

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/145636>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

What Kind of Learning Is Machine Learning?

Tyler Reigeluth, Insititute of Philosophy, Université de Grenoble-Alpes

Michael Castelle, Centre for Interdisciplinary Methodologies, University of Warwick

To appear in: Roberge, J., & Castelle, M. (2020). *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Palgrave.

Abstract

While social theories of human learning have proliferated in the last century, machine learning is a rather less reflexive enterprise. What conception of learning do the techniques of machine learning—especially in its recent connectionist forms—imply or induce? To answer this question, we explore the perspectives of the Soviet cultural-historical psychologist Lev Vygotsky, contrasting his socially-grounded understandings of mediated concept learning and the “zone of proximal development” with the methodologies of supervised and unsupervised machine learning. Such a comparison highlights the dependence of machine learning on microgenesis (repetitive, behaviorist training processes) and phylogenesis (the architectural “evolution” of models) at the expense of ontogenesis (the lifelong, interactional development of an individual in society), and thus provides new insights into the fundamental limits of contemporary artificial intelligence.

Introduction

At the outset of his 1803 lectures on pedagogy, Kant (1803/2007) famously stated that “The human being is the only creature that must be educated.” He explains that human nature is indeterminate and needs to be given care, discipline, and instruction; i.e., humans can only achieve their full potential by developing a *second* nature. We have since grown accustomed to framing this indeterminacy of human development in terms of debates around “nurture versus nature,” a dichotomous trope which Enlightenment philosophy can be seen as foreshadowing (Keller, 2008). Kant did not, of course, foresee that humans would one day build artificial entities that could *themselves* learn—but if he had, he might have argued that they would also need to be educated in a similar manner to humans. While the idea that learning machines require an “education” seems strange from the perspective of contemporary machine learning, such a notion is in fact present near artificial intelligence’s inception. Specifically, in the 1948 paper “Intelligent Machinery”—which would remain unpublished until the late 1960s—Alan Turing (1948/1969) acknowledges that learning by machines would,

just as in the case of learning by humans, necessarily involve the *social* roles of teachers, peers, and the larger community:

Although we have abandoned the plan to make a ‘whole man’, we should be wise to sometimes compare the circumstances of our machine with those of a man. It would be quite unfair to expect a machine straight from the factory to compete on equal terms with a university graduate. The graduate has had contact with human beings for twenty years or more. This contact has been modifying his behaviour pattern throughout that period. His teachers have been intentionally trying to modify it. At the end of the period a large number of standard routines will have been superimposed on the original pattern of his brain. These routines will be known to the community as a whole. He is then in a position to try out new combinations of these routines, to make slight variations on them, and to apply them in new ways.

What Turing understood (i.e., the fundamentally social and situated nature of learning), contemporary approaches to machine learning (including the twenty-first-century revival of neural networks known as *deep learning*) have all but forgotten or neglected. Our claim in this essay is that we need a *social theory of machine learning*: i.e., a theory that accounts for the interactive underpinnings and dynamics of artificial “learning” processes, even those thought to be “unsupervised.” But in doing so, however, we must be careful not to resort to the term “social” as a self-explanatory term that is simply opposed to the “technical” and does not require further elaboration, for to do so would only reiterate a longstanding—and philosophically problematic, as pointed out by Simondon (1958)—antagonism between the realms of culture and technique. In other words, our attempt to sketch a social theory of machine learning should not be interpreted as simply another demonstration that the models learned by algorithms bear the mark of the humans who design and train them (Diakopoulos, 2014); nor are we simply advancing another claim that opening the algorithm’s “black box” will reveal the social norms embedded within its operations (Pasquale, 2015).

What we are interested in instead is how machine learning behaviors unfold and develop meaning and purpose as a socially recognizable form of activity (i.e., as Turing says above with respect to human-based education, “these routines will be known to the community as a whole”). That is, we want to approach machine learning *with reference to, and in comparison with, our understandings of human learning in society*. But to do so we must acknowledge that machine learning algorithms are not merely executors or implementors of prior or external social norms or knowledge; instead, their activity reshapes collective activity as much as it is shaped by it (Grosman & Reigeluth, 2019).

Developing such a social theory of machine learning is necessarily a critical enterprise—not in the sense of attacking claims of machine “learning” from the outside, but of establishing the conditions upon which claims of “machine learning” could be convincingly held. In order to do this, we will compare and contrast contemporary machine learning in practice with one of the most distinctive sociotechnical approaches to pedagogy and learning, namely, what has come to be called *activity theory*. This multidisciplinary

perspective builds from the transformational works of the Soviet cultural-historical psychologists Vygotsky, Leont'ev, and Luria. Activity theory differs from American behaviorist and cognitivist theories of learning in that it does not depend on a dualism between behavior and mind, between individual action and sociocultural practice, or between action and interaction.¹ We suggest that activity theory—aspects of which have also been taken up and expanded by other contemporary theories such as situated learning (Lave & Wenger, 1991)—can provide an essential, albeit general, framework for developing a social theory of machine learning.²

For the purposes of this essay, however, we will mostly be drawing on Vygotsky's programmatic contributions around the processes of *concept development* and analogizing them to the classification practices of machine learning models. Vygotsky's (1987) central idea is that concepts are not abstract content that is acquired, copied, or directly incorporated by an individual; instead, concepts develop through learning and through a process of *generalization* (p. 47)—a term which, as we will discuss below, is also used extensively in machine learning. For the learner, a concept has value and meaning by representing a solution to a *lived* problem, which may further enable the learner to solve *future* problems. Vygotsky thus offers a dialectical account of conceptualization in which learners and concepts *develop together*. The activity of learning drives the individual's development forward, and both the individual and social meaning of concepts transform through that same learning activity. In other words, concept learning is a *genetic* (i.e., developmental and processual) social activity in which individuals take part in meaning-making and, in doing so, transform both themselves and society.³

While concept development and generalization are certainly not the only learning processes addressed by Vygotsky's work, they do offer an ideal problem space to explore machine learning, insofar as the latter—in the form of object recognition models in computer vision, for example—is overwhelmingly presented in terms of being able to *generalize* the detection of higher-level *concepts* such as “giraffe” or “butterfly” in previously unseen images (Chollet, 2018). Vygotsky's approach, as well as those sociocultural approaches to learning that grew out of it, can help renew our understanding of what it means to learn in ways other than the “optimization”-oriented approach that is generally used by machine learning practitioners. Our hope is that by applying socially oriented theories of learning to the field of machine learning, we can help provoke deeper discussions about what exactly we should expect from machine “learning.” For example, one central (yet

¹ On the dualism between behavior and mind see Vygotsky (1997a, p. 65); on the dualism between the individual and the social see Cole and Wertsch (1996); for a discussion of the key assumptions of activity theory, see Chaiklin (2019).

² Our reasons for using activity theory rather than other dominant educational theories are twofold. First, more conventional theories of learning from the late twentieth century—which consider learning as a hierarchy of individual behaviorist and/or information-processing-like achievements ranging from simple associations to logical problem solving (Gagné, 1970)—tend to reduce or eliminate the social, semiotic, contextual, and technological qualities of any human learning situation. Second, alternative approaches, which vary from constructivism (Piaget, 1954) to constructionism (Papert, 1991) to situated learning (Lave & Wenger, 1991), frequently incorporate one or more of the dialectical insights of activity theory.

³ Vygotsky's term *genetic* is unrelated to the branch of biology concerned with DNA or RNA nucleotides, but instead refers to the way his experiments focused on uncovering the developmental phases of a learning process (Ageyev, 2003).

largely implicit) normative premise that polarizes debates around machine learning is the idea that, ideally, machines would ultimately automate the process of learning because, after all, automation is what machines do best. This in turn implies that “full” machine learning will only be achieved when the machine learns “by itself” (in what is called *unsupervised learning*). Where does this fetish for self-sufficiency come from? Why *should* we expect that from machine learning techniques, however “deep” they may prove to be, when such a fundamental asociality is considered a contradiction in the context of human learning?⁴

A brief history of human theories of learning in machine learning and artificial intelligence

The recent enthusiasm on the part of researchers and the popular media for artificial intelligence represents the result of the increased relevance in recent decades of the field and methodology of *machine learning* (ML) and, specifically, *deep learning* (DL). While not all contemporary machine learning techniques are “deep” and/or connectionist in nature, it is the case that the particular methodologies of ML/DL—in which a model is trained and progressively evaluated on subsets of a large corpus of data, whether manually labeled or not—induces an experimental approach that diverges from previous AI regimes in the 1970s and early 1980s (Langley, 1988). Specifically, the importance of the term “learning” and its metaphors was at a low point in that previous era of cognitivism and “good old-fashioned AI” (GOF AI), in part due to well-known critiques—such as the attacks by Chomsky (1959) on Skinner and by Minsky and Papert (1969) on the Perceptron, which rejected behaviorist and connectionist approaches, respectively—that set the stage for an intellectual environment in which a framework largely (if not exclusively) devoted to rule-based symbol processing could become “the only game in town” (Sejnowski, 2018, p. 250).

