A Cybernetic Theory of Persons: How Sellars Naturalized Kant

Carl B. Sachs
Program of Philosophy in the School of Humanities
Marymount University
csachs@marymount.edu

Abstract: I argue that Sellars's naturalization of Kant should be understood in terms of how he used behavioristic psychology and cybernetics. I first explore how Sellars used Edward Tolman's cognitive-behavioristic psychology to naturalize Kant in the early essay "Language, Rules, and Behavior". I then turn to Norbert Wiener's understanding of feedback loops and circular causality. On this basis I argue that Sellars's distinction between signifying and picturing, which he introduces in "Being and Being Known," can be understood in terms of what I call cybernetic behaviorism. I interpret picturing in terms of cycles of cybernetic behavior and signifying in terms of coordination between cybernetic behavior systems, or what I call triangulated cybernetic behavior. This leads to a formal, naturalistic understanding of personhood as the capacity to engage in triangulated cybernetic behavior. I conclude by showing that Sellars's thought has the resources, which he did not exploit, for introducing the concept of second-order cybernetics. This suggests that Sellars's philosophy of mind could be developed in the direction of autopoiesis and enactivism.

0. Introduction

It is one thing to say that we can or should 'naturalize Kant', but quite another to specify in any detail what (if anything) that means – not least of which because the very phrase seems to be a contradiction in terms. Nevertheless, there is also a surprisingly long tradition of philosopher-scientists who aspired to do exactly this, beginning at least with early neo-Kantians such as Helmholtz. I do not think it controversial to suggest that Wilfrid Sellars belongs to this tradition, though it may be controversial to suggest that Sellars not only attempted to naturalize Kant, but to a remarkable extent that has not yet been fully appreciated, actually succeeded in doing so.

The linchpin of my interpretation relies on what Michael Friedman (2001) calls "philosophy as metascience". On Friedman's account, one important role for philosophical speculation is the

generation of new candidate explanatory frameworks during a Kuhnian scientific crisis.[1] I suggest that Sellars's philosophy of mind should be read as a *metascience of mind* during an interregnum between behaviorism and cognitive science, when anomalies within behaviorism were accumulating and the gathering trends that would become the cognitive revolution had not yet fully congealed. Yet Sellars makes extensive use of the history of Western philosophy, from Plato through the moderns to Kant, Hegel, pragmatism, and positivism for the resources his metascience of mind requires. Hence what follows is but a preliminary sketch of how Sellars's metascience of mind established some conceptual foundations of cognitive science by translating key insights of Kantian transcendental psychology into a behavioristic-cum-computational register.

In what follows, I shall begin Sellars's first attempt at 'naturalizing Kant' in his "Language, Rules, and Behavior" (1949), which turned on a remarkable and very suggestive synthesis between Edward Chace Tolman's "cognitive behaviorism" and an emphasis on "symbolic activity" that has a strongly Kantian flavor (§1). This will be followed by a somewhat longer explication of cybernetics, which has a significance for Sellars that is unfortunately almost universally neglected, and as a result of which his centrally important concept of picturing has been (I shall contend) misunderstood (§2). On this basis I will suggest a somewhat novel interpretation of Sellars's contributions to semantics and philosophy of mind (§3) before remarking on the extent to which Sellars retains any contemporary relevance to philosophy of cognitive science (§4).

[1] "Science, if it is to continue to progress through revolutions, therefore needs a source of new ideas, alternative programs, and expanded possibilities that is not itself scientific in the same sense – that does not, as do the sciences themselves, operate within a generally agreed upon framework of taken for granted rules. For what is needed here is precisely the creation and stimulation of new frameworks or paradigms, together with what we might call meta-frameworks or meta-paradigms – new conceptions of what a coherent rational understanding of nature would amount to – capable of motivating and sustaining the revolutionary transition to a new first-level or scientific paradigm" (Friedman 2001: 23).

