Irreducible worlds of inexhaustible meaning: early 1950s machine learning as subjective decision making, creative imagining and remedy for the unforeseen

Aaron Mendon-Plasek

Information Society Project, Yale Law School, USA
Email: aaron.mendon-plasek@yale.edu

Abstract

Little historical work examines the problems, practices and values of 'machine learning' as it was understood and justified within pattern recognition research communities prior to the 1980s. This omission has led to a failure to appreciate how the efficacy of machine learning was often justified by its perceived capacity for 'originality' rooted in machine (and human) subjectivity. This paper examines why and how early 1950s pattern recognition researchers came to see 'machine learning' as a technical and epistemological set of nominalist strategies for performing induction and abduction given incomplete, complex or contradictory information that might also spur 'creative' insight in such diverse activities as political judgement and scientific inquiry. I document local research problems, epistemological commitments, institutional contexts and the circulation of 'machine-learning' practices and values through three cases of early-career researchers imagining, building and programming digital computers to 'learn' from 1950 to 1953. This machine learning implemented in learning programs came to be seen by some researchers as more efficacious descriptions of the natural and social world because these descriptions were perspective-dependent, profoundly contingent and contextually non-exhaustive. Often indexed as colloquial appeals to greater 'generality', conceptions of machine learning's efficacy as the capacity to make meaning from contradictory information continues to inform contemporary debates regarding artificial intelligence, society and possibility.

If programs identified what was 'significant' for a particular decision, Oliver Selfridge argued in 1955, they might make good decisions even if the programmer's understanding of those decisions was inadequate.[1] The problem to be solved, Selfridge wryly contended, was identifying the relevant 'stimulus' in order to respond to it.[2] How did a dog (or a human) 'classify all cases of bell ringing' as equivalent when 'bell ringing never gives the same set of nervous impulses into the brain twice'?[3] Selfridge saw this 'extraction of the *significant* features from a background of irrelevant detail' as *the* fundamental problem of pattern recognition.[4] To that end, a pattern recognition program that identified what was 'significant' might identify many dissimilar things as equivalent, such as many different handwritten As as an *A*. But it might also categorize identical data in

---

1 Oliver Selfridge, 'Pattern recognition and modern computers', in *Proceedings of the March 1–3, 1955, Western Joint Computer Conference*, AFIPS '55 (Western), Los Angeles: Association of Computing Machinery, 1955, pp. 91–3, 92.

2 Oliver Selfridge, 'Pattern recognition and learning', in Colin Cherry (ed.), *Information Theory: Papers Read at the Symposium on 'Information Theory' Held at the Royal Institution, London, September 12th to 16th 1955*, London: Butterworths Scientific Publications, 1956, pp. 345–53, 345.

3 Selfridge, op. cit. (2), p. 345.

4 Selfridge, op. cit. (1), p. 92, original emphasis.

different ways depending on context – say, for example, identifying the same handwritten symbol as the letter O in one context but as the number 0 in another. '[C]ontext is a function of experience', Selfridge reasoned, '[b]ut more than that, experience alone affects the *kind* of thing we regard as significant'.[5] Accordingly, Selfridge suggested such programs learned 'new patterns' – effectively '[n]ew guesses' to try out – built up through the experience of new input data, clever combinations of existing patterns, and a maximization principle he called 'hedony'.[6] This, he argued, might enable machines to develop a repertoire of actions not explicitly programmed or even anticipated by their creators that would be useful for decision making involving poorly understood causal relations, extraordinarily complex systems and contradictory evidence.[7] Even more exciting to Selfridge was that these learning programs might inductively generate novel hypotheses to aid in reimagining scientific and social possibility.[8]

Selfridge's mid-1950s popularizations of 'significance' occurred in a context where artificial intelligence, cybernetics and pattern recognition researchers had already separated into distinct but mutually overlapping communities of practice in the anglophone world, and within each there was a considerable heterogeneity of beliefs, values and motivations. Engaging each other in debates from the 1950s to the 1980s about the meaning, role and value of learning, little agreement existed between or within AI, cybernetics or pattern recognition research communities about what research problems were worth prioritizing. Rather, a constructive disunity of beliefs and values emerged – analogous to the productive disunities that Peter Galison has traced in physics subcultures – regarding what 'learning' was, how it might be implemented and the respective meanings it entailed within these three communities.[9] The term 'machine learning' was increasingly used among pattern recognition researchers pursuing 'non-numerical' applications of computing by 1953 to denote both technical implementation and epistemological investigation.[10] These researchers working on machine learning saw their own investigations as providing descriptive tools for examining contingent human categories, a method for scientific discovery and a decision-making strategy when objectivity was neither desired nor possible.[11] I explain how this collection of 'machine-learning' concerns emerged by describing *what* 'learning' was to an early 1950s coterie of researchers who saw it as valuable, *how* this conception of 'learning' emerged contingently from implementing programs on 1950s analogue and digital computing machinery, and *why* learning programs came to be seen by these researchers as a strategy for acting when the relevant causal mechanisms were ill-defined or simply unknown.

I argue that early 1950s researchers seeking to build learning programs brought together a distinctive and contingent arrangement of practices linking (1) local technical

---

5 Selfridge, op. cit. (1), p. 92, original emphasis.

6 Selfridge, op. cit. (2), p. 349.

7 Aaron Mendon-Plasek, 'Mechanized significance and machine learning: why it became thinkable and preferable to teach machines to judge the world', in Jonathan Roberge and Michael Castelle (eds.), *The Cultural Life of Machine Learning*, Cham: Palgrave Macmillian, 2021, pp. 31–78, 40–2.

8 Selfridge, op. cit. (2), pp. 349–52; see also Mendon-Plasek, op. cit. (7), pp. 48–9, 53–4.

9 Peter Galison, *Image and Logic: A Material Culture of Microphysics*, Chicago: The University of Chicago Press, 1997, pp. 781–844.

10 For early uses of the term 'machine learning' in relation to analogue and digital computing see Andrew Booth and Kathleen Booth, *Automatic Digital Calculators*, London: Butterworths Scientific Publications, 1953, p. 211; Floyd Steele, 'An evaluation of basic machine processes', in *Proceedings of the Western Computer Conference held by the Joint IRE–AIEE–ACM Computer Conference Committee, Los Angeles, California, February 4–6, 1953*, New York: Institute of Radio Engineers, Inc., 1953, pp.28–33, 33.