The early-1980s *Handbook of Artificial Intelligence* describes how, after Minsky and Papert’s 1969 devastation of (single-layer) connectionism, “those . . . who continued to work with adaptive systems ceased to consider themselves AI researchers . . . [A]daptive-systems techniques are presently applied to problems in pattern recognition and control theory” (Cohen & Feigenbaum, 1982, p. 326). Instead, by the 1970s, according to the authors, AI researchers mainly “adopted the view that learning is a complex and difficult process and that, consequently, a learning system cannot be expected to learn high-level concepts by starting without any knowledge at all” (p. 326). This period, then, was characterized by a belief that tabula rasa-style systems that learned by example were infeasible; and indeed, the idea that the acquisition of new knowledge required an existing base of knowledge was a good fit with Chomsky’s “innate grammar” view of language. This was

⁴ While self-learning in humans is of course possible in some situations (Gibbons et al., 1980; Rousseau, 1762), few learning theorists would advocate for the elimination of teachers entirely. Vygotsky argued that instructional experience is internalized, and thus “when the school child solves a problem at home on the basis of a model that he has been shown in class, he continues to act in collaboration, though at the moment the teacher is not standing near him . . . It is a solution accomplished with the teacher's help. *This help—this aspect of collaboration—is invisibly present. It is contained in what looks from the outside like the child's independent solution of the problem*” (Vygotsky, 1987, p. 216; emphasis added).

combined with a specific ideology, highlighted by the historian Cohen-Cole (2005), in which cognitive science and AI came to be unconsciously and reflexively based on a conception of intelligence inspired by the intellectual pursuits of cognitive scientists and AI researchers themselves (e.g., highly trained in the physical and mathematical sciences, good at chess, etc.).⁵ It was in part because the goals for AI were so lofty—and specifically constrained to a domain of symbolic-centric rationality—that starting from scratch seemed intuitively implausible (and we see this again today in the demands for “artificial general intelligence”).⁶

With cognitive science and AI practitioners in the United States considering themselves part of a revolution against the previous behaviorist tradition in psychology, which was in part characterized by a dependence on experiments on animals, it should not be surprising that there is little reference to theories of animal learning in those fields. There were, of course, some in cognitive science and AI who made explicit references to *human* theories of learning in their work, such as Seymour Papert, who studied under the Swiss psychologist and learning theorist Jean Piaget—best known in the United States as a proponent of a discrete, staged development of child intelligence—before joining MIT’s AI Lab (Boden, 1978; Papert et al., 1963; Piaget, 1970). However, the cognitive science community remained largely indifferent or hostile to Piaget, a situation illustrated by a fizzled 1975 debate between Piaget and Chomsky in France (Piattelli-Palmarini, 1980).⁷ Even Papert (1980), in his work on the programming language Logo, described his main influence from Piaget as “a model of children as builders of their own intellectual structures” and his view of “Piagetian learning” as being equivalent to “learning without being taught” (p. 7)—emphasizing the AI field’s intrinsically asocial, individualist, and innatist perspective (Ames, 2018). References to Vygotsky were also, unsurprisingly, rare in the cognitive science and AI literature, in part perhaps because of an early, disparaging misreading of Vygotsky by the cognitivist philosopher Jerry Fodor, one of the more extremist representatives of the symbol-processing ideology of intelligence who would later go on to attack connectionism in the late 1980s (Fodor, 1972; Leontiev & Luria, 1972).

R.S. Michalski, a Polish-American émigré at the University of Illinois, thus described this 1970s milieu as a time in which learning was “a ‘bad’ area to do research in” (Boden, 2006, p. 1047). However, by 1983, Michalski and others (1983) introduced a collection of papers that would be seen as the first significant textbook on machine learning, though Michalski and his coeditors largely hewed to the hegemonic line

⁵ “As cognitive scientists like George Miller and Herbert Simon crossed back and forth between scientific descriptions of the human and normative discussions of the best way for scientists to think, they borrowed from the folk and social psychological image of right thinking to inform their own personal and public images . . . These very same scientific self-images would form the basis for the image of human nature that cognitive science produced” (Cohen-Cole, 2014).

⁶ By contrast, DL architectures do begin in a tabula rasa manner (while the neural architecture is fixed, the parameter weights are set randomly), and their goals are arguably more limited, to what would have then be dismissively known as a kind of pattern recognition (e.g., the classification of a bitmapped image as containing a tiger).

⁷ This debate, staged at the Abbaye de Royaumont outside of Paris, was later described as one in which “Chomsky’s argument focused exclusively on complex details of the learning of syntax, about which Piaget had virtually nothing to say . . . [and] Piaget’s ground for argument was conceptual learning, about which Chomsky had nothing to say” (Jackendoff, 1995).

regarding knowledge acquisition being a largely symbol-centric activity, requiring a base of preexisting symbolic knowledge. The nascent machine learning field instead distinguished itself by emphasizing not just deductive processes but those of analogy and induction, with the latter including under its umbrella the “learning by example” paradigm maligned by the GOFAI community (Carbonell, Michalski, & Mitchell, 1983).⁸ Regardless, an awareness of developmental and/or socially oriented perspectives on learning is rather lacking in this era of machine learning literature, where the term “teacher” merely indicates the presence of labeled examples (Michalski, 1983)—the scenario now commonly known as supervised learning. This literature also frequently refers to the wholly individualistic concept of “unsupervised” learning or learning without a “teacher” or outside of society—a concept which, if applied to human learning, a theorist like Vygotsky would find very strange.

The lack of reflection on the relevance of existing theories of human learning to machine learning did not significantly improve with the reintroduction of connectionist methods in the mid-1980s (exemplified by the “Parallel Distributed Processing” or PDP volumes of (Rumelhart & McClelland, 1986)). However, the PDP authors were fully aware of the paradigm-shifting quality of their connectionist conception of learning:

In recent years, there has been quite a lot of interest in learning in cognitive science . . . All of this work shares the assumption that the goal of learning is to formulate explicit rules (propositions, productions, etc.) which capture powerful generalizations in a succinct way . . . The approach that we take in developing PDP models is completely different . . . [W]e do not assume that the goal of learning is the formulation of explicit rules. Rather, we assume it is the acquisition of connection strengths which allow a network of simple units to act as though it knew the rules. (McClelland et al., 1986)

While this revival of connectionism was accompanied by some genuine technical improvements—e.g., the unsupervised learning techniques of the fully-connected “Boltzmann machines” (Ackley et al., 1985) and a working backpropagation algorithm for real-valued hidden layers (Rumelhart et al., 1986)—any hint of a connection between this revised approach to neural learning and developmental theories of learning remained absent for a few years.⁹ This began to change in the early 1990s with a succession of texts coming from San Diego and the south of England, beginning with the cognitive scientist Jeffrey Elman at the University of California, San Diego, who popularized what we now know as the “simple” version of a *recurrent neural network* (RNN) in a paper suggestively titled “Finding Structure in Time” (Elman, 1990). Unlike the (single-layer or multi-layer) perceptron, in which input data is successively transformed by matrix multiplications and sigmoid functions, the RNN works with sequence data (such as sequences of words or phonemes), and each

⁸ For internalist depictions of the history of machine learning of this era, see (Carbonell et al., 1983b) as well as (Kodratoff, 1992).

⁹ It is remarkable to read one early text attempting to bridge PDP and cognitive psychology (Kehoe, 1988) describing how “cognitive research and theory has focused largely on the already highly productive performance by highly experienced human subjects.” On PDP and classical conditioning see also (Klopf, 1988).

input is accompanied by the previous output, allowing a “hidden state” vector of values to evolve over time. The recognition of this “developmental” quality of RNN learning led to collaborations between Elman and the developmental psychologist Elizabeth Bates (Bates & Elman, 1993); a work bridging developmental child psychology with connectionism from a former Piaget student, Annette Karmiloff-Smith (1992), who had visited UCSD and worked with Bates and Elman; and an argument for an “epigenetic developmental interpretation” of “connectionist modelling of human cognitive processes” (Plunkett & Sinha, 1992). These authors would all come together for the volume *Rethinking Innateness* (Elman et al., 1996), which appears to mark the end of this brief developmental enlightenment in neural network research. This interesting intersection of connectionism and developmental psychology seemed to wane with the (second) decline of connectionism, induced in part by the rise of support vector machines (Cortes & Vapnik, 1995) and decision tree ensembles (Breiman, 2001)—also forms of supervised learning, but ones which operated primarily on tabular data, as opposed to the sequence-like data that had inspired Elman. It would then be another two decades before the multilayered neural network would make its most recent resurgence (Cardon et al., 2018).¹⁰

We contend, however, that the current deep learning “revolution” has already begun to inspire a new intellectual shift comparable to the impact and hegemony of the cognitivists, who were able to popularize an individualistic information-processing metaphor of the mind that ultimately influenced educational practice itself—as in the case of the educational psychologist Robert Gagné’s (1977) approach to cognitive processes (Ertmer, Driscoll, & Wager, 2003), as well as Papert’s constructionism approach (inspired by a partially asocial reading of Piaget’s constructivism). As such, we believe that it will once again become increasingly common to attempt to understand human learning with reference to the new generation of success stories of twenty-first-century AI, perhaps ultimately in the form of radical shifts in educational policy.¹¹ But at the same time—and quite unlike the early 1990s—we are at a moment in which the sociotechnical embeddedness of these machine learning and deep learning systems, radically increased in scale and scope if not in their basic architectural underpinnings, are making their limitations increasingly salient to a wide population—and the audience for understanding the nature and source of those limitations is increasing. The stakes are now arguably higher for understanding what kind of “learning” machine learning—and deep learning specifically—really represents.