One of the interesting features of Sellars's philosophy that can be brought out more clearly through a careful analysis of his engagement with the sciences of his time is his complex view of intentionality. In Haugeland's (1998) famous baseball metaphor of the positions about intentionality, he notes that there is an intermediate position between "second-base" neobehaviorism and "third-base" neopragmatism. About this, however, he says only: "Wittgenstein may have been a short-stop".[2] On the reading of Sellars I develop here, Sellars develops the short-stop position into a sophisticated and defensible view because he shows how to affirm *neopragmatism* about intentionality with respect to the manifest image and *neobehaviorism* about intentionality with respect to the scientific image – a contrast that he articulates in his distinction between "signifying" and "picturing".[3]

1. Symbolic Activity as Cognitive Behavior

To assess the importance of Sellars's philosophy of mind as the metascience of psychology, I want to begin where Sellars himself began: with a retrospective analysis of the debates over psychologism. These debates played a significant role in the formation of philosophy as an academic discipline, including the establishment of both phenomenology and logical positivism, both of which were formative influences on Sellars.[4] In an early text (Sellars 1947/2005a) Sellars begins by announcing that the founding move of analytic philosophy is to distinguish philosophical from psychological problems, and though he never rejected the need to distinguish

[2] In the game of baseball as played in the United States, the short-stop is a defensive position located between the defensive positions at second base and at third base. For this reason, Haugeland uses the short-stop as a metaphor for a theoretical position between neobehaviorism ("second base") and neopragmatism ("third base").
[3] A comprehensive treatment of how Wittgenstein and Sellars differ in how they occupy the short-stop position is beyond the scope of this essay.
[4] For the role of psychologism in shaping philosophy as an academic discipline, see Kusch (1995).

normative from empirical concepts, he was also, consistently, concerned to establish the legitimacy of this very distinction.

I would like us to pay careful attention to how Sellars takes up, in the late 1940s, once again the problem or question of 'psychologism'. Although Sellars does not explicitly notice the connection, it is worth stressing that the critique of psychologism took for granted a specific conception of psychology itself: experimental introspectionist psychology in the grand tradition of Wundt, Titchener, and others. But we must notice (even if, perhaps, Sellars himself did not) that not all of the original arguments against psychologism can proceed once the paradigm of psychological research is no longer introspectionist but behavioristic. For example, Frege's complaint that publicly valid assertions or thoughts cannot be reduced to private mental episodes does not work when the objects of psychological research are themselves publicly observable behavioral acts. If there is something importantly right about the critique of psychologism, it must nevertheless be substantially revised in order to be applicable to behaviorism. And this is in effect what Sellars sets out to do.

In this early and unpublished note entitled "Psychologism", Sellars begins by articulating, in Kantian terms, the basic problem of the fate of epistemology at the midpoint of the 20th century.[5] The classical conception of epistemology was by this time beleaguered on two sides by well-respected and prominent campaigns aimed at overcoming epistemology in the traditional sense: logical positivism and American pragmatism. Logical positivism aimed at replacing epistemology insofar as they pursued a radical anti-psychologism that transformed epistemology into the logical analysis of science. What had been epistemology became, in the hands of the Vienna Circle and by their own admission, analytic *a priori* assertions – and hence, on the

[5] This text is now published as an Appendix to Olen (2018).

Tractarian account of analyticity that they also accepted, tautologous assertions. If the logical positivists replace epistemology with the tautologies of logical analysis, then perhaps, Sellars ventures, we should side with the pragmatists like John Dewey. Here, in sharp contrast to the anti-psychologism that shaped the context within which logical positivism emerged, we find an extremely sophisticated biologically grounded social psychology of scientific inquiry. Taking up the project developed in Dewey's *Logic* would also be a rejection of epistemology as classically conceived. The question, then, is whether there is a way of avoiding the replacement of epistemology by either logic or by science – that is, whether we could somehow salvage the very idea of synthetic *a priori* assertions, and with it, the distinct status of epistemology as not something that could be replaced by the analytic *a priori* assertions of logic or by the synthetic *a posteriori* assertions of psychology.

Though Sellars was already by 1949 tentatively sketching out the revival of the Kantian alternative to both positivism and pragmatism, he nevertheless understood the imperative of undertaking a careful examination of what positivism and pragmatism had contributed to epistemology, especially with regard to the whole question of "psychologism." To assess the viability of the arguments against psychologism, and therefore to examine whether epistemology as a synthetic *a priori* enterprise could even be vindicated, Sellars needed to first carry out a careful construction and analysis of the most sophisticated (at the time) psychological explanation of our cognitive capacities. It is only by considering what is missing from the most sophisticated psychology of cognition that we would be in the right position to vindicate the need for a genuinely *a priori* element to epistemology. At the same time, however, Sellars accepts that we need, somehow, to reconcile Kant's emphasis on the *a priori* with Dewey's naturalism: we need to sketch an *Aufhebung* between Kant and Dewey.