11 Aaron Mendon-Plasek, 'Genealogies of machine learning, 1950–1995', PhD dissertation, Columbia University, 2022, at https://doi.org/10.7916/a2t9-r965. See also Matthew Jones, 'How we became instrumentalists (again): data positivism since World War II', *Historical Studies in the Natural Sciences* (2018) 48(5), pp. 673–84.

implementations on existing computing infrastructures, (2) models of scientific inquiry writ large and (3) visions of good governance in the early 1950s. This arrangement of concerns spanned local, professional and social spheres while incorporating, as Samuel Franklin documents, then-new Cold War conceptions of 'creativity', and forms of scientific identity and organization, as Steven Shapin discusses, tied to 'big science' projects.[12] Given these concerns, justifications for learning programs offered by pattern recognition researchers prior to the 1980s rarely, or only incidentally, involved the percentage of labels 'correctly' identified for a data set by a program. Given the variability of training and test data and of the methods of learning, pattern recognition researchers were painfully aware of how profoundly inadequate percentage-correct scores were for comparing different machine-learning systems. Mid-century pattern recognition researchers wove together technical and popular ideas of learning, science and creativity into an explicit strategy for imagining *new* possibilities and interpretations. Jamie Cohen-Cole has argued that human-sciences researchers in the early 1950s engaged in a 'reflexivity' that jumped between 'technical' and 'folk' concepts to both 'attack their foes' and justify and even spur their scientific imagining.[13] Similarly, those working on 'machine learning' in the early 1950s intermixed practical computing experience with popular notions of 'science' and 'learning' in an attempt to seek out genuine scientific novelty that could reshape their vision of reality much as they saw quantum mechanics or relativity as having done. Such mechanized creativity that might allow machine learning to reorganize, recategorize and reconstitute knowledge of the natural world, some pattern recognition researchers hoped, might also facilitate a social 'unity in diversity' through a 'liberal pluralism' that Cohen-Cole has discussed.[14]

Early machine learning in the 1950s provides an opportunity to explore forms of power through changing forms of human description. Ian Hacking has pithily called this 'making people up', in which the creation of new 'human kinds' that were thinkable at some times but not others emerge in lockstep with people who inhabited these new forms of life.[15] Janet Abbate and Stephanie Dick have argued that contemporary histories of computing grapple with power in how such histories reveal the 'social relations' and political economies that are enacted through technologies that produce 'a specific configuration of some peoples' minds and other peoples' bodies'.[16] This emphasis on social relations is reflected in much recent historical work on AI and machine learning, including historical critiques of AI as discussed by Ekaterina Babintseva, Olessia Kirtchik and Stephanie Dick, and in investigations of AI by Matthew Jones and Xiaochang Li.[17] All of the

---

12 Samuel Franklin, *The Cult of Creativity: A Surprisingly Recent History*, Chicago: The University of Chicago Press, 2023, pp. 6–10; Steven Shapin, *The Scientific Life: A Moral History of a Late Modern Vocation*, Chicago: The University of Chicago Press, 2008, pp. 165–208.

13 Jamie Cohen-Cole, 'The reflexivity of cognitive science: the scientist as model of human nature', *History of the Human Sciences* (2005) 18(4), pp. 107–39, 110.

14 Jamie Cohen-Cole, 'The creative American: cold war salons, social science, and the cure for modern society', *Isis* (2009) 100, pp. 219–62, 220.

15 Ian Hacking, *Historical Ontology*, Cambridge, MA: Harvard University Press, 2002, pp. 99–114, 99, 106.

16 Stephanie Dick and Janet Abbate, 'Beyond abstractions and embodiments', in Janet Abbate and Stephanie Dick (eds.), *Abstractions and Embodiments: New Histories of Computing and Society*, Baltimore: John Hopkins University Press, 2022, pp. 435–41, 439.

17 Ekaterina Babintseva, 'Engineering the lay mind: Lev Landa's algo-heuristic theory and artificial intelligence', in Abbate and Dick, op. cit. (16), pp. 319–40; Olessia Kirtchik, 'The Soviet scientific programme on AI: if a machine cannot "think", can it "control"?', *BJHS Themes* (2023), this issue, doi:10.1017/bjt.2023.4; Stephanie Dick, 'The Marxist in the machine', *Osiris* (2023) 38, pp. 61–81; Matthew Jones, 'Decision trees, random forests, and the genealogy of the black box', in Morgan Ames and Massimo Mazzotti (eds.), *Algorithmic Modernity*, Oxford: Oxford University Press, 2022, pp. 190–215; Xiaochang Li, 'The measure of meaning: automatic speech recognition and the human–computer imagination', in Abbate and Dick, op. cit. (16), pp. 341–59.

aforementioned position their studies of AI within larger debates about political organization, political economy and human values. Early machine-learning researchers examined here linked local technical practices concerned with efficacious decision making given contradictory data to popular conceptions of creativity that they saw as essential to successful scientists and democracies alike – which led some to question the very possibility of objective knowledge of self, science or society.

To understand this we need examples. In what follows I provide three that all occurred in the first few years of the 1950s. First, through an examination of Donald MacKay's work on digital and analogue computers, I trace how early machine learning invoked a radically constructivist epistemology in which machine 'originality' often precluded machine (and even human) objectivity, and was justified largely through analogies to narratives about how scientists created new science. Second, through an examination of Anthony Oettinger's 1952 reinforcement learning program, I show how the 'generality' of learning programs came to be seen as a solution to problems of contextual ambiguity and epistemic possibility. Finally, in examining Andrew and Kathleen Booth's 1953 digital-computing textbook, we see how the value of creative work performed by machine learning programs is remade into a notion of 'generality' as the capacity to redefine the scope of the tasks or decisions they were assigned.

Learning to be prejudiced and creative

After graduating with an undergraduate degree in natural philosophy from St Andrews in 1943, Donald MacKay got a radar job with the British Admiralty. Tasked with automating 'pilots' and gunfire, MacKay designed 'sense organs' and electronics 'to simulate situations … so as to provide goal-directed guidance for ships, aircraft, missiles and the like'.[18] Examining how combinations of humans and computers might be optimally deployed to perform specific tasks, MacKay wondered what activities might never be 'conceived in mechanistic terms' and subsequently automated.[19] Human brains struck him as quite dissimilar from either analogue or digital computers.[20] MacKay sought to evaluate different arrangements of analogue and digital computers applied to a particular task by recognizing the unique affordances specific to different forms of calculation. Digital computers facilitated calculations of high accuracy using physical phenomena to represent 'symbols in one–one correspondence with the set of *numerals*'.[21] In practice, calculation accuracy was limited by the difference between a number's value and the number of numerals that could be represented by the computer in memory. By contrast, analogue computers used physical magnitudes such as length or voltage to represent numerical magnitudes, which limited accuracy to usually just a few decimal places. Nevertheless, calculations like integration could be performed faster and cheaper with analogue machines than with digital ones.[22] Such trade-offs could make it difficult to know when to prefer one form of computing to the other, or, more importantly, how to best combine both to solve specific questions.