In the next three sections, we will describe three significant aspects of Vygotsky’s “cultural-historical” psychology that will, later on, help us address the implicit theory of “learning” in machine learning. First, we

¹⁰ In the intervening years, a field known as *epigenetic robotics* emerged that integrated developmental perspectives with robotics, which even inspired occasional reference to learning theorists like Piaget and Vygotsky (Berthouze & Ziemke, 2003). Because our focus here is on supervised learning and not reinforcement learning or interactive robotics, we do not discuss this work in detail, but it will undoubtedly be relevant to future histories of “social” reinforcement learning.

¹¹ This inevitable analogical reframing of deep learning toward human learning is a traditional “universalizing” quality of AI, identified by Goldstein and Papert (1977), who explained that “it may seem paradoxical that researchers in the field [of AI] have anything to say about the structure of human language or related issues in education. . . . [But] [a]lthough there is much practical good that can come of more intelligent machines, the fundamental theoretical goal of the discipline is understanding intelligent processes *independent* of their particular physical realization” (p. 85).

will discuss the intrinsically developmental and specifically *sociogenetic* qualities of learning in the work of Vygotsky and his followers; then, we will show how Vygotsky conceived of *concept development* in his psychological experiments with children, which will be useful because contemporary machine learning classifiers excel in recognizing objects in images or in learning vectorial “meanings” of words; and we will discuss how Vygotsky’s *zone of proximal development* highlights that one must always understand “learning” as it is instituted in specific educational systems with specific forms of instruction.

What is “social” about learning?

Unlike in machine learning—where, for instance, the notion of “unsupervised” learning implies that it is possible to learn without a teacher—the study of human learning (sometimes referred to as *pedagogy*) explicitly acknowledges the necessity of some kind of teaching agent. However, there has for some decades been a debate in pedagogical thought that can be characterized, broadly speaking, as about whether the best learning model is the “guide by the side” or the “sage on the stage” (Ferster, 2014). The former tends to involve a naturalistic or spontaneist stance whereby children learn best when “left to their own devices.” Proponents of this “child-centric” approach claim that learning is optimal and also most enjoyable when unhampered by the limitations and constraints that rigid and homogenizing education systems impose through methods, techniques, and normative expectations; a child’s best teacher is experience—a view prefigured by writers such as Rousseau (1762), Pestalozzi (1827), and Montessori (1912). Education in this case is about accompanying and guiding natural learning processes in a nonintrusive manner, namely by adapting content and methods to the learner’s developmental stage. The work of Piaget (1954) represents one of the clearest expressions of this position in which education is about making the world available to what the child can do at a given stage of its natural development.

Conversely, the “sage on the stage” perspective reflects what could be called an institutional stance, according to which the central site of learning is, and needs to be, a formal educational structure where individuals develop their cognitive abilities alongside their moral character (Hegel, 1808–1811/1984; Kant, 1803/2007). In this sense, and given human nature’s inherent indeterminacy, education is seen as instituting a certain *kind* of individual, as “molding,” “shaping,” or “structuring” individuals through their learning. This more “content-centric” approach tends to see education as the process through which culture is reproduced; and when viewed critically, as in Bourdieu and Passeron (1977/1990), additionally as a mechanism for the reproduction of social stratification in general. Although the uncritical version of this perspective is generally considered outdated or conservative when compared to the “child-centric” approach—and is disparaged as “instructionism” by Papert (1993) and others—this perspective helps us foreground the fact that education is always, at least in part, about bringing into being a specific kind of individual living within a given culture, and, that education *is possible*; i.e., as Kant suggested, there is something in human development which is indeterminate and thus requires a process of education.

But as with most binary oppositions, this presentation is a caricature. Few thinkers or researchers in education theory or pedagogy would unilaterally call for one of these two approaches to pedagogy, and most would recognize that children need curriculum-based knowledge *to learn* as well as more open-ended situations in which their creativity and problem-solving skills are actively stimulated. But simply taking the middle road between two extremes does not lead us to a robust social theory of learning. Vygotsky's cultural-historical psychology—in which the term “historical” refers to a focus on *development* at various interrelated biological, individual, and social levels—offers a serious basis for thinking of alternatives to this mired debate through focusing on the role of *mediation* in the construction of psychological and behavioral functions and processes (Engeström, 1999, p. 28).

Vygotsky developed a research program—cut short by his premature death—that his successors Leont'ev and Luria would carry on and expand (van der Veer & Valsiner, 1991). According to Vygotsky, behavioral and cognitive development through learning is not a process of socialization by which the individual *becomes* social (as is the case for Piaget); instead, the individual is itself the site of a process through which society is transformed (Brossard, 2012). This genetic and relational view of behavioral and cognitive development is Vygotsky's way of expressing the Marxist doctrine of historical materialism, which tells us that we are not so much thrown into “the World” (as phenomenologists would have it) as we are thrown into determinate forms of social activity. As the Vygotsky commentator Roy Pea (1985) neatly summarizes, “our productive activities change the world, thereby changing the ways in which the world can change us.”

In this view, development is not simply about successive cognitive functions replacing or evicting one another as the individual accommodates to the pressures of communication and social life, but about transformations in the *organization* of cognitive functions as they relate to one another (Vygotsky, 1987, p. 175). Cognitive functions do not transcend the specificities of social activity, but instead modulate and reorganize themselves accordingly (Prot, 2012, pp. 308–309). These transformations, in turn, bring about new forms of cognitive activity. For instance, Vygotsky is not interested in concepts in and of themselves, but in how their meaning develops within social activities, on the one hand, and how the individual's cognitive activity is transformed through the use of certain concepts, on the other.

This brings us to a fundamental aspect of Vygotsky's sociogenetic theory of psychological development: all higher psychological concepts (such as, for example, the meanings of words) were at some point concrete social relations. As Roth and Jornet (2017) explain, drawing on Vygotsky's 1934 book *Thinking and Speech*:

In [Vygotsky's] approach, there is a primacy of the social. Whatever higher psychological function can be identified first *was* a social relation with another person. Vygotsky did not suggest that something (e.g. meaning, rule) initially existed *in* a social relation, something that the participants may have constructed together to be internalized by individuals. Instead, the higher psychological function is identical to that earlier social relation. (p. 106)

To illustrate this general idea, Vygotsky (1979) shows how a toddler's spontaneous movements and gestures become the basis of a social activity in which meaning is progressively constructed through the mediating presence of an adult, which in turn paves the way for semiotic and linguistic mediations. Following Vygotsky's account, the act of pointing is first an attempt to grab or reach for an object. Or rather, what matters is that this gesture *is interpreted* as such by an adult who responds by moving the object closer to the child or helping the child reach the object.¹² The child's spontaneous gesture becomes an indicative one that now appropriates the adult's mediating presence, which will ultimately serve as a basis for the child being able to say "I want." The toddler (putatively) reaching for an object is doing something more than failing to reach it. With the help of an adult, one could say that it is reaching a new developmental stage by performing actions it cannot yet do alone but some day will.

Correspondingly, when a child learns the word "apple," it is not learning some fixed descriptor of a given object in reality, it is using the word as an instrument within a social activity. The word uttered by the child might mean "I want an apple" or "I'm hungry" or "I want to hold that red thing and throw it"; it takes the place of an entire chain of gestures and actions that have become relatively stabilized through interactions with others. Before ever becoming free-floating signifiers, words congeal and condense concrete social relations as they are used in activities. The density of social relation, the obstacles encountered in activities and the ways in which words, as instruments, help direct action and regulate behavior, are the basis upon which concepts are developed. We can begin to see that the way a child learns the meaning of "apple" may be quite significantly different, for example, from the way a convolutional neural network "learns" to detect the presence of apples in bitmap images, or the way a word embedding model learns the vectorial "meaning" of the word "apple" in high-dimensional space.

For Vygotsky, the development of "higher psychological functions" is essentially a *mediated* process, insofar as these functions involve *signs* as a constitutive dimension of their activity.¹³ Vygotsky's understanding of the term "sign" is broad and not strictly limited to a Saussurean perspective where the sign incorporates a preestablished or fixed relationship between a signifier and a meaning (as in Saussure [1915]). In addition to all the culturally relevant mediations (i.e., language, maps, arithmetic, writing, plans, mnemotechnics, etc.) involved in social activities, for Vygotsky any object can potentially become a sign, or a

¹² The semiotic novelty of Vygotsky's account of word meaning, in which indexical signs are seen as a foundation for the development of conceptual meaning, has been noted by Wertsch (1985): "The indicatory or indexical function of speech makes it possible for adults to draw young children into social interaction because it allows intersubjectivity on perceptually present objects even though complete agreement on the interpretation or categorization of these objects is lacking" (p. 57).

Silverstein's (1985) characterization is even more complete: "Vygotsky's account [on the development of concepts] is really a logical reconstruction of the passage of words from being indexicals connected to the things they ultimately will truly denote, through an 'egocentric' stage in which they serve as performatives of a sort, to their ultimate emergence as sense-valued elements of propositional schemata, each stage being a functional enrichment, not a replacement, of the cognitive utility of language" (p. 231).

