

Introduction: Where Does Intelligence Come From?

There is a moment, early in Jane Austen's *Emma*, when the novel's heroine reflects on her own qualities with considerable satisfaction. Emma Woodhouse is clever, she knows it, and the people around her have always agreed. She has been tutored at home and cushioned by comfort, and she has always taken her intelligence for granted. What the novel demonstrates, across nearly four hundred pages of social comedy and moral reckoning, is that she is wrong about this, not about being clever, but about what cleverness is and where it comes from.

On a scenic summer afternoon at Box Hill, Emma does something she cannot take back. Surrounded by friends and neighbors on a picnic, she makes a joke at the expense of Miss Bates, a woman older and poorer than herself, and entirely without defenses. The joke draws exactly the laughter Emma intended, and it also reveals, in a single moment, the precise shape of what she does not yet know how to do. Mr. Knightley takes her aside afterward and tells her, without raising his voice, that Miss Bates deserved her compassion, not her wit, and that she should have known better. On the carriage ride home, Emma cannot stop the tears running down her face. She will think about Box Hill for a long time, and it will change her.

What changes, by the end of the novel, is not her quickness. Emma is as quick and perceptive on the last page as on the first. What changes is the hard-won knowledge of her own limits, arrived at through consequence rather than instruction. The process is incremental and dependent on experience that corrects: moments where existing understanding proves insufficient and something more has to form in its place. Whether Austen understood it in these terms matters less than what the novel reveals about it.

Something similar happened in a very different setting. In 2017, a program called AlphaZero was given the rules of chess and nothing else: no record of human games, no accumulated theory, no guidance about what good play looked like. It began by making moves essentially at random, losing constantly, receiving a single signal in return: win, lose, or draw. Over nine hours of training, playing millions of games against itself, it developed a style of chess that leading grandmasters, including Garry Kasparov, described as unlike anything they had encountered before, sacrificing material for positional advantages that human players had never learned to value, because no human player had ever learned chess the way AlphaZero did. Its intelligence was, in the most literal sense, the product of its specific experience. The prescription is to change the experience to change what forms.

The marquee example is Emma Woodhouse on a carriage ride home, unable to stop crying. The machine learning analog is a chess program playing millions of games in nine hours, developing preferences that were not programmed into it. These two images have no obvious connection, and making one is the aim of this book. The argument made here is that intelligence is not a property that can be installed in a system, but rather a process that builds through experience. The necessity of experience is as true of a person as it is of a machine. What

determines the quality of mind that develops is not the underlying material, but how the material has been processed through experience: the encounters that surprised it, the errors that cost it something in real terms, and the feedback from the world that helped to shape it..

For most of human history, one of the strongest answers to the question of where intelligence comes from pointed inward, toward whatever a system was made of materially. Whether it was the soul, the mind, the brain, or the genome, as each era had its preferred substance, this was the favored location for the mechanism that gave creatures their level of capability. A change in the substrate (the underlying material), suggested a change in what might be possible. A rock could not think, but a brain could. The material was the explanation.

However, the evidence of recent decades has made it increasingly difficult to sustain the assumption that substrate is destiny. The same essential process of learning, it turns out, can occur in very different physical systems, and what matters in both cases is not the material but the quality of the experience that has shaped it. That observation does not settle every question about minds and machines, and this book does not pretend that it does. But it shifts the weight of the argument away from what a system is made of and toward what it has been through.

This argument is developed through two bodies of evidence that at first blush may not seem to belong together: the fiction of Jane Austen and Joseph Conrad, and the research programs of contemporary machine learning. The novel, it turns out, is an unusually good way to understand what the research describes through its rich development of the inner lives of characters over time.

Austen's topic, across her work, is the formation of judgment: the long, unsteady process by which a person, in the *Emma* example, develops from several overly-confident misreadings of situations towards, instead, someone much closer to heartfelt understanding. This parallels what machine learning researchers describe when they talk about how intelligence forms: not through sudden insight, but through accumulated correction over time. Conrad works in the same territory of character development, under somewhat opposite conditions. There is a moment in Conrad's *Heart of Darkness* when Marlow, a seasoned sailor of twenty years, arrives in the Congo and discovers that everything his experience has taught him is suddenly useless. The moral frameworks he brought with him are useless, and the man he has traveled so far to find, Kurtz, turns out to be a case study in what happens to an intelligent mind when everything that normally keeps it intact is taken away.

Whereas Austen's *Emma* has time, security, and the corrective friction of honest relationships, Conrad's Kurtz ostensibly lacks these, and what forms in their absence is not wisdom but its wreckage. Conrad's fiction returns to this contrast in other works too, for example, in *The Secret Sharer* and *Under Western Eyes* as much as in *Heart of Darkness*. The difference between Austen and Conrad's examples is that of intelligences that develop through gradual encounters with a world that resists them; and intelligences that are forced, compressed, or corrupted by environments that do not provide constructive feedback.

The contrast between the two writers maps directly onto a distinction the book develops between formation under conditions of time and reflection on one hand, and formation under pressure and truncation on the other. Both produce problem-solving agents of intelligence, but of different kinds, with different limits and different costs. The distinction turns out to have a precise parallel in machine learning, where researchers have found that the conditions under which a system learns matter as much as the architecture on which it runs.

For the early part of the machine learning field's history, artificial systems were built on the assumption that intelligence could be specified in advance. If a system was provided with the right rules, the appropriate architecture, and the requisite amount of data drawn from human experience, it would perform intelligently. However, decades of effort have revealed that this assumption has a ceiling. Systems built on human-generated data can reproduce the kinds of knowledge that humans have already acquired, but they struggle to exceed it, and they cannot learn in the way a mind learns, through consequence, through the pressure of a world that pushes back, through the knowledge that comes from acting rather than observing. The argument now gaining ground in the field, though not without contest, is that genuinely capable systems will have to learn from experience in a richer sense: interacting with environments that resist them, accumulating the kind of knowledge that only contact with a variable world can produce.