MacKay explored these trade-offs in an unpublished but widely circulated 1949 report, 'On the combination of digital and analogue computing techniques in the design of

18 Donald MacKay, *Behind the Eye*, Oxford and Cambridge: Basil Blackwell, 1991, p. 40. For Wiener see Peter Galison, 'Ontology of the enemy: Norbert Wiener and the cybernetic vision', *Critical Inquiry* (1994) 21(1), pp. 228–66.

19 MacKay, op. cit. (18), p. 40.

20 MacKay, op. cit. (18), p. 40.

21 Donald MacKay and Michael Fisher, *Analogue Computing at Ultra-high Speed: An Experimental and Theoretical Study*, New York: John Wiley & Sons, Inc., 1962, p. 3, original emphasis.

22 MacKay and Fisher, op. cit. (21), p. 4.

analytical engines'. He defined 'information' as that which 'changes' the 'valid[ity] in a logical pattern' and computation as 'transformations of information'.[23] This scheme allowed for comparing combinations of analogue and digital computers. More exciting to MacKay was that describing the world as information transformations might facilitate the incorporation of human meaning into non-numerical uses of computing. MacKay suggested 'human intercourse' pertaining to the 'compresence of [scientific] facts' as a useful thought experiment.[24] He imagined an 'autonomous artefact' that would collect disciplinary information, reduce that information to a 'logical vocabulary', identify '*fresh conclusions*' and 'fresh operations', and 'pursue an active, responsive, and logically-disciplined existence independent of human intervention'.[25] This artefact would engage in something like 'purposeful activity' as the 'attainment of a certain equilibrium configuration ... of incoming or present information'.[26] It might generate information from its 'logical "mill" to a human interlocutor', and use information provided by that 'interlocutor' as part of its subsequent input.[27] And the artefact would incorporate 'the *reaction of the field to the error*' as new information such that 'the machine after a period of active operation' would 'develop subsidiary characteristic patterns of behavior or "purposes" depending on its history'.[28]

MacKay's thought experiment led him to suggest 'purely metaphorical' comparisons between 'artefact' behaviour and 'analogous manifestations in human beings' in terms of signal, noise and 'threshold' responses triggered when the combination of signal and noise exceeded a certain value.[29] An artefact with a high signal-to-threshold ratio that incorporated statistical noise might be like a person gaining 'inspiration' from a 'random thought' which 'leads somewhere'.[30] Too little noise and high signal made an artefact predictable but without 'imagination'; too much noise and too little signal might foster 'distorted and hallucinatory impressions'.[31] The clear winner, for MacKay, was when signal was high and noise was opportunistically increased to produce 'bright ideas' that would be 'followed-up in a logical fashion' while not so high as to lead to 'insanity'.[32] MacKay noted that this speculative exercise was a 'totally inadequate' description; however, he argued, it suggested that a 'quite abstract general theory of information' describing calculators might 'have at least formal analogues in the mechanism of human thought'.[33] The artefact was interesting, MacKay concluded in his report, because it encouraged us to explore 'the science of knowledge itself'.[34] The artefact might, MacKay suggested, even generate 'original and unpredictable contributions to [a dialogue with people] in the terms and categories of human thought'.[35] MacKay gave his 1949 report to Warren McCulloch during the first Ratio Club meeting, at

---

23 Donald MacKay, 'On the combination of digital and analogue computing techniques in the design of analytical engines', mimeographed, 1949, reprinted in *Proceedings of the National Physical Laboratory Symposium on the Mechanisation of Thought Processes*, vol. 1, London: Her Majesty's Stationery Office, 1959, pp. 53–65, 56, 65.

24 MacKay, op. cit. (23), p. 59.

25 MacKay, op. cit. (23), pp. 59–60, original emphasis.

26 MacKay, op. cit. (23), p. 60.

27 MacKay, op. cit. (23), p. 61.

28 MacKay, op. cit. (23), pp. 62, 61, original emphasis.

29 MacKay, op. cit. (23), p. 63, 61.

30 MacKay, op. cit. (23), pp. 63.

31 MacKay, op. cit. (23), p. 64.

32 MacKay, op. cit. (23), p. 64.

33 MacKay, op. cit. (23), p. 64.

34 MacKay, op. cit. (23), p. 65.

35 MacKay, op. cit. (23), p. 65.

which McCulloch was an invited guest. McCulloch subsequently circulated the report in the United States.[36] (Nine years later McCulloch recalled MacKay's 1949 report as '[o]ne of my old treasured possessions'.[37])

MacKay would earn his PhD in physics on analogue computing for solving differential equations at Kings' College London in 1951 at the age of twenty-nine. MacKay's interest in information caught the attention of Warren Weaver, and Weaver provided MacKay with a Rockefeller Foundation scholarship to visit the US in 1951.[38] MacKay spent half of the year touring the US, visiting more than 120 US labs as well as attending the eighth Macy conference in New York City.[39] The other half-year was spent working in McCulloch's University of Chicago laboratory examining how information might be treated in neurology. MacKay and McCulloch developed a close friendship, despite MacKay's scepticism 'that the brain could be plausibly thought of as a digital computer à la McCulloch–Pitts', referencing a 1943 paper on neurons.[40]

MacKay increasingly saw his work on information theory as an empirical and philosophical exploration of human communication, originality and agency. This was exemplified in his 'Mindlike behaviour in artefacts', an article submitted to the *British Journal for the Philosophy of Science* in March 1951. Rejecting simplistic analogies between computers and brains as having 'little merit', MacKay saw the qualities of '[o]riginality, independence in opinion, and the display of preferences and prejudices' as 'characteristic of a human mind'.[41] He used the word 'artefact' in place of 'machine' to suspend the reader's own prejudice regarding the possibility of an artificial construct exhibiting similar creative agency.[42] MacKay repeated Wiener's 1948 premise that 'goal-directed behavior' of an artefact necessitated that artefacts receiving 'feedback' on itself, its activity and its environment.[43] 'Goals' such as optimizing efficiency, maintaining equilibrium and developing 'superficially unpredictable characteristics capable of rational description in purposive terms' were hobby horses of cybernetics, which MacKay well knew, duly citing Ashby, Wiener, Pitts and McCulloch.[44] Yet MacKay noted that the simplifying 'assumption that all the data used are exact, and lead to deductions which are unique and certain', depended on an 'assumption that decomposition into elementary "yes-or-no" propositions is possible without distortion'.[45] Social activities like '[h]uman intercourse', MacKay noted, countermanded the latter assumption, and that '[d]ata are only moderately certain, and [that] the closest attainable approach to a unique conclusion is often an estimate of the relative probabilities of several'.[46]

This multiplicity of answers depended on potential hypotheses, and the credibility of these hypotheses changed given the existence of new hypotheses. An artefact's capacity

---

36 Steve Heims and Donald MacKay, 'Interview of Professor Donald M[a]cKay', pp. 1–30, 15–16, Steve Heims Papers, 1946–1987, interviews for 'The Cybernetics Group' by Steve J. Heims – Gerald—MacKay, 1968–1973, Box 1, Folder 6, MC 361, Department of Distinctive Collections, MIT Libraries.