¹³ While some interpreters, following in the line of Leont'ev, hold that signs such as speech are just one type of "tool" that includes external tools and techniques, Vygotsky focused on speech as the most fundamental form of mediation (Miller, 2011, pp. 20–24; Vygotsky, 1987, pp. 60–61).

“psychological instrument,” if it is integrated into action as a means of controlling one’s behavior and planning one’s actions (Friedrich, 2012, p. 261; Vygotsky, 1930/1997c).¹⁴

Vygotsky provides a striking illustration of this thesis in his critique of the Piagetian conception of the child’s internal or “egocentric” language. Whereas Piaget sees egocentric language as a transition between pre-social (“autistic”) language and “socialized” language, Vygotsky does not take the beginning of the development process to be pre-social. For Vygotsky, the language a child develops when they seem to be “speaking to themselves,” using words that are directed at no one in particular, as though gratuitously accompanying an activity (e.g., coloring, building with blocks), in fact corresponds to the process of reorganizing the activity. Talking over the activity is not then an idle accompaniment, aimed simply at telling oneself what one is doing, but a way of making sense of the forces or objects that resist the child’s activity, a way of coping with the world’s *will*, so to speak. Vygotsky (1987) gives the example of a child he studied who was drawing a tram and pressed too hard on the paper with his pencil, declared “broken,” and proceeded to take up a paintbrush to depict a broken tram car while continuing to talk to himself about the new scenario (p. 70). As Yves Clot (1997) comments, words are the instruments children use to think through the obstacles they encounter, but these instruments are not simply lying around waiting to be used; instead, their appropriation transforms the very activity within which they are taken up, thereby actively broadening the meaning they are supposed to “have”.

Conceptualization as problem solving and meaning transformation

Vygotsky focuses on concept *development*, rather than concept *acquisition*, as one of the central forms of mediation through which meaning emerges. In this sense, framing learning as a social activity does not merely imply that the learner progressively *acquires* socially constructed signifiers, but that a concept’s meaning actually develops through the social relations that underpin the learning process. As Vygotsky (1987) describes, “in the problem of interest to us, the problem of concept formation, [the] sign is the word. The word functions as the means for the formation of the concept. Later, it becomes its symbol” (p. 126). The meaning of a signifier is not presupposed nor is it intrinsically attached to a word; rather, it is the result of a dialectical process through which meaning develops socially.

Crucially, the correspondence between a word and the concept it enfolds is not learned once and for all but is itself developed through learning. A child can learn the word “tree” without having mastered “tree” as a concept. Vygotsky (1987) explains that “in itself, learning words and their connections with objects does not lead to the formation of concepts. The subject must be faced with a task that can only be resolved through the formation of concepts” (p. 124). Vygotsky is explicitly taking aim at the associationist belief that concepts somehow grow out of repeated associations between an object and a word, that particular traits are gradually superimposed to form a general concept or category that includes all the particular traits or qualities. Learning

¹⁴ See also Davydov and Radzikhovskii (1986) for a discussion of signs as psychological tools.

a concept, Vygotsky insists, is not about thickening the associations or increasing the number of connections. Instead, it implies a qualitatively different form of cognitive activity that is not reducible to quantitative reinforcement of associative connections (pp. 123–124). It isn't that the child using the word "tree" has a half-baked or poor understanding of what a tree is and that when she reaches the developmental stage in which abstract concepts are mastered (i.e., adolescence) she will have a complete understanding of "tree." For Vygotsky, a concept's meaning is not self-contained but intimately connected to the activity within which it is used. A concept comes to life through learning as the subject is confronted with problems it is incapable, in its given developmental state, of resolving by itself.

Vygotsky's insistence on the social nature of word meaning can be read as analogous to Marx's critique of the fetishization of exchange-value over use-value, in which, e.g., a coat is valued as an abstract monetary quantity as opposed to an expression of the social nature of its use and the labor involved in its production and distribution. Specifically, Vygotsky (1987) is attacking contemporaneous approaches in psychology and linguistics for the way they reified meaning by matching a word to its signifier in the same way that Marx undermined classical economic theories of value by exposing the inherently social relationship economic value translates (pp. 162–163, 169–173). What Vygotsky is trying to get at is the *genesis* of meaning and concepts, as well as the activity this genesis entails. Traditional experimental setups in child psychology proceeded by isolating words from their sentences and contexts, thereby depriving the subject of its ability to effectively think through its activity. For example, as Vygotsky describes, "the experimenter selects an isolated word and the child must define it. This definition of the isolated word taken in a congealed form tells us nothing of the concept in action. It tells us nothing of how the child operates with the concept in the real-life process of solving a problem, of how he uses it when some real-life need for it arises" (p. 123).

By contrast, for Vygotsky, every word, insofar as it bears meaning, enfolds a generalization which cannot be understood as a preestablished relationship between a signifier and its meaning but must be studied as an "act of thought": "The word does not relate to a single object, but to an entire group or class of objects. Therefore, every word is a concealed generalization. From a psychological perspective, word meaning is first and foremost a generalization. It is not difficult to see that generalization is a verbal act of thought; its reflection of reality differs radically from that of immediate sensation or perception" (Vygotsky, 1987, p. 47). One can contrast this notion of generalization to the simpler notion of generalization in machine learning, which merely indicates that large numbers of association tasks (e.g., labeling cats vs. dogs in a corpus of images) can lead to greater performance on unseen images, regardless of whether or not the model truly has a coherent high-level "concept" of a cat or dog. A convolutional neural network can be exceptional at distinguishing between bitmap images of cats and dogs, but it could tell you little about the relationship between cats and dogs in Western society.

Vygotsky would thus likely be dubious of this notion of generalization in machine learning; for him, a concept develops because it can help a learner solve a problem, and the learner must experience the *need* for a concept as it informs its activity. As he takes care in emphasizing, however, spontaneously solving problems

as we happen to run into them is not enough for a concept to truly develop. Concepts emerge and come to matter for the learner because he or she was presented with tasks or problems to solve within certain situations. As Vygotsky (1987) explains:

Where the environment does not create the appropriate tasks, advance new demands, or stimulate the development of intellect through new goals, the adolescent's thinking does not develop all the potentials inherent in it. It may not attain the highest forms of intellect or it may attain them only after extreme delays. Therefore, it would be a mistake to ignore or fail to recognize the significance of the life-task as a factor that nourishes and directs intellectual development in the transitional age. However, it would also be a mistake to view this aspect of causal-dynamic development as the basic mechanism of the problem of concept development or as the key to this problem. (p. 132)

Learning as instituted in specific educational systems: The Zone of Proximal Development (ZPD)

This leads us to a pivotal point of our argument: *concepts are learned, but they also need to be taught*. And this is particularly true, as Vygotsky points out, of *non-spontaneous* or *scientific* concepts, i.e., concepts that relate to systematic forms of knowledge (e.g., “gravity” as it forms a system of concepts in physics with “mass” and “force”) and specific problem-solving skills (e.g., solving a mathematical equation). Unlike *spontaneous* concepts which develop through the child’s daily activities and encounters, one cannot reasonably expect scientific concepts to develop spontaneously; they require certain social conditions and forms of activity typically found in schools or other educational institutions in which learners are confronted with specific types of problems and in which their motivation to develop and use concepts is actively stimulated and nurtured.

Concepts that are developed out of daily experience—such as money, to use a particularly notorious example—are often difficult to define (Brossard, 2012, p. 102; Simmel, 1907/2004). This does not imply that scientific concepts are completely distinct or disconnected from spontaneous ones. On the contrary, scientific concepts *need* spontaneous concepts even though they synthesize their contradictions on a different cognitive level (Brossard, 2012, p. 103). This entails that education is not so much about inculcating content as it is about mobilizing students’ previous experiences (inside and outside of school) into knowledge and skills they can use in new situations. Although critical of the spontaneist maxim that children learn best when left to their own devices and pedagogical intervention should be used parsimoniously, Vygotsky (1987) does recognize a kernel of incontrovertible wisdom in Tolstoy’s belief that “consciously transferring new concepts or word forms to the pupil is as futile as attempting to teach the child to walk through instruction in the laws of equilibrium” (p. 171).¹⁵ He explains:

¹⁵ The passage Vygotsky cites is from Tolstoy (1967, p. 278).

The development of concepts or word meanings presupposes the development of a whole series of functions. It presupposes the development of voluntary attention, logical memory, abstraction, comparison, and differentiation. These complex mental processes cannot simply be learned. From a theoretical perspective, then, there is little doubt concerning the inadequacy of the view that the concept is taken by the child in completed form and learned like a mental habit. The inadequacy of this view is equally apparent in connection with practice. No less than experimental research, pedagogical experience demonstrates that direct instruction in concepts is impossible. **It is pedagogically fruitless. The teacher who attempts to use this approach achieves nothing but a mindless learning of words, an empty verbalism that simulates or imitates the presence of concepts in the child.** Under these conditions, the child learns not the concept but the word, and this word is taken over by the child through memory rather than thought. Such knowledge turns out to be inadequate in any meaningful application. (p. 170)

Vygotsky explicitly takes aim at what he considers to be an empirical and theoretical error: we cannot expect to get at learning either through personalization—i.e., simply matching “knowledge” with the child's current state of development—or standardization, i.e., setting up a rigid learning plan that is supposed to correspond to the students' developmental stages. In either case, the error is to think tautologically of learning as what can be learned alone given an individual's developmental level. Ultimately, education should be about situating learning within social activities that demand more of the learner than what the learner can perform by themselves.