The next major stage of Sellars's work in which he undertakes the synthesis of Kant and Dewey is in his "Language, Rules, and Behavior" (hereafter LRB).[6] The initially stated goal of this text is "to explore from the standpoint of a philosophically oriented behavioristic psychology the procedures by which we evaluate actions as right or wrong, arguments as valid or invalid, and cognitive claims as well or ill grounded" (211) – in short, we are to construct a "psychology of the higher processes" that makes contact with the structure of normativity as such, whether ethical, logical, or epistemic norms. That is, we are to begin where Dewey left off, with a naturalistic, behavioristic psychology, and construct a bridge that allows us to make contact with Kantian concerns. We cannot determine by mere intuition whether or not normativity can be naturalized; we can only determine whether normativity can be naturalized by attempting to naturalize it and then evaluating whether or not the attempt is successful.

What Sellars aspires to here is a *via media* between "rationalistic apriorism" and the idea that "all meaningful concepts and problems belong to the empirical or descriptive sciences". That is, we want to avoid "descriptivism" –a tendency into which pragmatism tends to lapse – while also avoiding "rationalistic apriorism" and its accompanying "pseudo-psychology of cognitive givenness". Thus, on the one hand we must reject the error at the very heart of rationalism: the pseudo-psychology on which it depends. It depends on the error that one can, through an act of mere noticing, of the sort that would be nicely botanized by introspectionist psychologists, come to awareness of the basic underlying structures of the world (or indeed of the mind itself). In calling the cognitive given a 'pseudo-psychology' Sellars is implicitly relying on how behavioristic psychologists would evaluate introspectionist psychology.[7] Yet on the other hand,

[6] Originally published in 1950. All page references are to the 1980 reprint in *Pure Pragmatics and Possible Worlds* edited by Jeffrey Sicha.

[7] For a behavioristic polemic against introspectionism, and one that perhaps influenced Sellars, see Tolman (1932: 233-234).

"a sound pragmatism must reject descriptivism in all areas of philosophy, and that it can do so without giving one jot or tittle to what has so aptly been called the New Failure of Nerve" (213). Here Sellars is referring to Sidney Hook's article of that title in *Partisan Review* (1943), where Hook uses this phrase to refer to the tendency amongst those otherwise committed to a secular, scientific worldview to give in too readily whenever a need for pious reverence for eternal verities is announced. Thus, Sellars is explicitly aligning himself with Hook (who edited the volume in which LRB first appeared) and with Dewey (in whose honor the volume was written) while at the same time cautioning pragmatism not to reject all of the insights of the rationalism that it had come to oppose.

To advance the rapprochement between rationalism and pragmatism, Sellars admits that he needs to address the philosopher's concern that psychology is not even relevant to philosophy. Why does the philosopher need to be concerned with a psychologist of symbolic behavior? "What would the relevance of an adequate empirical psychology of rule-regulated symbol activity to the task of the philosopher?" (218). If one were to insist that the philosopher and the psychologist are engaged in different enterprises, why should the philosopher pay attention to the psychologist? The answer is that "bad psychology may give aid and comfort to bad philosophy" (ibid.) – that is, when we are correcting bad philosophy, we should notice how much it depends on bad psychology. But we cannot do this unless we have at least a passing acquaintance with good psychology, especially with what scientific psychology might come to say about the higher processes. In short, we will not know what we shall need to say when doing epistemology until we know more about the conceptual resources that are missing from the cognitive psychology of rational behavior.

Though my use of the term cognitive psychology is anachronistic, a closer reading of the text suggests that this is precisely Sellars's concern. For his purposes, it will not suffice to carve the distinction between epistemology and psychology as between the higher, more sophisticated processes and those more primitive behaviors that we share with other animals: "To content oneself with glib phrases about stimulus-response conditioning is to give the rationalist armor and armament. … It is easy to shape the psychology of the higher processes as embodied in common sense into the direction of intuitionism and rationalism. Philosophers have been doing just that for over two thousand years" (220). For this reason, the embattled empiricist has urgent need of "an adequate psychology of rational behavior" (ibid.).