37 National Physical Laboratory, *Proceedings of the National Physical Laboratory Symposium on the Mechanisation of Thought Processes*, vol. 1, London: Her Majesty's Stationery Office, 1959, pp. 68–9.

38 Heims and MacKay, op. cit. (36), pp. 16–17.

39 Ronald Kline, *The Cybernetics Moment: Or Why We Call Our Age the Information Age*, Baltimore: Johns Hopkins University Press, 2015, p. 59.

40 Heims and MacKay, op. cit. (36), p. 17, see also pp. 14–17, 24–25. For the 1943 paper, see Warren McCulloch and Walter Pitts, 'A logical calculus of ideas immanent in nervous activity', *Bulletin of Mathematical Biophysics* (1943) 5, pp. 115–33.

41 Donald MacKay, 'Mindlike behaviour in artefacts', *British Journal for the Philosophy of Science* (1951) 2, pp. 105–21, 105.

42 MacKay, op. cit. (41), p. 105.

43 MacKay, op. cit. (41), pp. 105–6.

44 MacKay, op. cit. (41), p. 108.

45 MacKay, op. cit. (41), p. 109.

46 MacKay, op. cit. (41), p. 109.

to learn a 'pattern or *Gestalt*' might generate additional subjectivities useful for generating novel premises, 'including hypotheses about its own mechanism for generating hypotheses'.[47] Using the consequences of previous actions to update its future activity might produce tendencies in the artefact's behaviour analogous to 'originality, independence of opinion, and, for example, illogical human characteristics as prejudice, and preference' and 'other "emotional" effects'.[48] MacKay speculated that such idiosyncrasies might allow the artefact to produce creative and surprising hypotheses, and facilitate an artefact's sensible response 'in the face of contradictory subsequent experience'.[49]

MacKay linked the subjectivity of an artefact to its capacity to perform useful abduction. This implicitly linked the value of a learning program to its capacity to respond in novel ways to unseen observations, contradictory or otherwise. Consider strategies for mechanically identifying *any* triangle *as* a triangle. MacKay attributed to McCulloch and Wiener a strategy in which an artefact identified triangles by testing candidates against a 'template'.[50] That this strategy invoked an 'ideal' meant that it might poorly handle cases not anticipated at the outset.[51] Mackay's own strategy worked via an analogy to a blindfolded person identifying triangles by running his finger along the shapes to be tested. To such a person, MacKay argued, 'the concept of triangularity is invariably related with and can be defined by the sequence of elementary responses necessary in the act of replicating the outline of the triangle'.[52] '*The elementary acts of replication*', MacKay stressed, '*define the basic vocabulary in terms of which the artefact describes its own experience*'.[53]

Rearranging known sequences of actions to produce novel sequences of actions, what MacKay called 'replications', was, by definition, abduction. Replications were effectively new hypotheses to be tried out in the world.[54] The framing of experience and originality in terms of replications meant that *any* question, decision or problem framed in such terms could be examined by such artefacts. *Generality* was the learning of 'hierarch[ies] of abstraction' in which replications were 'formed by the responses of each level [of abstraction] becom[ing] the subject matter of the response-vocabulary of the next higher [level]'.[55] The replications composed of sequences of actions useful for one context – say, identifying a letter on a page – became the vocabulary of useful actions for composing new sequences at a higher level of abstraction – such as parsing the grammar of a sentence.

Cognizant that building up replications across different scales and contexts would produce paradoxes, MacKay drew on a paradox with which he was intimately well acquainted: wave–particle duality.[56] Such a paradox, he wrote, was 'resolved neither by arbitrary denials of "reality" nor by "explanations" of one as "nothing but" an aspect of the other', but by deciding 'the appropriateness of a description' for 'any given situation'.[57] Multiple 'descriptions' could be 'valid', he argued, but '[p]aradoxes arise when concepts defined for one logical background are mixed carelessly with those defined for another'.[58]

47 MacKay, op. cit. (41), pp. 112, 121, original emphasis.

48 MacKay, op. cit. (41), pp. 109, 111.

49 MacKay, op. cit. (41), p. 116.

50 MacKay, op. cit. (41), p. 113.

51 MacKay, op. cit. (41), p. 113.

52 MacKay, op. cit. (41), p. 114.

53 MacKay, op. cit. (41), p. 114, original emphasis.

54 For 'replications' see MacKay, op. cit. (41), pp. 114–16.

55 MacKay, op. cit. (41), p. 116.

56 Donald MacKay, *Information, Mechanism, and Meaning*, Cambridge, MA and London: MIT Press, 1969, pp. 1–3.

57 MacKay, op. cit. (41), p. 118.

58 MacKay, op. cit. (41), p. 118.

These 'logical backgrounds', for MacKay, were the 'organization of action in the world' that a particular artefact built up through experience and that interwove the conditions of possibility for originality, creativity and generality.[59]

That MacKay's 'originality' entailed a radically constructivist epistemology can be seen in how MacKay thought a person's or an artefact's capacity to recognize originality was itself contingent on individual experience. What might appear to one person as another person's meaningless activity – what MacKay calls 'the "completely random component" of the human behaviour pattern *from the observer's point of view*' – may be seen 'as the exercise of free choice' when viewed from the perspective of the doer of that activity.[60] Actions appearing as random or devoid of meaning to one artefact may be of primary importance to another artefact as a consequence of the differences in each artefact's repertoires of replications. Such perspectivalism meant that an artefact's originality was predicated on its past experiences. Such an artefact would be deeply fallible, as people too were fallible. Two artefacts trained on the same data could develop different (but perhaps) complementary worlds through their respective 'experiences' of those data.