This leads us to one of Vygotsky's (1987) famous concepts, the *Zone of Proximal Development (ZPD)*, an idea that critiques the concept that a child inhabits a particular stage of development based on their present abilities in isolation:

If the gardener decides only to evaluate the matured or harvested fruits of the apple tree, he cannot determine the state of his orchard. Maturing trees must also be taken into consideration. The psychologist must [also] not limit his analysis to functions that have matured. He must consider those that are in the process of maturing. If he is to fully evaluate the state of the child's development, the psychologist must consider not only the actual level of development but *the zone of proximal development*. (pp. 208–209)

For Vygotsky, instruction must aim to be just ahead of the child's potential zone of development, keeping in mind that this potential includes not just what the child can do independently but also what the child can do with others. In this process, *imitation* is taken to be essential to learning and not as its perversion or inauthentic expression. “Instruction is possible,” he says, “only when there is a potential for imitation” (p. 211); as we will see below, this idea that an interactional form of imitation is potentially productive and not simply “bad” learning finds parallels in the recent ML technique of generative adversarial networks.

In this perspective, learning does not *coincide* with development, then, but rather *pulls it forward*. Children learn how to do more than they can do independently because of the social mediations embedded in activity, which entails that learning can only occur where the child is able to imitate what lies just beyond what it is capable of doing on its own.¹⁶ The ZPD is thus the site for those concrete social relations which for Vygotsky, as described above, become internalized as words, concepts, or other cognitive mediations which are subsequently deployed in new social situations.

This view of the intrinsically social quality of the learning process is seemingly at odds with the way institutionalized education tends to evaluate “true” or authentic learning by testing the agent’s ability to solve a problem on its own—be that without the help of peers, teachers, or even technical mediations—as well as the way models in machine learning are evaluated on their predictive performance in isolation. On a broader level, this points to the deep interrelation between the way AI in general and ML in particular mirror certain normative or even political stakes involved in defining what learning *should* be by producing a model of what we *think* learning is. Proponents of Vygotsky-influenced educational psychology instead argue that determining the learner’s proximal rather than actual developmental stage allows for a differentiated approach to education whereby a learner is confronted with problems that require it to move beyond what it can do by itself, but which are accompanied by the support of social and technical mediations (Moll, 1990). In addition, higher order “scientific” concepts, which can only be learned within formal educational settings, should not all be taught exactly the same way to every student. Making scientific concepts relevant involves having a grasp of each learner’s everyday or “spontaneous” concepts that they use to make sense of the world, which the newly acquired scientific concepts will in turn help organize in new ways. Vygotsky’s account of the ZPD allows for a form of “personalization” of content and rhythm that does not atomize the learner and recognizes that the very things it is learning are social by nature. When it comes to machine learning, this perspective points to the trouble the field has in evaluating model performance with abstract metrics in what would otherwise be highly social tasks such as translation and image generation, and it implies the need for a radical shift in our way of evaluating what should count as learning and what kind of activity it involves.

Is machine learning a social form of learning?

We have seen how Vygotsky’s cultural-historical psychology provides an original understanding of what is social about learning; we have seen how he applies this to concept development in particular; and we have learned about the zone of proximal development. Now we can begin to ask: is machine learning—and deep learning in particular, which has over the last decade promised repeatedly to lead to a new paradigm of artificially intelligent agents—actually a kind of *learning* in Vygotsky’s sense?

We can consider a now-prototypical case of contemporary (connectionist) machine learning, that of the

¹⁶ A similar conclusion has been reached, although with different social and cognitive implications, by advocates of the extended mind theory (Clark & Chalmers, 1998; Hutchins, 1994; Wheeler, 2011).

multilayer convolutional neural network classifier model in the lineage of those developed by Yann LeCun in the 1980s on the MNIST dataset or by Geoff Hinton's students in the early 2010s using Fei-Fei Li's ImageNet dataset.¹⁷ The models that are trained in these scenarios are said to have distinct *architectures*, a kind of morphology in which each "layer" consists of one or more matrices (if two-dimensional) or tensors (if three-dimensional or higher) (with corresponding bias vectors). These architectures generally do not *change* throughout the "lifespan" of the model and thus DL models can be said to have a highly restricted form of "development" that psychologists would call *microgenetic*: namely, the potential to modify, during the training process, their weight parameters, today sometimes numbering in the hundreds of millions or billions.¹⁸ (Like the field of architecture for buildings/physical structures, neural architectures also "evolve" in a *phylogenetic* way, but this is highly dependent on the actions of their [human] architects and the technical environment in which they both reside.¹⁹) As with all forms of supervised learning, these models update their weight parameters through the use of a simple loss function (e.g., categorical cross-entropy) and optimizer (e.g., stochastic gradient descent) on labeled training data; i.e., each example image is accompanied by a categorical label indicating that it represents a picture of a butterfly and not a picture of a tiger or 998 other objects.

So what kind of learning *is* supervised learning? Clearly it does not easily correspond to the "sage on the stage" or "guide by the side" caricatures mentioned above—we find no lecturer or dynamic accompaniment, only labels and the mechanical instructions of stochastic gradient descent. Supervised learning is instead what early machine learning researchers called "learning by example"—which is ideally not a rote memorization but a learning process most closely akin to the stimulus-response animal experiments of behaviorism. Specifically, supervised machine learning is more similar to the classical conditioning of Pavlov than the operant conditioning of Thorndike and Skinner, in which the animal subject has the freedom to engage in a variety of actions; the latter corresponds more closely with reinforcement learning.²⁰ But if we appreciate supervised learning (as well as reinforcement learning) as essentially behaviorist—despite its complex ideological and technical underpinnings in a cognitivism that previously revolted against behaviorism as well as (in the case of deep learning) a connectionism that previously revolted against aspects of said cognitivism (Baars, 1986; Bechtel & Abrahamsen, 2002)—then we can ask: well, what did Vygotsky think about

¹⁷ We describe connectionist ML over the "classic" tabular-oriented ML of Breiman (2001) so that our critiques have greater purchase on claims for twenty-first-century AI, although in principle our critiques could apply to the simpler architectures of decision trees or SVMs (the latter of which, as pointed out by LeCun [2008], can be seen as a particular genre of two-layer neural network or "glorified template match[ing]").

¹⁸ Microgenesis is "based on the assumption that activity patterns, percepts, thoughts, are not merely products but processes that, whether they take seconds, or hours, or days, unfold in terms of developmental sequence" (Werner, 1957, pp. 142–143).

¹⁹ See Grosman and Reigeluth (2019) on how the "lineages" of DL architectures impose certain constraints on human invention.

²⁰ Rescorla and Wagner's (1972) mathematical formalization of Pavlovian classical conditioning can be shown to be similar to simple supervised neural network models that typically update their weights based on a learning rate hyperparameter and a least-mean-squares loss function (Sutton & Barto, 1981). (Today, this similarity of machine learning to behaviorism is only really appreciated in the reinforcement learning community, where some of the language is in a direct lineage.)

behaviorism?

Indeed, the conference talk that kicked off Vygotsky's career in psychology in the mid-1920s was a critique of the Soviet genre of behaviorism known as *reflexology* specifically associated with Pavlov and the neurologist Vladimir Bekhterev—a field whose proponents believed that “all human conduct could be conceived as combinations of conditional reflexes” but which “deemed the scientific study of [subjective experience] impossible” (van der Veer & Valsiner, 1991, p. 41).²¹ Vygotsky saw this approach—which Bekhterev explicitly intended to supplant all other forms of psychology—as doomed to fail, especially if the ultimate goal involved an understanding of human consciousness. This was ironic, Vygotsky explained, because these researchers often had spoken interactions with their human subjects in practice, and speech for Vygotsky indeed represented a form of “reflexes” connected to the subject's internal experience; but the reflexologists simply did not consider those interactions a valid form of behavioral data. (At the time, Vygotsky [1997a] instead believed that one might be able to understand higher psychological processes and even consciousness through these “reflexes of reflexes” [p. 79].)

This inattention to speech, in turn, meant that behaviorism—locked in as it was to the “stimulus-response” or “S—R” framework—was blind to the importance of *mediation*, which was a fundamental feature of Vygotsky's later studies of child development. In this transitional stage of Vygotsky's work, he and L. S. Sakharov performed experiments to determine how children learned to form concepts, given a set of differently colored and differently shaped objects along various dimensions that had been assigned nonsense names like “dek” (small and tall objects) or “mup” (large and tall objects) by the experimenters (van der Veer & Valsiner, 1991, p. 261). They argued that children specifically navigate from nonsensical “syncretic” groupings to arrangements into what are called “complexes”—in which groupings are based on objective features, but these features may be irrelevant to adults—and eventually to what Vygotsky then called “true concepts,” although this latter stage could only be reached by adolescents and adults.²² Vygotsky's (1987) insight is that younger children thinking in *complexes*—specifically the highest form of complex, called the “pseudoconcept”—and adults thinking in *concepts* can nevertheless “establish mutual understanding and verbal interaction” (p. 145), despite the child only being able to conceive of, e.g., a word's meaning as a “bundle of attributes or features” (Blunden, 2012, p. 231).²³ Vygotsky (1997c) referred to the transitions to complexes and then concepts as a process of *generalization* that went beyond pattern-matching and to the level

²¹ “Pavlov was certainly interested in behavior (a term that acquired almost as many varied meanings as the word ‘objective’), but he was not a behaviorist. Unlike John Watson and other American behaviorists of his day, he consistently acknowledged the existence and paramount importance of subjective phenomena—of the internal emotional and intellectual experiences of humans and other animals—and he always believed that science should seek to explain them” (Todes, 2014, p. 295).