The most important philosophical function of "an adequate psychology of rational behavior" – of cognitive psychology *avant la lettre* – is to help the pragmatist philosopher overcome the bad "pseudo-psychology of cognitive given-ness" on which rationalism and intuitionism have traditionally relied. On that model, the mind has, as it were, a *single* kind of cognitive relation: it can directly apprehend the objects referred to by terms occurring in syntactico-semantic structures (sentences, theories). Thus, one apprehends abstract entities of all sorts -- universals, generals, kinds, etc. – in exactly the same way that one apprehends physical objects described by the common and proper sensibles. Sellars raises several objections to this "pseudo-psychology" over the course of his work, but I want to focus on what I shall call *the circularity objection.* The circularity objection hinges on the following thought: in order to *begin* to apprehend abstracta or universals, we would need to be able to *notice* them. But we cannot notice them without having the requisite concepts. But according to this pseudo-psychology, the requisite concepts are directly apprehended. Thus, we cannot directly apprehend abstracta or universals, as abstracta

and universals, *unless we already have*.[8] Put otherwise, the advocate of the Given cannot avoid a "dormitive virtue" pseudo-explanation, and that is why the psychology of givenness is a pseudo-psychology.

The beginning of an alternative to the introspectionist pseudo-psychology on which rationalism depends lies in taking seriously behavioristic psychology, beginning with the thought that "most if not all animal behavior is tied to the environment in a way in which much characteristic human behavior is not" although learned habits of response "remain the basic tie between all the complex rule-regulated symbol behavior which is the human mind in action, and the environment in which the individual lives and acts" (ibid., 217). Crucial here is the naturalistic conviction that we are to envision the human individual as an animal in an environment, although we should consider the environment to be 'social' as well as 'physical'. But what, exactly, does Sellars have in mind by "animal behavior" here? Although Sellars refers to behavioristic psychology in general terms, there is one specific reference that deserves closer scrutiny: the idea of a cognitive map.

Shortly before Sellars wrote LRB, the American psychologist Edward Chace Tolman published what was to become a foundational text in the transition from behavioristic to cognitive psychology: "Cognitive Maps in Rats and Men" (1948).[9] Here Tolman summarizes experiments on maze learning in rats, carried out by his graduate students and himself, to show that, contrary to the widespread view of animal behavior at the time, animal learning cannot be explained exclusively through reward-driven associations. Rather, he argued, we need to think of

[8] Sellars makes this line of thought explicit in EPM §45, SPR p. 176.
[9] Tolman uses the concept of a map for the methodology of science as early as 1932, which he seems to have borrowed from his friend the pragmatist philosopher Stephen Pepper; see Tolman 1932: 424-426. The innovation represented by the 1948 paper is that maps are not only a metaphor for scientific theories but also an analogy for animal (and human) cognition generally.

animals as having a map-like model of their environments that they are testing against experience and revising as necessary in order to achieve their goals and satisfy their needs. Animal behavior is not only purposive (as Tolman argued in his 1932 text) but genuinely cognitive. Hence, I shall follow Baars (1986) in referring to Tolman's position as "cognitive behaviorism", though this is not a term that Tolman himself used.[10] Though I do not mean to marginalize the importance of naturalized teleology for Tolman's purposive behaviorism (as he called his view), for present purposes I want to focus on importance of cognitive processes. On my reading, Tolman's cognitive behaviorism inspires Sellars to imagine an adequate psychology of the higher processes: one that begins with cognitive behaviorism and tries to explain *rational* behavior in terms of *cognitive* behavior.

The second prong in Sellars's strategy is to think of our symbolic activity as essentially rule-governed or rule-regulated. Here too Sellars is treading on familiar ground he has inherited from Charles Morris on signs, the Wittgenstein of the *Blue and Brown Books*, and what he learned of Cassirer from (at least) Langer's translation of *Sprach und Mythos*. What matters most to Sellars about this kind of activity is that is, in a sense difficult to articulate precisely, "free" activity – which is not to say that it is "uncaused" but rather to say that (1) it is concerned with imagining or conceptualizing non-actual possibility, and indeed with different kinds of possibility (logical, mathematical, physical), which is crucial to counterfactual reasoning and experimental testing, and also (2) the constraining rules of symbolic activity are *themselves* grounded in our acquired (but revisable) commitment to those rules. We can revise those normative constraints themselves – not by abandoning all rules, but by changing one rule for another. Hence our "rule-regulated

[10] There was – and remains – a lively debate as to whether Tolman was committed to realism about cognitive maps or accepted them on merely instrumentalist grounds. Though a fascinating chapter in the history of cognitive psychology, exploring it is beyond the scope of this paper. However, I believe that Sellars's own philosophy of science commits him to realism about cognitive maps regardless of the best interpretation of Tolman.

symbolic activity" includes the intellectual summits of Einstein, Leibniz, and Cantor: the freely undertaken construction of new domains of syntactical and semantic structures through which our comprehension is enlarged and transformed.