The very thing that gave an artefact the ability to learn new things, to learn in redundantly different ways and to handle contradictory inputs was also precisely what guaranteed that an artefact could not be objective. And it was precisely the artefact's subjectivity that gave it the potential to act efficaciously in ill-defined or poorly understood systems. By the same light, the activities of an artefact could not be considered apart 'from its environment, when the concepts of interest are essentially properties of the system-plus-environment'.[61] Properties of 'minds' were exactly of this sort such that the 'system-plus-environment' included 'conceptual frames of reference, defined respectively from the viewpoint of the actor and spectator' where '[d]escriptions in terms of only one group or the other may both be valid'.[62] This descriptive relativism pertaining to perspective also included an exchangeability between the artefact's repertoires of 'abstractions' and the artefact's input data: both could be 'exemplified by replication-operations of one kind or another'.[63] Such artefacts could have 'analogues of concepts such as emotion, judgment, originality, consciousness, and self-consciousness'.[64] Yet what appears to have been even more exciting to MacKay than simulating 'consciousness' was that his artefact might inductively and abductively formulate 'hypotheses about its own mechanism for generating hypotheses'.[65] This would soon be explored in a learning program for the Electronic Delay Storage Automatic Calculator (EDSAC) written by a new college graduate named Anthony Oettinger.

Generality as an efficacious response to novelty

Anthony Oettinger visited the University of Cambridge's Mathematical Laboratory as a Henry fellow during a gap year between pursuing his undergraduate and doctoral degrees at Harvard in applied physics and applied mathematics respectively. By 16 October 1951, Maurice Wilkes, in his capacity as Oettinger's supervisor at Cambridge's Mathematical Laboratory, introduced Oettinger to Alan Turing's 'Computing machinery and intelligence' (1950), which Oettinger described as 'funny as hell in parts' and posing 'a rather

---

59 Heims and MacKay, op. cit. (36), p. 6.
60 MacKay, op. cit. (41), p. 119, original emphasis.
61 MacKay, op. cit. (41), p. 118.
62 MacKay, op. cit. (41), p. 118.
63 MacKay, op. cit. (41), p. 118.
64 MacKay, op. cit. (41), p. 121.
65 MacKay, op. cit. (41), p. 121.

interesting problem'.[66] Between attending Bach concerts, sitting in on Paul Dirac's quantum mechanics course, and staying up nights to execute programs on the EDSAC, Oettinger wanted 'to get at some general principles' pertaining to a 'machine simulat[ing] a learning process' in order 'to determine whether it is possible to give the machine more freedom, without getting sheer nonsense out of it'.[67] Frequently working from 10 a.m. to midnight in the lab and reading about 'learning' in 'bio[logy], psych[ology], phil[osophy], and math', Oettinger's efforts culminated in 'Programming a digital computer to learn' (1952), which discussed his implementation of two learning programs on a digital computer.[68] Oettinger argued that the first program, which he called the 'shopping programme' or 's-machine', would pass a restricted version of Turing's 'imitation game' involving one question: 'In what shop may article *j* be found?'[69] However, such a constrained version of the test, Oettinger contended, highlighted the shopping program's 'limitations' as a less 'general programme'.[70] The program, barring computer-mechanical error, did exactly what it was instructed so that its 'range of operation [was] consequently limited'.[71]

Oettinger's more general solution was a second program he called the 'response-learning s-machine' that could learn new tasks through 'trial and error', to which he devoted the bulk of his 1952 article.[72] The response-learning program developed a particular 'response', what he called 'a habit', to particular repeated inputs.[73] The program wasn't told *what* task to perform or *how* to perform a task. Learning generalizability was what made a program better, and this quality was not defined by its performance at any specific task but by its ability to do what it wasn't told. Response-learning programs were to be judged by a different set of standards since external observation of a program's behaviour was inadequate.[74] A 'general' learning program's behaviour had the status of an experiment: patterns could be proven false but never true. To evaluate his response-learning program's habits he compared them to 'conditioned reflexes'.[75] The program accepted inputs (i.e. a number) and returned outputs (i.e. another number) that changed based on repeated approvals and disapprovals (i.e. a third number) that a person provided the program in response to the appropriateness of its outputs. The program could even learn some habitual responses from input stimuli even when no person-guided training was performed. Oettinger explored this behavior 'from the privileged point of view of the designer' by providing the system of equations that governed the response-learning behavior, including (1) changing 'threshold' states driven by 'random fluctuations' that triggered different responses, and (2) a memory 'decay' function in which learned habits were slowly forgotten.[76] These two features, Oettinger argued, facilitated more general learning.

The virtue of the response-learning program was in the program's capacity to learn new (i.e. not predetermined by the programmer) responses that might exceed the abilities

66 Anthony Oettinger, letter to Marilyn Tanner, 16 October 1951, 'Letters from Anthony Oettinger to Marilyn Tanner, 1951–1952', Acc. #2017.542, Box 1, Papers of Anthony Oettinger, HUGFP 59, Harvard University Archives; Alan Turing, 'Computing machinery and intelligence', *Mind* (1950) 59(236), pp. 433–60.

67 Oettinger, op. cit. (66), letters to Marilyn Tanner, 12 October 1951, 15 November 1951.

68 Oettinger, op. cit (66), letters to Marilyn Tanner, 29 November 1951, 26 January 1952; Oettinger, 'Programming a digital computer to learn', *Philosophical Magazine* (1952) 7, pp. 1243–63.

69 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1250.

70 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1250.

71 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1251.

72 Oettinger, 'Programming a digital computer to learn', op. cit. (68), pp. 1250–1.

73 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1254.

74 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1257.

75 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1251.

76 Oettinger, 'Programming a digital computer to learn', op. cit. (68), pp. 1251, 1255.

or understanding of the programmer. Using '[random] fluctuations the machine can make mistakes', but, Oettinger noted, '[t]his makes possible the teaching of a new response, or, when no response is favoured over the others, produces an interesting variety of responses'.[77] Oettinger noted that the decay 'introduces some lethargy into the behaviour of the machine, by causing all [threshold states] to drop [slowly over time], hence requiring ever increasing stimuli', effectively giving the learning program a short-term memory and the ability to forget.[78] This combination of 'random fluctuations' and forgetfulness made it possible for the program to exhibit 'decay or habit formation, or the steady state of frequent or varied responses'.[79] Producing valuable responses was not just a strategy of cultivating ever greater menageries of heterodox responses to data and seeing which worked. A complete model of the world was neither required nor desired, nor perhaps even possible. But a machine that learned directly from its environment, however limited, could develop 'habits' to do useful tasks.

These programs, Oettinger argued, might also serve as 'elementary building block[s] for a wide variety of more sophisticated systems' such that 'some [programs] serve as stimuli for others' to collectively perform more complex tasks.[80] For Oettinger, such collections of learning programs might learn how to do a task without being told how to do it. By invoking a whole repertoire of simple response-learning programs capable of developing habits that interdependently fed inputs and outputs to each other, Oettinger was suggesting not just that learning was a strategy for developing complex behaviors (think 'shopping'), but, more importantly, that learning was a way of acclimating machine actions to handle contextual ambiguity and human categories.