²² It is important to recognize that Vygotsky's discussion of syncretism, complexes, and true concepts in the fifth chapter of *Thinking and Speech* (1987) was originally written around 1930, but his sixth chapter on spontaneous and scientific concepts was written in 1934, the year of his death (van der Veer & Valsiner, 1991, p. 257).

²³ Blunden points out that the distinction between pseudoconcepts and true concepts is related to the influence from Hegel's *Logic*: “Dialectical logic is in fact nothing more than the art of dealing with concepts, that is, *true* concepts, rather than simplified, impoverished pseudoconcepts” (Blunden, 2012, p. 253).

of a scientific concept seen in a relational network with other concepts:

I will give an example. Let us compare the direct image of a nine, for example, the figures on playing cards, and the number 9. The group of nine on playing cards is richer and more concrete than our concept "9," but the concept "9" involves a number of judgments which are not in the nine on the playing card; "9" is not divisible by even numbers, is divisible by 3, is 3^2 , and the square root of 81; we connect "9" with the series of whole numbers, etc. Hence it is clear that psychologically speaking the process of concept formation resides in the discovery of the connections of the given object with a number of others, in finding the real whole. That is why a mature concept involves the whole totality of its relations, its place in the world, so to speak . . . the concept is not a collective photograph. It does not develop by rubbing out individual traits of the object. It is the knowledge of the object in its relations, in its connections. (p. 100)

We can contrast this distinction between complexes and concepts in Vygotsky with the way that deep convolutional models train on MNIST data, which involves correctly labeling scanned images of handwritten digits. Just as in Vygotsky's examples, such a model can reliably emit the label '9' when confronted with an image of a handwritten number nine, but it does not understand the *concept* of 9 as described above. This highlights the difference between *generalization* in Vygotsky and *generalization* in machine learning. For supervised machine learners, "generalization" merely means high accuracy in the model's ability to label previously unseen input data. For Vygotsky, such a model has perhaps learned a pseudoconcept of the ten digits, but it has not learned to generalize to the *true concept* of a digit.²⁴ From his perspective, even the supposedly "super-human" object classifiers of today do not have these scientific concepts—those acquired through explicit instruction guided by a teacher, like the term "square root." (One can test this proposition today by asking a state-of-the-art generative text model like GPT-2 [OpenAI, 2019] to complete the phrase "The square root of 4 is," a task at which it is typically unsuccessful.)

What are the greater implications of Vygotsky's critique of behaviorism and his understanding of concept development on our understanding of machine learning? We can go back to our earlier characterization of machine learning as essentially the study and practice of *microgenesis*: a process of training that (in the case of contemporary AI), depending on the task, might take anywhere from an hour to a couple of months.²⁵ And it is also, to some extent, a practice of *phylogenesis* in which new architectures are designed and "evolved" by human researchers. But what Vygotsky's work shows is that machine learning is *not* about, in general, *ontogenesis*, i.e., the long-term development of a conscious individual, acting in the world, and "trained" by their social surroundings in both informal and formal ways. And machine learning research is also

²⁴ Such an objection recalls the recent critiques of Gary Marcus: "Deep learning doesn't recognize a dog as an animal composed of parts like a head, a tail, and four legs, or even what an animal is, let alone what a head is, and how the concept of head varies across frogs, dogs, and people, different in details yet bearing a common relation to bodies" (Marcus & Davis, 2019).

²⁵ For a discussion of microgenesis in the context of Vygotsky, see Wertsch and Stone (1978).

only tangentially concerned with *sociogenesis*, meaning the development of social groups and identities—although in many cases it is clearly implicated in such development, as in the influence of YouTube’s recommendation algorithm on political preferences (Ribeiro et al., 2020). We summarize this comparison in Table 1.

	Time scale (approx.)	Human Development	Machine Learning
<i>microgenesis</i>	minutes to months	a given pedagogical lesson or psychological experiment	training an ML model with a given loss function
<i>ontogenesis/ sociogenesis</i>	days to decades (up to ~100 years)	the natural and sociocultural development of the individual	<i>n/a</i> (but see transfer learning and generative adversarial networks)
<i>phylogenesis/ ethnogenesis</i>	up to hundreds to thousands/millions of years	biological and cultural evolution of the human species	technical and cultural development of new ML architectures

Table 1. The role of different developmental processes in a human development framework and in a machine learning “model development” approach.

In our descriptions in the previous sections, it is clear that it is the complexities of microgenesis and ontogenesis—specifically the ontogenesis of the higher psychological functions and of true and/or scientific concepts—with which Vygotsky is most concerned. It is also relatively clear that, to a large degree, **the natural and sociocultural development of the individual is not the kind of learning with which machine learning is concerned.** This indicates that as long as the individual learning subject in machine learning is taken to be the ML model/architecture, someone like Vygotsky would not recognize it as a form of learning due to its lack of emphasis on the ontogenesis of a conscious individual, who develops over time in sequences of dialogical processes, each of which permit transitions from *intermental* to *intramental* functions (Vygotsky, 1997b, p. 106).

The technical exceptions to this ontogenetic lack in ML/DL, as noted in the table, are quite interesting in that our framing manages to surface the most intriguing and even underrated features and/or innovations of contemporary deep learning techniques: the ability to retrain or *fine-tune* certain types of models on new corpora of data that differ in some way from (and are often much smaller than) the original “pre-training” datasets, which is known as *transfer learning* (Pan & Yang, 2010; Pratt & Jennings, 1996), as well as distinctively interactional training architectures like those of *generative adversarial networks* (GANs). The transfer learning case demonstrates a limited way in contemporary machine learning in which trained models can have an ontogenetic “lifespan” beyond an individual training session by permitting the reuse of parameters

learned over longer periods of training. In the case of GANs, a distinct discriminator model, D , and a distinct generator model, G , are initialized randomly, but D is gradually trained in a supervised manner to recognize, e.g., “real” images of human faces, and G repeatedly attempts to “fool” D with generated face images by learning from D ’s assessment of its generated images’ plausibility (Castelle, 2020; Goodfellow et al., 2014). This training process differs from traditional supervised learning in that the discriminator model acts as a dynamic “instructor” of sorts, who (in the ideal GAN training process) is just slightly ahead of the student in its ability to determine the validity of the generated face images and can thus pull or guide the generator model forward. This can, arguably, represent the use of a kind of *zone of proximal development* within the otherwise sociogenetically degenerate world of machine learning techniques.

However, what if the subject of machine learning is *not* simply the model? If we instead perform an act of reconfiguration (Suchman, 2007) and situate the subject of machine learning as a *human-machine activity*—i.e., including the ML researcher, their models, and their associated tools and organizational resources—then we have a situation which permits closer analogies to Vygotsky’s cultural-historical approach. This is the approach taken by Adrian Mackenzie (2017) when he uses the phrase *machine learners* to refer “to both humans and machines or human-machine relations.” In this latter perspective, which observes the intrinsic technicity of human action and becoming—and thereby of ontogenesis and phylogenesis—we would instead argue that a supervised model that finds objects from the space of ImageNet categories permits the coexisting “user” of that model to ontogenetically improve their development of concepts. That is, as a socialized, technical human with an awareness (if not an expertise) in bird species, I can leverage deep learning classifiers to accelerate my own learning and ability to take action in new situations (such as birdwatching). In other situations, such leveraging of ML classification may be seen as hazardous; consider the court judge who, as a machine learner using an algorithm for pretrial risk assessment (Barabas et al., 2020), effectively “learns” a new scientific concept of “pretrial risk” and, by deploying their preexisting agency and authority, leverages it to take further action against specific groups. Vygotsky (1930/1997c) would consider such devices yet another psychological tool that individuals can use to master one’s mental processes; in this case “machine learning” would refer to the development of these psychological tools that extend the ontogenetic development of the machine learner. In order for machine learning to truly embrace theories of learning, such a reconfiguration may indeed be necessary.

Conclusion: Who is learning in machine learning?