The distinction between "tied behavior" – habitual responses to the environment – and "symbolic activity" – rule-regulated symbolic structures that comprise our intellectual life – is the opening move in the critique of the pseudo-psychology of cognitive givenness. As Sellars understands the state of play, the rationalist has the advantage over the naturalist for their emphasis on the inspiring intellectual achievements made in mathematics and science – but the naturalist has the advantage over the rationalist for diagnosing the cognitive given as a pseudo-psychology, the Achilles' heel of rationalism. The alternative, which Sellars emphasizes is little more than a promissory note (or at least it was in 1948), replaces the single-function account of intentionality or mindedness with a dual-function account. The crux of this account, it should be emphasized, is not simply the distinction between "tied behavior" and "rule-regulated symbol behavior" – after all, even the rationalist who has read Watson would allow for that much. Rather, what matters is that these two kinds of behavior are inextricably meshed together. If symbolic activity were not meshed together with tied behavior, it would have no causal hook-up to the environment and consequently it would be wholly irrelevant to both perception and action. If not for its meshing together with tied behavior, symbolic activity could neither structure sensory input in the form of observations nor structure motor outputs in the form of volitions. In the absence of structuring both observations and volitions, symbolic activity would be idle if it were innate (since it could not affect perception and action) and unacquirable if it were not (since no one could learn it via observation and imitation).

What, then, does the meshing together of tied behavior and rule-regulated symbol behavior require? As Sellars sees it, "in order for the above mentioned meshing of rule-regulated language with tied symbol behavior to take place, *certain intra-organic events must function as symbols in both senses, as both free and tied symbols*" (220). That is, we need to posit neurological events – or at least neurological/non-neurological biological events – that can function as both (1) belonging to a system that coordinates purposive responsiveness to the ambient environment and (2) belonging to system characterized as a syntactico-semantic structure constituted by its own logical and material rules of inference. Let us call these *hinge events*.[11] In other words, we need to replace the single-function model of the rationalist with a dual-function model, as long as we understand that there must be hinge events: some neurological events must participate in both cognitive functions in order for them to remain coordinated (however loosely) sufficient for symbolic activity have causal bearing on the world in perception and action.[12]

Thus far I have argued for the important role of LRB in Sellars's search for an *Aufhebung* of rationalism and pragmatism, looking to both Kant and to Dewey for inspiration and guidance (among many others). The account offered in LRB is, however, a promissory note in several notable respects. In order to contextualize the route that Sellars's thought took subsequent to LRB, I want to underscore two crucial issues that Sellars neglects in LRB. First, though Sellars introduces the concept of a cognitive map and suggests that symbolic activities (including but not limited to logic, mathematics, and science) can be transposed into a naturalistic framework by

[11] The distinct status of hinge events is resumed in Sellars's much later discussion of "natural-linguistic objects" in *Naturalism and Ontology*.

[12] It is also true that the dual-purpose model is crucial to Sellars's nominalism, and it allows him to say what the rationalist wants to say about universals or kinds without a commitment to a non-naturalistic metaphysics. But while this is a strength of the Sellarsian view – if one endorses metaphysical naturalism – I shall treat it as a corollary rather than an objection.

seeing them as tools for constructing better cognitive maps much like those posited by Tolman, he does not articulate any causal mechanism whereby cognitive maps can be constructed and revised – without which, Sellars's naturalization of rationalism must be half-baked by his own lights. Second, Sellars does not fully articulate how we should think about the relation between the ineliminably normative and *a priori* nature of epistemology and the 'adequate psychology of rational processes' that LRB has begun to sketch. Both of these issues occupied much of Sellars's subsequent philosophical development. I shall argue that the solution to both of these problems can be found in his mature conception of the distinction between signifying and picturing, especially in the version of that distinction that Sellars develops in "Being and Being Known".