That Oettinger's program echoed some of the ideas we have seen in MacKay's 1951 paper – especially pertaining to the uses of signal, memory and decay, habit formation, and noise to generate more complex behaviors in learning programs – is no coincidence. Oettinger writes of attending a Ratio Club meeting on 8 February 1952, in which he notes that Turing would present a paper and W. Grey Walter and D.M. MacKay would be hosts.[81] Ten days later, at precisely the time when Oettinger and Wilkes were revising Oettinger's article, Oettinger visited MacKay at King's College, reporting that MacKay was 'very much interested in the same ideas as I, and we had a good time'.[82] Oettinger would attend the Ratio Club again, but his disposition towards his own learning program grew more vexed throughout the spring as the EDSAC frequently broke down and Wilkes urged new rewrites of his paper. Much worse, Oettinger wrote, was that 'the man I most wanted to talk to, as he'd done a good deal of work on this thinking business', Alan Turing, 'had to plead guilty to charges of "gross indecency" with another man as a result of some blackmail'.[83] After many revisions, Wilkes gave his approval for Oettinger to submit his paper for publication by 27 May 1952, with Oettinger writing that his final version was 'somewhat pessimistic' and that 'I had hoped for better results'.[84] His response-learning program associated certain numerical inputs with other numerical outputs – but this was hardly shopping.

---

77 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1255.
78 Oettinger, 'Programming a digital computer to learn', op. cit. (68), pp. 1255–6.
79 Oettinger, 'Programming a digital computer to learn', op. cit. (68), pp. 1255, 1257.
80 Oettinger, 'Programming a digital computer to learn', op. cit. (68), p. 1261.
81 Oettinger, op. cit. (66), letter to Marilyn Tanner, 7 February 1952.
82 Oettinger, op. cit. (66), letter to Marilyn Tanner, 19 February 1952.
83 Oettinger, op. cit. (66), letters to Marilyn Tanner, 27 March 1952, 2 April 1952.
84 Oettinger, op. cit. (66), letter to Marilyn Tanner, 27 May 1952.

Machine learning as the production of completely new ideas not deducible from known data

Andrew and Kathleen Booth's 1953 discussion of 'machine learning' appears to take their examples without attribution from Oettinger's work just discussed.[85] Having already built multiple digital computers at Birkbeck College, University of London – often by cannibalizing earlier computers for parts to make subsequent ones – Andrew and Kathleen Booth documented their efforts in a 1953 textbook entitled *Automatic Digital Calculators*.[86] Apologizing at the outset for the use of 'anthropomorphic term[s]', their use of 'machine learning' in the concluding chapter pre-dates by more than two years John McCarthy's introduction of 'artificial intelligence' in McCarthy's co-authored Dartmouth workshop proposal.[87] The Booths' linguistic choice was less rooted in aspirations of giant mechanical brains, androids or reproducing human intelligence, but, like MacKay, in the complications entailed by the switch from 'analogue' to 'digital' computers.

Digital computers were too fast. Analogue computers, as we have seen in discussing MacKay, were limited in the numerical precision they could provide because they represented numbers via material analogies.[88] Even a numerical error of less than one in a hundred, the Booths noted, was difficult to achieve. In contrast, digital computers could be 'arbitrarily' precise given appropriate implementation such that 'one in ten million' was not unthinkable.[89] This precision was necessary to solve problems that required many repetitive calculations. Rounding up or down the last digit of thousands of calculated numbers quickly led to an accumulated numerical error that could dwarf the computed value, rendering the entire calculation useless.[90] Since the potential precision of automatic digital calculators made it feasible to perform iterative calculations using previously computed values, the Booths noted that the bottleneck in the speed of digital computers was often the time it took for humans to interact with the machine.[91]

Six years earlier, they had argued that rather than waiting for human input, a digital computer could be ordered to make 'judgements' using the results it had computed from input data. Their incisive summary of the state of computing in the US and the UK was submitted for publication after visiting the Institute of Advanced Study at Princeton from March to August 1947.[92] Defining the 'translating of a problem in terms of the available functions of the machine … as programming', the Booths contended that a 'programme' was 'capable of controlling its own operations and exerting a certain amount of independent judgement'.[93] This was necessary because 'human agencies [were] too

---

85 Oettinger, 'Programming a digital computer to learn', op. cit. (68). For EDSAC information see Maurice Vincent Wilkes, 'The Edsac computer', 10–12 December 1951, Joint American Institute of Electrical Engineers–Institute of Radio Engineers Computer Conference: Review of Electronic Digital Computers, ACM, New York, USA, pp. 79–83; Wilkes, 'Automatic calculating machines', *Journal of the Royal Society of Arts* (1951) 100, pp. 56–90.

86 Booth and Booth, op. cit. (10), p. 211.

87 Booth and Booth, op. cit. (10), pp. v, 216; John McCarthy, Marvin Minsky, Claude Shannon, Nathaniel Rochester and Dartmouth College, 'A proposal for the Dartmouth Summer Research Project on Artificial Intelligence', 1955.

88 For the preference of analogue computers over digital computers see Matthew Jones, 'Calculating devices and computers', in Bernard Lightman (ed.), *A Companion to the History of Science*, Chichester: John Wiley & Sons Ltd, 2016, pp. 472–87; see also Joseph November, *Biomedical Computing: Digitizing Life in the United States*, Baltimore: Johns Hopkins University Press, 2012.

89 Booth and Booth, op. cit. (10), pp. 2–3.

90 Booth and Booth, op. cit. (10), pp. 2–3.

91 Booth and Booth, op. cit. (10), p. 3.

92 Andrew D. Booth and Kathleen H.V. Britten, 'Principles and progress in the construction of high-speed digital computers', *Quarterly Journal of Mechanics and Applied Mathematics* (1949) 2(2), pp. 182–97.

93 Booth and Britten, op. cit. (92), pp. 183–4.

slow to control machines which are much faster than those already in use'.[94] For the Booths, the machine 'judgement' constituted by the computer's ability to take different actions in response to its calculations (using input data and criteria specified at the outset by the human operator) already constituted a form of learning.