The aim thus far has been neither to provide a comprehensive introduction to Vygotskian psychology nor a full critique of machine learning, but merely to emphasize three essential starting points for undertaking a social theory of machine learning. **The first essential point is the idea that learning occurs when a problem is encountered and resolved by the learner, thereby transforming the problem space itself.** To resolve a problem, a child uses instruments (in activity theory, semiotic and/or technical mediations) that are not simply

extraneous to its activity but constitute and transform that activity on a much deeper level. When it comes to machine learning, this gives a novel perspective on what it means to evaluate an algorithm's learning on an already-given problem space. From a Vygotskian point of view there is little sense in evaluating learning performances on ready-made problems without taking into account the activity within which the problem is encountered and resolved. Developmentally speaking, there is little sense in unleashing algorithms on abstract image classification or word substitution problems and hoping that their results will lead to a "generalization of generalizations" necessary for higher-level thinking (Vygotsky, 1987, p. 229). While some deep learning models are often held as examples of successful generalization, it is still the case that reinforcement learning models with superhuman performance on a majority of Atari games (Mnih et al., 2015) are still wholly limited to the world of Atari games and would be useless outside that domain. From a Vygotskian perspective, the difficulty most machine learning techniques have in generalizing models beyond their given problem space is linked to the fact that practitioners tend to consider their problem spaces as equivalent and comparable when in fact they should be approached as qualitatively different learning experiences directed by the individual's development. The excitement around "transfer learning" reflects an implicit recognition that AI, in its current form, is ontogenetically weak: its models can only "fine tune" and—unlike the child learner over the course of years of instructional microgenetic interactions—cannot undergo repeated reformations of past generalizations.

The second essential point is that learning is not something done on one's own, but something that occurs in the space between what one can do on one's own and what one can do with the help of another. In this sense, learning is always a differential experience and implies being presented with certain problems to resolve rather than others and not simply running into problems randomly. Again, Turing's quote can serve as a valuable guide when he reminds us that the college graduate has spent twenty years or so having his or her behavior intentionally modified by teachers. One of the roles of education as an institution is to determine which problems are worth solving, i.e., which problems will have a value for the learner once they are solved, which will be truly transformative in shaping their cognitive development. From this standpoint the ideal or "optimal" learning process is not one where content is made available for a learner's given developmental stage, but one where the learner reaches beyond their developmental stage through a series of sociotechnical mediations that are not mere crutches but make up the very stuff of learning. This, however, implies that we change our ethical and epistemological expectations by not only demanding optimization and automation of ready-made problems (i.e., evaluating based solely on objective functions), but by engaging in concrete learning processes as an educator would, that is an educator who is not simply a test administrator. One can see hints of this today in generative adversarial networks and even more in the experimental genres of developmental robotics that take ontogenesis and sociogenesis more seriously (Oudeyer & Kaplan, 2006).

The third essential point is that the words, concepts, or images that are learned are never fixed relationships between a signifier and signified but are in fact social relations that have been abstracted from their original activity. Crucially, this means that learning is not merely about acquiring abstract content, knowledge, or symbols and applying them to concrete situations, but it is about turning a concrete social

relation into abstract technical and semiotic mediation that can be used in other situations to regulate thinking and behavior. In other words, this is where we see learning as an *activity* in and of itself and not only something that merely happens within a given social context; this is where we see learning transforming the world, to paraphrase one of Vygotsky's primary inspirations. We argue, then, that in order to formulate a learning theory of machine learning, it may be necessary to move from seeing an inert model as the machine learner to seeing the human researcher or developer—*along with*, and not separate from, his or her model and surrounding social relations—as the machine learner. Only then can we see the sociotechnical process of empirically observable machine learning in the light of cultural-historical psychology and/or activity theory. *These* machine learners are developing ontogenetically, whereas the model in isolation is largely developing only microgenetically (although see our ontogenetic caveats in Table 1 above). They are taking the social relations from which their data has been constructed (e.g., the massive human labor of the Mechanical Turk workers who labeled ImageNet), internalizing the labeled pseudoconcepts into their models, and incorporating those internalizations into their own ontogenetic development (whether by fine-tuning the existing models or by moving on to training more interesting architectures). On public and private forums, they educate each other in their techniques at levels appropriate to their development; through the various communicative activities of releasing research papers, codebases, and pre-trained models, they pull each other's development forward. Machine learning as a *technique* is a hermetic process of training individual models—but machine learning as a *cultural activity* is a vast and social process of training other machine learners.

References

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann Machines. *Cognitive Science*, 9(1), 147–169. [https://doi.org/10.1016/S0364-0213\(85\)80012-4](https://doi.org/10.1016/S0364-0213(85)80012-4)
- Ageyev, V. S. (2003, September). Vygotsky in the mirror of cultural interpretations. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 432–450). Cambridge University Press. <https://doi.org/10.1017/CBO9780511840975.022>
- Ames, M. G. (2018). *Hackers, computers, and cooperation: A critical history of Logo and constructionist learning*. Association for Computing Machinery. <https://doi.org/10.1145/3274287>
- Baars, B. J. (1986). *The cognitive revolution in psychology*. Guilford Publications.
- Barabas, C., Doyle, C., Rubinovitz, J., & Dinakar, K. (2020). Studying up: Reorienting the study of algorithmic fairness around issues of power. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 167–176. <https://doi.org/10.1145/3351095.3372859>
- Bates, E. A., & Elman, J. L. (1993). Connectionism and the study of change. In *Brain development and cognition: A reader* (pp. 623–642). Blackwell Publishers.
- Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks* (2nd ed.). Blackwell Publishing.
- Berthouze, L., & Ziemke, T. (2003). Epigenetic robotics—Modelling cognitive development in robotic systems. *Connection Science*, 15(4), 147–150. <https://doi.org/10.1080/09540090310001658063>
- Blunden, A. (2012). *Concepts: A critical approach*. Brill.
- Boden, M. A. (1978). Artificial intelligence and Piagetian theory. *Synthese*, 38(3), 389–414.
- Boden, M. A. (2006). *Mind as machine: A history of cognitive science*. Oxford University Press.
- Bourdieu, P., & Passeron, J. (1990). *Reproduction in education, society and culture* (2nd ed.). SAGE Publications Ltd.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231. <https://doi.org/10.1214/ss/1009213726>
- Brossard, M. (2012). Le développement comme transformation par appropriation des œuvres de la culture. In Y. Clot (Ed.), *Vygotski maintenant* (pp. 95–116). La Dispute.
- Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983a). An overview of machine learning. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 3–23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-12405-5_1

- Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983b). Machine learning: A historical and methodological analysis. *AI Magazine*, 4(3), 69–69. <https://doi.org/10.1609/aimag.v4i3.406>
- Cardon, D., Cointet, J.-P., & Mazières, A. (2018). Neurons spike back: The invention of inductive machines and the artificial intelligence controversy. *Reéseaux*, 5(211), 173–220.
- Castelle, M. (2020). The social lives of generative adversarial networks. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 413. <https://doi.org/10.1145/3351095.3373156>
- Chaiklin, S. (2019). The meaning and origin of the activity concept in Soviet psychology—With primary focus on A. N. Leontiev’s approach. *Theory & Psychology*, 29(1), 3–26. <https://doi.org/10.1177/0959354319828208>
- Chollet, F. (2018). *Deep learning with Python*. Manning Publications.
- Chomsky, N. (1959). A review of B. F. Skinner’s “Verbal Behavior.” *Language*, 35, 26–58.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1111/1467-8284.00096>
- Clot, Y. (1997). Avant-propos. In L. S. Vygotski, *Pensée et langage*. La Dispute: SNEDIT.
- Cohen, P. R., & Feigenbaum, E. A. (1982). *The handbook of artificial intelligence* (Vol. 3). HeurisTech Press.
- Cohen-Cole, J. (2005). The reflexivity of cognitive science: The scientist as model of human nature. *History of the Human Sciences*, 18(4), 107–139. <https://doi.org/10.1177/0952695105058473>
- Cohen-Cole, J. (2014). *The open mind: Cold War politics and the sciences of human nature*. University of Chicago Press.
- Cole, M., & Wertsch, J. V. (1996). Beyond the individual-social antinomy in discussions of Piaget and Vygotsky. *Human Development*, 39(5), 250–256. <https://doi.org/10.1159/000278475>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1023/A:1022627411411>
- Davydov, V. V., & Radzikhovskii, L. A. (1986). Vygotsky’s theory and the activity-oriented approach in psychology. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 35–65). CUP Archive.
- Diakopoulos, N. (2014). *Algorithmic accountability reporting: On the investigation of black Boxes*. Academic Commons. <https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2>
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press.
- Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Miettinen, & R.-L. Punamäki (Eds.), *Perspectives on activity theory*. Cambridge University Press.
- Ertmer, P. A., Driscoll, M. P., & Wager, W. W. (2003). The legacy of Robert Mills Gagné. In *Educational psychology: A century of contributions* (pp. 303–330). Lawrence Erlbaum Associates Publishers.