2. The Scientific Image of Intentionality

At the end of "Being and Being Known" (hereafter BBK) Sellars remarks that "recent cybernetic theory has begun to shed light on how cerebral patterns and dispositions picture the world".[13] This remarkable claim tells us that Sellars sees a deep connection between his account of picturing and what was once called cybernetics. Much like behaviorism, cybernetics has been largely forgotten because the revolution that it began has become mainstream (even though, in both cases, some of the deepest insights were forgotten along the way). What began as the science of "control and communication in animal and machine" – the subtitle of Wiener's 1948 monograph-manifesto – relatively soon evolved into computer science, information theory, and AI. Ironically, by the time that Sellars started making substantive use of cybernetic ideas, it was

[13] Originally published in 1960; all citations to reprinted as 1963a.

already beginning to be eclipsed as a serious science. For this reason (among others) the importance of cybernetics for Sellars's philosophy of mind has been, until recently, wholly neglected. Yet I shall argue that a better understanding of cybernetics is the key to Sellars's scientific image of intentionality, what he calls "picturing".

The term "cybernetics" was coined by the American mathematician-philosopher Norbert Wiener from the Greek word "kubernetes", a steersman or helmsman on a boat. The basic idea of cybernetics at the time was to refer to what were also called, at the time, "teleological mechanisms," or mechanisms capable of self-governance or self-control. An exceptionally crude precursor of such systems is the Watts governor used in steam engines. The Watts governor enables the steady production of heat by preventing too much heat from being produced: when the system overproduces, the governor closes off the supply of fuel until the pressure has decreased. The invention of electronic relays in the 20th century obliged engineers to design circuits with feedback loops so that noise can be filtered out and signals amplified relative to noise – at the same time mathematicians needed to develop a sophisticated analysis of the very concepts of "information" and "noise" that were of concern to engineers. Cybernetics was born from the need to conceptualize, operationalize, and realize the concepts central to information theory, computer science, and their sequelae.[14]

The crucial notion that Sellars absorbs from cybernetics is the idea of feedback. In Wiener's formulation, feedback is indispensable "when we desire a motion to follow a given pattern the difference between this pattern and the actually performed motion is used as a new input to cause the part regulated to move in such a way as to bring its motion closer to that given by the

[14] See Kline (2017) for the history of cybernetics, but especially the personal and political factors that led to its eclipse. In large part cybernetics was re-branded as information theory and as computer science; it also led directly to chaos theory, complexity theory, evolutionary robotics, autopoiesis, artificial intelligence, and cognitive science.

pattern" (Wiener 1948: 6-7).[15] In light of this, feedback is essential to patterned behavior in general: patterned behavior is possible due to feedback that corrects actual deviations, errors, or noise relative to what expected or desired. In the case of designed systems, it is the designers who know what pattern they want to see generated and institute feedback loops in order to generate the behavior that they intend. In the case of naturally evolved cognitive systems, there is no designer, and yet can say that patterned behavior emerges from the feedback loops between the cognitive map (as a spatio-temporal map of the environment and the place of the organism in that environment) and the environment to which that map is structurally coupled via transducers and effectors. The concept of feedback is also crucial here because it allows us to understand how Sellars transforms the concept of picturing that he has borrowed from Wittgenstein's *Tractatus.* Put much too simply, Sellars uses the concept of feedback to give Tractarian picturing a cybernetic twist.[16]

A corresponding change in the basic metaphysics is required by the new science of cybernetics, since we now must understand change not only in terms of energy but also in terms of the then-new concept of information:

> the newer study of automata, whether in the metal or in the flesh, is a branch of communication engineering, and its cardinal notions are those of message, amount of disturbance or 'noise' – a term taken over from the telephone engineer – quantity of information, coding technique, and the like. In such a theory, we deal with automata effectively coupled to the external world, not merely by their energy flow, their

[15] I am focusing on Wiener partly because of his historical importance and partly because Sellars had a copy of Wiener (1948) in his personal library, though it was not his only source of information about cybernetics.
[16] This becomes the key move in how to understand rule-regulated behavior – a rule is a generalization that tends to make itself true by virtue of how the norm is enacted through feedback loops between members of the community.