Their extensive mechanical knowledge and mathematical acumen meant that few were better prepared to equip digital computers with such 'judgement' than the Booths. Andrew's father had trained Andrew to be comfortable doing calculus by the age of ten. By the time he earned his PhD in physics from the University of Birmingham in 1943, Andrew had built three analogue computers for X-ray crystallographic work, something that brought him to the US in 1946.[95] Kathleen earned her PhD in applied mathematics from the University of London in 1950, focusing on supersonic aerodynamics. Alongside Zenia Sweeting, she was responsible for the bulk of the construction of a computer that the Booths and Sweeting co-created for the British Rubber Producer's Research Association prior to the Booths' Princeton trip in 1947.[96]

While the bulk of the Booths' 1953 *Automatic Digital Calculators* provided the reader with a mechanical description of the components of a digital computer and theoretical description of some relevant computational methods, its concluding section discusses applied problems. After the mathematical problem of X-ray crystallography, the remaining three applications of 'mechanical translation' (of science articles), 'games' (e.g. cards, noughts and crosses and chess) and finally 'machine learning' were all 'non-numerical'. 'Machine learning', they wrote, was of at least two kinds.[97] The first was analogous to teaching a dog through reward and punishment. The learner remains unaware of the teacher's objectives. The second involved goal-oriented learning; that is, specifying a goal and how to achieve it, which the Booths compared to teaching 'partially educated' children specific tasks like multiplication, reading poetry or playing the piano.[98]

This second kind of learning facilitated faster computation, analogous to the 1947 version of computer 'judgement' discussed earlier. It involved a computer program that remembered the outcomes of earlier actions and inputs. It essentially 'learned' by updating its information given inputs, reconfiguring its actions according to different numerical or logical conditions defined by the programmer. Being goal-directed meant that the programmer could state the conditions the program could use to determine whether the program had achieved its goal. The specific program the Booths provided to illustrate this type of learning was a 'shopping' program, detailed in the form of (1) a diagrammed sketch, (2) a series of program orders and (3) an inventory of memory addresses and contents. The program attempted to obtain specific papers at different simulated shops, both represented in memory by numerical codes, by following a given algorithm and by remembering which articles were at which store.[99] The shopping program adjusted its actions by remembering which stores did or did not have articles, avoiding or frequenting them as appropriate when retrieving future articles.

In contrast, the Booths' called the first kind of learning they discussed 'conditioned reflex', much as Oettinger had done.[100] And like Oettinger's response-learning program,

---

94 Booth and Britten, op. cit. (92), p. 186.

95 Andrew Booth, 'Andrew D. Booth, an autobiographical sketch', *IEEE Annals of the History of Computing* (1997) 19(4), p. 58. Regarding X-ray crystallography and early molecular biology see Soraya de Chadarevian, *Designs for Life: Molecular Biology after World War II*, Cambridge: Cambridge University Press, 2002.

96 Booth, op. cit. (95), p. 60.

97 Booth and Booth, op. cit. (10), p. 211.

98 Booth and Booth, op. cit. (10), pp. 211–12.

99 Booth and Booth, op. cit (10), pp. 212–13.

100 Booth and Booth, op. cit (10), p. 214.

the Booths' program was trained through a series of interactions with its environment and a corresponding series of 'encouragement[s]' and 'punishment[s]' provided by a trainer (i.e. a mathematical function or a human operator) after each interaction, until the program produces a desired performance.[101] Such a process, the Booths argued, could teach a program that initially outputs a string of random numbers to output π.[102] But the point was to produce a repertory of 'genuine learning programmes inside the machine so that its memory faculty and elementary reasoning powers (e.g., shopping) could effectively be used to speed up its education' and that this might 'constitute a valid reason for asserting that a machine can learn from experience'.[103]

The Booths did not restrict intellectual activity to logical operations. They noted that 'intelligence' was 'a matter of taste, since intelligence would mean two entirely different things to, say, a dock labourer and a university professor'.[104] Definitions of intelligence, they hinted, provided little insight and missed the point. They instead prefered 'to define relative intelligence by relative capacity for creative work' via either (1) the '[d]eduction of new conclusions from existing facts' or (2) the '[p]roduc- tion of completely new ideas not deducible from known data'.[105] The latter form of cre- ative activity, they suggested, was 'extremely rare' and 'a result of random processes', but had the potential to identify 'a new law of nature' for the machine's 'human ancil- laries' to investigate.[106] For a machine to be 'creative', the Booths implied, you must allow the machine to redefine the form of the answer you expect, just as you would any person. A human 'genius', they argued, might consider the statement 'the moon is made of green cheese', and not merely either accept or reject the claim, but 'con- sider it in detail and thus possibly discover a fundamental natural law'.[107] Implicit in their concluding example was that a creative act involved speculative inference, and an appropriate inference machine could ascertain or redefine the appropriate scope of relevance directly from its input data and its interactions with its environ- ment. Their machine learning was not just another version of naive behaviourism. Rather, the Booths were pre-emptively and unknowingly flipping the script of much future 1950s and 1960s AI work: instead of employing limitations of scope as a neces- sary evil to make an AI system pragmatically usable, machine learning could turn the inability to specify the scope of a problem into an epistemological virtue by defining the relevant scope as whatever caused the machine to generate acceptable answers in test cases.

The Booths' use of 'machine learning' as a term of art prior to McCarthy's popular- ization of the term 'artificial intelligence' contradicts the contemporary understanding of machine learning as a sub-field of artificial intelligence, pointedly suggesting that there were other traditions of learning machines not tied to the commitments and values of early AI that privileged human imitation. The Booths' 1953 discussion sug- gests that a distinct form of learning-machine epistemology was already well estab- lished specifically to address problems for which multiple 'correct' answers might be permissible. This approach relied on the available data to define the appropriate scope of the phenomena of interest, and did not require a definition of the objects of inquiry but just the specification of relevant data to the inquiry. The ability to not specify the object of inquiry encouraged a different form of answer to social

---

101 Booth and Booth, op. cit (10), pp. 214–215.
102 Booth and Booth, op. cit. (10), p. 214.
103 Booth and Booth, op. cit. (10), p. 215.
104 Booth and Booth, op. cit. (10), p. 216.
105 Booth and Booth, op. cit. (10), p. 216.
106 Booth and Booth, op. cit. (10), p. 216.
107 Booth and Booth, op. cit. (10), p. 216.

problems that emphasized continuity with the past while allowing a computational system to intervene in the present.

Epilogues and new directions: meritorious descriptions of the political and social world

Oettinger's subsequent doctoral work after returning to Harvard was on machine translation, 'automatic data processing' and computer design and logic. He went on to be the youngest professor to get tenure at Harvard at that time. The Booths pursued similar research subjects throughout the 1950s. Andrew's efforts on machine translation were already well known by 1953 and Kathleen apparently worked on neural networks for character recognition in 1958 or 1959.[108] They continued to build a variety of commercially funded computers. MacKay's questions about representation led him away from information theory and towards philosophical- and physiology-inflected studies of perception, communication, choice and free will. For each researcher considered here, their work on learning programs was professionally transitory but intellectually generative. They are all better known today for what they did after 1953.

