surprising reference to earlier material, when you realize that you share a mutual friend with someone, when you learn that your favorite musician was influenced by your favorite writer—these things are just satisfying. Similarly, when you learn that mass-spring-damper systems follow the same governing equations as resistor-inductor-capacitor circuits, one gets an exciting feeling that everything in the world is connected. So explicitly emphasizing connections in the course material is another way to foster an attitude of playfulness. Perhaps projects could be designed for students to explore connections between the course and other things that they care about—whether it is past courses, their hobbies, current events, etc.

Another way to play with course material is to code things up yourself. I am aware that many students don't think of coding as particularly playful, as I was one such student in the past. But it is easier than ever to get started coding with large language models. In engineering, one simply does not get a hands-on feel for mathematics until programming a computer to solve problems. One reason this is true is that there is a lot of mathematical abstraction that computers cannot handle—functions become vectors, integrals become simple addition, ordinary differential equations become recursion relations. Generally, everything makes a lot more sense. And the beauty of coding things up yourself is that you get to try things out. You suddenly have agency to ask and answer your own questions about the course material. You get to bring the material to life with plots and animations. Assuming the task is not overly complicated, coding things from scratch is more playful and informative than using existing code. I believe that experimenting with ideas through the medium of code is one of the best forms of play for college students studying engineering. Another useful strategy I have learned is to use the online graphing software called Desmos, which can be used to quickly build animations and play with ideas.

Learning engineering should be fun, and encouraging a playful attitude toward the material is one way to make this happen. Play does not mean doing childish science experiments at the expense of hard skills, it means finding "enjoyment and recreation" in the process of cultivating those skills. A playful attitude towards ideas means seeing things from multiple perspectives, challenging assumptions, making connections, and fostering a sense of agency in one's learning by bringing to life ideas in the form of code. Just as play with friends builds relationships, playing with ideas creates a sense of care and connection to what you are learning. Play is only possible when external pressures are mitigated. If students believe that grades are the only things that matter, that the quality of their future life depends entirely on their performance on assignments, that they will be seen as unintelligent if they don't know the answer, the playful attitude will be snuffed out. In order for a playful attitude to flourish, learning must be conceptualized as a recreational activity, rather than a "serious or practical" thing. And I believe that better work of the serious and practical sort would get done if students were given the space to approach their education in this way.

3 Teaching to non-majors

In my case, teaching non-majors would probably mean teaching calculus, or perhaps a science elective like the "physics of music." As I understand it, the distinction between majors and non-majors is not expertise—presumably, the two groups have approximately the same background when entering an introductory course. Instead, the distinction has to do with whether the student plans to continue taking courses in that field. Is it obvious that there should be any difference in how you teach to these two groups? I am not so sure, but let's think through it. My first thought is that students majoring in the field require less persuasion to care about the course material. They already have a vested interest in absorbing the content of the course, knowing as they hopefully do that it will be drawn on later in their college journey. On the other hand, the non-majors are probably sitting in the back with their arms crossed, counting down the minutes of each lecture, waiting to check the box of their general education requirement. This stereotype is easy to reach for at least. I haven't taught a course of this sort, so I can't say for sure whether it holds or not. But I imagine that some non-major students are genuinely interested in the course (they elected it, after all), and many yes-major students are just going through the motions. In my mind, the question really comes down to: can the students be assumed to care about the course? In the non-major case, it is tempting to say "no" and in the major case to say "yes." After all, if the students have similar backgrounds, the way the material is presented is likely not that different. So the whole question seems to pivot on the students caring.

Intuitively, the assumption of caring frees the instructor from *persuading* the students to care. Adopting a *homo economicus* model of the yes-major students, it is pretty obviously in their self-interest to have some stake in the course. But is this a good model? And is it really such a boon to let the musculature of persuasion atrophy with the yes-majors? Is not every opportunity to speak on your chosen area of expertise an exercise in polemics, a theater in which to tout the many virtues of your field? Can anyone be assumed to care about anything, really? If the looming threat of planetary extinction isn't enough for us to cancel our beloved cross-country road trip in our gas-guzzling SUV, can we be sure that a random collection of teenagers has miraculously set aside myriad teenaged distractions, in service of a whole-hearted commitment to a career path they have more-or-less blindly chosen, and therefore also whole-heartedly committed to *your particular class*, totally free of the need for persuasion? No, that cannot be assumed. So, effectively, in spirit at least, everyone is a non-major. No one can be assumed to care, nor have any particular stake besides getting a good grade, as it probably can be safely assumed that the world's vices—in this case, a mania for metrics and outcomes—have already penetrated these impressionable teenaged heads.

The question is not how to teach to non-majors vs. yes-majors, it is how to teach students that may or may not care. At the time of writing, I have only worked with students who care. They are students who have sought out opportunities to learn beyond their coursework. These ones are easy to teach, as you can condition on the fact that they think the material is interesting, important, worthy of their time. Regarding the complement of this set of already-bought-in students, I can only speculate. Perhaps their lack of interest is a form of clear-seeing—is it self-evidently true that a teenager *should* care about calculus? I can imagine the protestations of a cross-armed student in the back, saying that this is all just wrangling a bunch of made-up symbols, with nothing to do with the real world. Can a convincing counter-argument be made to this? The only honest response I see is that, made up though the symbols are, they have *a little bit* to do with the real world, though you will probably never get to see that firsthand. The uninterested students probably have as much to teach us about our own quirks and fetishes as we have to teach them about symbol wrangling.

All of the machinery of persuasion should be deployed to convince students that the material is interesting, but with no hope of it working. If I went to a talk on snail biology, I would like the speaker to attempt to convince me that this is an interesting subject, but I reserve the right to insist that it is not. We're all entitled to our preferences. But I do think it is worthwhile to always teach like the audience is non-majors. To make connections to the outside world, to provide concrete examples that pique students' interest, to highlight the interesting and fun aspects of the field while minimizing tedium, to discuss explicitly how the field has influenced our daily lives, to make the material memorable, to broadcast your own excitement about the field, to state explicitly why you find the material worthwhile. These are some of the things that might convince the skeptical student to care, and will deepen the interests of students who are already interested. If many remain unconvinced, then at least you tried, and hopefully reminded yourself why you chose to study these topics in the first place.

References

[1] Richard Korf. Does Deep Blue use AI? *AAAI Technical Report*, 1997.

[2] Robert K. Lindsay, Bruce G. Buchanan, Edward A. Feigenbaum, and Joshua Lederberg. DENDRAL: A case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence*, 61(2):209–261, June 1993.

[3] John Searle. *Minds, Brains, and Programs*. 1980.

[4] Edward H. Shortliffe. Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 66–69, October 1977.

[5] William P. Thurston. On proof and progress in mathematics, April 1994. arXiv:math/9404236.