> metabolism, but also by a flow of impressions, of incoming messages, and of the actions of outgoing messages. (Wiener 1948: 42)

Of particular interest to the history of the scientific image of mind is how Wiener insisted on conceptualizing the central nervous system in cybernetic terms. Two passages are noteworthy for the parallel between Wiener and Sellars:

> The central nervous system no longer appears as a self-contained organ, receiving inputs from the senses and discharging into the muscles. On the contrary, some of its most characteristic activities are explicable only as circular processes, emerging from the nervous system into the muscles, and re-entering the nervous system through the sense organs, whether they be proprioceptors or organs of the special senses. (Wiener 1948: 8)

and

> [F]or effective action on the outer world it is not only essential that we possess good effectors, but that the performance of these effectors be properly monitored back to the central nervous system, and that the readings of these monitors be properly combined with the other information coming in from the sense organs to produce a properly proportioned output to the effectors. (Wiener 1948: 96)

As we shall see, this is precisely how Sellars characterizes the 'anthropoid robot of the future' in "Being and Being Known" as having internal computational states that covary with the states of the environment and its body due to feedback loops between processors, effectors, and transducers.

It is also noteworthy, I think, to stress that Wiener regards cybernetics as bearing directly on the question as to whether logic is reducible to psychology – that is, to "psychologism":

> The science of today is operational; that is, it considers every statement as essentially concerned with possible experiments or observable processes. According to this, the study of logic must reduce to the study of the logical machine, whether nervous or mechanical, with all its non-removable limitations and imperfections. … any logic which means anything to us can contain nothing which the human mind – and hence the human nervous system – is unable to encompass. (Wiener 1948:125)

By appealing to a version of operationalism, Wiener is able to suggest that the meaning of logical statements is equivalent to the procedures used by a computing machine – whether metal or meat – used to verify those statements. This does not reduce logic to psychology, nor psychology to logic – but it does transform logic into the science of the formal properties of cognitive machinery. All of this is nicely taken on board by Sellars throughout the 1950s, so that by the early 1960s, Sellars is finally in a position to use cybernetics for conceptualizing a scientific image of mind that allows him to articulate a comprehensive, philosophically adequate alternative to "the pseudo-psychology of cognitive givenness" upon which rationalism depended.[17]

In the development of this alternative, "Being and Being Known" (Sellars 1963a) deserves special status because it is here that Sellars explicitly invokes cybernetics in his scientific image of intentionality. This conception is developed through a close criticism of the Aristotelian philosophy of mind located in (among other places) Thomism. Sellars does this for two main reasons. The first is that he thinks that there are important insights in the Aristotelian tradition that have been overlooked by the modern approach that begins with Descartes. The second, and

[17] Though Sellars lobbies this accusation in rationalism in LRB, by the time he writes "Empiricism and the Philosophy of Mind" he has realized that the myth of the given is a problem not only for rationalism and for empiricism but even for Kant and Hegel. For a brief reconstruction of the history of philosophy of mind aimed at making sense of this claim, see Sachs (2020).

more important, is that Sellars's philosophical method suggests the following commitment: the rational defensibility of his conception depends on its place within the dialectic of the history of philosophy of mind.

Sellars suggests that we accept something like the Aristotelian distinction between the sensitive soul and the rational soul, which he understands as a distinction between cognitive functions as systematically related to the environment and cognitive functions as governed by rules. But how is this distinction itself to be understood? As Sellars sees it, this distinction is not one between *kinds* of cognitive function but rather between different *ways of thinking* about what cognitive functions are. For while we have a few thousand years of theorizing about cognitive functions using the conceptual resources of the manifest image, we can also begin to compare those theories with the account of cognitive functions using the conceptual resources of the scientific image. To do this, Sellars engages in the thought-experiment of imagining an "anthropoid robot of the future" (Sellars 1963a: 51) – something that, perhaps falling short of genuine artificial general intelligence, might be within the next few generations of Mars rovers.

Sellars's starting point is to accept the traditional idea – going back at least to Aquinas – that there is an isomorphism between the intellect and the world – that *veritas est adaequatio intellectus et rei*. But he suggests that this isomorphism must be understood in two very different senses, and that nothing but confusion results from conflating these two distinct senses. The two senses refer to different "orders": "the logical order" and "the real order." The former is the explication of the order of understanding (*ratio cognoscendi*); the latter is the explication of the order of being (*ratio essendi*). In the real order, the isomorphism of intellect and world is what he calls "picturing"; in the logical order, the isomorphism of the intellect and the world is "signifying". Confusion between these two orders has led to the misbegotten Platonic-

Aristotelian idea that the intellect signifies the world by being informed by immaterial natures or "forms". Hence Sellarsian nominalistic materialism requires a sharp demarcation between signifying and picturing.