Yet their conceptions of learning as the capacity to respond appropriately to unexpected or contradictory new data by generating interpretations that might complement, surprise or challenge human interpretations informed and even defined what a valuable problem was in pattern recognition for the next three decades.[109] It would inform the possibilities for 'machine learning' in its rebranding as a sub-field of artificial intelligence in the late 1980s.[110] As the tasks of both interpreting identical patterns of data in different ways and recognizing different configurations of data as categorically identical were seen as different incarnations of induction, their implementation of induction in learning programs led 1950s pattern recognition researchers to conceive of 'creativity' as the generation of new hypotheses (i.e. abduction) to perform induction.[111] Rather than centring themselves as experts, as AI researchers often did, they foregrounded their own ignorance of the tasks they would have their machines perform in the hopes those machines might 'learn' to do something the researchers themselves did not know how to do. Having more data regardless of its provenance was not in the service of more accurate or faster prediction, but instead was motivated by a particular conception of generality as a machine's (and human's) capacity to respond in robust, relevant or even creative ways to what *was not* and *could not* be predicted.

Crucially, what a problem situation entailed for pattern recognition researchers was very different to what it entailed for the emulating-experts approach of artificial-intelligence research. Lots of data might provide an inventory of exceptions that a learning program might memorize for classifying future 'edge' cases – but better accuracy alone missed the point. Lots of data were more valuable because they facilitated induction. Mechanized induction could be performed by a learning program if 'knowledge' itself was the program's actions and the effects of these actions. In this way, for these researchers, the technical classifications performed by learning programs mirrored the *inductive* and *abductive* process of discovery by a person – including how one conducted

---

108 For Kathleen Booth's 1958–9 neural-network work on optical character recognition, see Roger Johnson, 'Kathleen Booth: UK Computer Pioneer', 24 April 2023, Computer Conservation Society, at https://youtu.be/kj1rBGf2i7k. I have been unable to substantiate this claim.

109 Mendon-Plasek, op. cit. (7).

110 Mendon-Plasek, op. cit. (11).

111 For machine learning as abduction see Luke Stark, 'Artificial intelligence and the conjectural sciences', *BJHS Themes* (2023), this issue, pp. 1–15, doi:10.1017/bjt.2023.3; Mendon-Plasek op. cit. (11), Chapters 3, 4.

a scientific investigation. Early machine learning is a particularly fruitful ground for reimagining the intertwined practices of computation and quantification in the twentieth century, and the new possibilities for 'making people up' in the twenty-first century.[112]

Early AI and early 'machine-learning' researchers' disagreements were less about the efficacy of one or another approach and more about the values and forms of knowledge that these approaches were seen as being capable of facilitating. Those working in AI, pattern recognition and cybernetics in the 1950s attended many of the same conferences and workshops, and read each others' publications, all while advocating for radically different technical practices and methodologies that belied different visions of self and society, and different possibilities pertaining to how each could be known. For example, Hunter Heyck, Stephanie Dick and Jonnie Penn have each examined how the Logic Theory (LT) Machine program reformulated mathematical proof proving as a deductive problem implemented on a digital computer that later came to be seen as an exemplar of AI research.[113] Simon's conception of 'bounded rationality' and the LT program engaged in the dual premise (1) that there was only one rational 'course of action' given all relevant information and values for a situation, and (2) that no person could consider all relevant information pertaining to a complex rational decision.[114] This view had political consequences. Simon saw organizations and institutions as the means by which individual subjectivity could be overcome through a kind of collective rationality – such rationality of collectives aligned with Simon's democratic ideals even as it de-emphasized individual agency.[115] It also led Simon to seek out problems which, as Heyck writes, 'could be described unambiguously', like chess and proof proving, which were formulated deductively, and, as Dick observes, led Simon to understand computers, human minds and organizations as instantiations of information processing.[116]

The early machine-learning researchers examined here came to information theory as a solution for reimagining what the world was and for escaping their own myopia. Their work on computers focused explicitly on ambiguous situations and poorly defined tasks that led them to see the problem of significance as a problem of nominalism, and that all other problems could be reimagined as problems of nominalism. The early machine-learning researchers I examine frequently had to draw upon a plurality of correct solutions, saw objective rationality as a special limiting case of subjectivity, and saw efficacious induction as facilitated by 'creativity' – where creativity itself was abduction that tacitly incorporated human values and commitments from examples. The contrasting positions of early machine-learning researchers and early AI researchers reflected different political and epistemological commitments, and yet these differences were articulated and clarified through debates about 'learning' that brought AI and machine-learning researchers into frequent and constructive contact despite their profound disagreements about the meaning and interpretations of their work.

To explain how learning programs came to be seen as offering credible descriptions of society and self – indeed if we want to understand the prominent role of machine learning in descriptions of society – we cannot resort to detailing sub-communities building different learning programs as if these efforts were used to compel the agreement of many

---

112 Hacking, op. cit. (15).

113 Hunter Heyck, 'Defining the computer: Herbert Simon and the bureaucratic mind – Part 2', *IEEE Annals of the History of Computing* (2008) 30(2), pp. 52–63; Stephanie Dick, 'Of models and machines: implementing bounded rationality', *Isis* (2015) 106(3), pp. 623–34. Penn argues that this program was labelled AI only after the fact. See Jonnie Penn, 'Inventing intelligence: on the history of complex information processing and artificial intelligence in the United States in the mid-twentieth century', PhD dissertation, University of Cambridge, 2021, p. 77.

114 Heyck, op. cit. (113), p. 59; Dick, op. cit. (113), p. 624.

115 Heyck, op. cit. (113), pp. 59–60.

116 Heyck, op. cit. (113), p. 59; Dick, op. cit. (113), p. 625.

other sub-communities. We instead need to understand how learning as a trading zone limited the forms of exchange and possibility across sub-communities. The LT program became a solution and a kind of evidence for realizing objectivity through institutions for Simon. For MacKay a learning program was a solution and a kind of evidence for recognizing what was significant to a task given a plurality of descriptions. MacKay, and the niche community of 'machine-learning' researchers in pattern recognition that he exemplifies, developed a nominalist-inflected vision of creativity through their efforts to implement learning programs on electronic computers, and linked their technical practices to larger political commitments. More broadly, the early histories of machine learning reframe the role of induction in twentieth-century sciences, and suggest that forms of social description and individual identity were often circulated and made durable because of a disunity of computing communities. Our histories of AI and machine learning have only begun to reflect this insight.