- Ferster, B. (2014). *Teaching machines: Learning from the intersection of education and technology*. Johns Hopkins University Press.
- Fodor, J. (1972). Some reflections on L.S. Vygotsky's thought and language. *Cognition*, *1*(1), 83–95. [https://doi.org/10.1016/0010-0277\(72\)90046-7](https://doi.org/10.1016/0010-0277(72)90046-7)
- Friedrich, J. (2012). L'idée des instruments médiatisants. Un dialogue fictif entre Bühler et Vygotski. In Y. Clot (Ed.), *Vygotski maintenant* (pp. 255–270). La Dispute.
- Gagné, R. M. (1970). *The conditions of learning* (2nd ed.). Holt, Rinehart and Winston.
- Gagne, R. M. (1977). *Conditions of learning* (3rd revised ed.). Thomson Learning.
- Gibbons, M., Bailey, A., Comeau, P., Schmuck, J., Seymour, S., & Wallace, D. (1980). Toward a theory of self-directed learning: A study of experts without formal training. *Journal of Humanistic Psychology*, *20*(2), 41–56. <https://doi.org/10.1177/002216788002000205>
- Goldstein, I., & Papert, S. (1977). Artificial intelligence, language, and the study of knowledge. *Cognitive Science*, *1*(1), 84–123. https://doi.org/10.1207/s15516709cog0101_5
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *ArXiv:1406.2661 [Cs, Stat]*. <http://arxiv.org/abs/1406.2661>
- Grosman, J., & Reigeluth, T. (2019). Perspectives on algorithmic normativities: Engineers, objects, activities. *Big Data & Society*, *6*(2), 2053951719858742. <https://doi.org/10.1177/2053951719858742>
- Hegel, G. W. F. (1984). *Hegel, the letters* (C. Butler & C. Seiler, Trans.). Indiana University Press.
- Hutchins, Edwin. (1994). *Cognition in the wild*. MIT Press.
- Jackendoff, R. S. (1995). *Languages of the mind: Essays on mental representation*. Bradford Book.
- Kant, I. (2007). *Anthropology, history, and education*. Cambridge University Press.
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.
- Kehoe, E. J. (1988). A layered network model of associative learning: Learning to learn and configuration. *Psychological Review*, *95*(4), 411–433.
- Keller, E. F. (2008). Organisms, machines, and thunderstorms: A history of self-organization, part one. *Historical Studies in the Natural Sciences*, *38*(1), 45–75. <https://doi.org/10.1525/hsns.2008.38.1.45>
- Klopf, H. A. (1988). A neuronal model of classical conditioning. *Psychobiology*, *16*(2), 85–125. <https://doi.org/10.3758/BF03333113>
- Kodratoff, Y. (1992). Ten years of advances in machine learning. In P. Dewilde & J. Vandewalle (Eds.), *Computer systems and software engineering: State-of-the-art* (pp. 231–261). Springer US. https://doi.org/10.1007/978-1-4615-3506-5_9
- Langley, P. (1988). Machine learning as an experimental science. *Machine Learning*, *3*, 5–8.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

- LeCun, Y. (2008, April 9). *Visual perception with deep learning*.
<http://web.archive.org/web/20131211154343/https://cs.nyu.edu/~yann/talks/lecun-20080409-google.pdf>
- Leontiev, A. N., & Luria, A. R. (1972). Some notes concerning Dr. Fodor's "reflections on L.S. Vygotsky's thought and language." *Cognition*, *1*(2), 311–316. [https://doi.org/10.1016/0010-0277\(72\)90024-8](https://doi.org/10.1016/0010-0277(72)90024-8)
- Mackenzie, A. (2017). *Machine learners: Archaeology of a data practice*. MIT Press.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Ballantine Books.
- McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1* (pp. 3–44). MIT Press.
<http://dl.acm.org/citation.cfm?id=104279.104284>
- Michalski, R. S. (1983). A theory and methodology of inductive learning. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 83–129). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-12405-5_1
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (1983). *Machine learning: An artificial intelligence approach*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-12405-5_1
- Miller, R. (2011). *Vygotsky in perspective*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511736582>
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Moll, L. C. (1990). Introduction. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 1–27). Cambridge University Press.
- Montessori, M. (1912). *The Montessori method: Scientific pedagogy as applied to child education in "the Children's Houses" with additions and revisions*. Stokes.
<http://archive.org/details/montessorimethod00montuoft>
- OpenAI. (2019, February 14). *Better language models and their implications*. <https://openai.com/blog/better-language-models/>
- Oudeyer, P.-Y., & Kaplan, F. (2006). Discovering communication. *Connection Science*, *18*(2), 189–206.
<https://doi.org/10.1080/09540090600768567>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.

- Papert, S. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 1–12). Ablex Publishing.
- Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. Basic Books.
- Papert, S., Apostel, L., Grize, J. B., Papert, S., & Piaget, J. (1963). Etude comparée de l'intelligence chez l'enfant et chez le robot. In *La filiation des structures* (Vol. 1–1, pp. 131–194). Presses universitaires de France.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist*, 20(4), 167–182. https://doi.org/10.1207/s15326985ep2004_2
- Pestalozzi, J. H. (1827). *Letters on early education. Addressed to J. P. Greaves, esq.* Sherwood, Gilbert and Piper. <http://archive.org/details/lettersonearlyed00pestiala>
- Piaget, J. (1954). *The construction of reality in the child*. Basic Books. <https://doi.org/10.1037/11168-000>
- Piaget, J. (1970). Piaget's theory. In L. Carmichael (Ed.), *Carmichael's manual of child psychology*. Wiley. <http://archive.org/details/carmichaelsmanua0001carm>
- Piattelli-Palmarini, M. (Ed.). (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Harvard University Press.
- Plunkett, K., & Sinha, C. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, 10(3), 209–254. <https://doi.org/10.1111/j.2044-835X.1992.tb00575.x>
- Pratt, L., & Jennings, B. (1996). A survey of transfer between connectionist networks. *Connection Science*, 8(2), 163–184. <https://doi.org/10.1080/095400996116866>
- Prot, B. (2012). Formation d'un concept potentiel et transformations de l'activité. In Y. Clot (Ed.), *Vygotski maintenant* (pp. 307–327). La Dispute.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 131–141. <https://doi.org/10.1145/3351095.3372879>
- Roth, W.-M., & Jornet, A. (2017). *Understanding educational psychology: A late Vygotskian, Spinozist approach*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-39868-6>
- Rousseau, J.-J. (1762). *Emile: Or on education* (A. Bloom, Trans.). Basic Books.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362).

- Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). MIT Press. <http://dl.acm.org/citation.cfm?id=104279.104284>
- Saussure, F. de. (1915). *Course in general linguistics*. McGraw-Hill.
- Sejnowski, T. J. (2018). *The deep learning revolution*. MIT Press.
- Silverstein, M. (1985). The functional stratification of language in ontogenesis. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 205–235). Cambridge University Press.
- Simmel, G. (2004). *The philosophy of money*. Routledge.
- Simondon, G. (1958). *On the mode of existence of technical objects* (C. Malaspina & J. Rogove, Eds.). University of Minnesota Press.
- Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions* (Vol. 2). Cambridge University Press.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88(2), 135–170.
- Todes, D. P. (2014). *Ivan Pavlov: A Russian life in science*. Oxford University Press.
- Tolstoy, L. N. (1967). The school at Yásnaya Polyána. In L. Wiener (Trans.), *Tolstoy on education*. University of Chicago Press.
- Turing, A.M. (1948). Intelligent machinery. In B. Meltzer & D. Michie (Eds.), *Machine intelligence 5* (pp. 3–23). Edinburgh University Press.
- van der Veer, R., & Valsiner, J. (1991). *Understanding Vygotsky: A quest for synthesis*. Blackwell Publishing.
- Vygotsky, L. S. (1979). The development of the higher forms of attention in childhood. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 189–240). Routledge.
- Vygotsky, L. S. (1987). Thinking and speech. In R.W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky, Vol. I: Problems of general psychology* (trans. N. Minnick) (pp. 39–285). Plenum.
- Vygotsky, L. S. (1997a). Consciousness as a problem for the psychology of behavior. In R. W. Rieber & J. Wollock (Eds.), *The collected works of L. S. Vygotsky: Problems of the theory and history of psychology* (Vol. 3). Springer US. <https://doi.org/10.1007/978-1-4615-5893-4>
- Vygotsky, L. S. (1997b). Genesis of higher mental functions. In R. W. Rieber (Ed.), *The collected works of L. S. Vygotsky: The history of the development of higher mental functions* (Vol. 4, pp. 97–119). Springer US. <https://doi.org/10.1007/978-1-4615-5939-9>
- Vygotsky, L. S. (1997c). On psychological systems. In R. W. Rieber & J. Wollock (Eds.), *The collected works of L. S. Vygotsky: Problems of the theory and history of psychology* (Vol. 3, pp. 91–107). Springer US. <https://doi.org/10.1007/978-1-4615-5893-4>
- Vygotsky, L. S. (1997d). The instrumental method in psychology. In R. W. Rieber & J. Wollock (Eds.), *The collected works of L. S. Vygotsky: Problems of the theory and history of psychology* (Vol. 3, pp. 85–89). Springer US. <https://doi.org/10.1007/978-1-4615-5893-4>

- Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development* (NED-New edition, pp. 125–148). University of Minnesota Press. <https://www.jstor.org/stable/10.5749/j.ctttj0x.12>
- Wertsch, J. V. (1985). The semiotic mediation of mental life: L. S. Vygotsky and M. M. Bakhtin. In E. Mertz & R. J. Parmentier (Eds.), *Semiotic mediation: Sociocultural and psychological perspectives* (pp. 49–71). Academic Press.
- Wertsch, J. V., & Stone, C. A. (1978). Microgenesis as a tool for developmental analysis. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 1(1), 8–10.
- Wheeler, M. (2011). Thinking beyond the brain: Educating and building, from the standpoint of extended cognition. *Computational Culture*, 1. <http://computationalculture.net/beyond-the-brain/>