What we need at this point is an answer to the question, "what are we talking about when we talk about how the intellect pictures the world?" And to this I suggest that Sellars's answer is: *cybernetics*. It is cybernetics that Sellars is alluding to when he writes, "I shall present the distinctions I have in mind as they appear when projected into discourse about computing machines, guided missiles, and robots" (Sellars 1963: 51). These are more or less standard examples in the cybernetics literature of the 1940s through 1970s.[18] Here is how Sellars himself describes a thought-experiment of cybernetics:

> Suppose such an anthropoid robot to be 'wired' in such a way that it emits high frequency radiation which is reflected back in ways which project the structure of its environment (and its 'body'). … Suppose such a robot to wander around the world, scanning its environment, recording its 'observations', enriching its tape with deductive and inductive 'inferences' from its 'observations' and guiding its 'conduct' by 'practical syllogisms' which apply its wired-in 'resolutions' to the circumstances in which it 'finds itself'. It achieves an ever more adequate adjustment to its environment, and if we permitted ourselves to talk about it in human terms (as we have been) we would say that it *finds out* more and more about the world, that it *knows* more and more *facts* about what took place and where it took place, some of which it *observed*, while it *inferred* others from what it did *observe* by the use of *inductive generalization* and *deductive reasoning*. (Sellars 1963a: 52-53; emphases original)

[18] See Rosenblueth, Wiener, and Bigelow (1943) and Wiener (1948).

We can, from the standpoint of the electronic engineer or cybernetician, consider the states of the robot as building up a picture of the environment – although "this picturing cannot be abstracted from the mechanical and electronic processes in which the tape is caught up" (Sellars 1963a: 53), or as we might say today: cognition is *both* computational *and* necessarily embodied and embedded. Just as the grooves on a record player cannot be understood apart from the procedures by which records are produced and played, so too the computational states of the robot cannot be understood apart from the physical *habitus* of the robot.[19]

It is widely accepted that Sellars borrows the concept of picturing from the early Wittgenstein's *Tractatus Logico-Philosophicus.* It has not been noted, however, that Wittgenstein and Sellars invoke the same examples of a picturing system. Specifically, Wittgenstein also uses the record as an example of picturing at *TLP* 4.014: "A gramophone record, the musical idea, the written notes, and the sound-waves all stand to one another in the same internal relation of depicting that holds between language and the world" (Wittgenstein 1974: 20). The crucial difference is that Sellars emphasizes that the grooves on a record picture sounds by virtues of the feedback loops involved in the production and use of records. Sellarsian picturing, unlike Tractarian picturing, is a cybernetic concept.

Sellars's invocation of computational states may seem to clearly anticipate what has become known as 'the computational theory of mind', especially in the versions promoted in Putnam in mid-1960s and by Fodor in the late 1970s. However, there is a crucial difference between CTM and Sellars's cybernetics. As Sellars sees it, the computational states that comprise the mind cannot be disentangled from the whole network of behaviors in which they are embedded, as made vivid by his comparison of the mind with a vinyl record. If one were to carefully examine

[19] See Huebner (2018) for a detailed explanation for why the physical *habitus* of the robot is necessary for understanding the analog computations and analog representations that comprise the robot's 'mind'.

the surface of a vinyl record, one can discern hundreds of thousands of grooves etched into it. But in order to understand why that record has the grooves that it does, one needs to understand the record in context, both as the result of a manufacturing process whereby sounds are converted into a semi-stable form and as something that can be inserted into an audio system designed to reproduce the sounds that were translated into the record when it was manufactured. The structure of the grooves is a consequence of the transposition of the structure of the music from an acoustic medium to a vinyl medium.[20] In the same way, the computational states of the mind are a 'materialization' of the features of the environment that caused those states via perceptual episodes. The key difference is that the structures are distorted or modulated at the same time that they are transposed from environment to mind, so that the relationship between them is not a simple matching but rather a highly dynamic structural coupling between computations and environmental features.[21]

I shall call this position *cybernetic behaviorism*. It differs from Tolman's "cognitive behaviorism", which was already important for LRB, by explicitly drawing upon cybernetics for concepts (e.g., feedback loops) and examples (e.g., guided missiles) in theorizing about how cognitive maps are constructed and updated. Thus, while Tolman argues that intelligent, purposive behavior is best explained by positing a map-like mental model of the features of the

[20] In one crucial respect the vinyl record analogy is misleading. The recording and production process allows for near isomorphism between the acoustic properties of the music and the grooves in the record, which is why vinyl is preferred even today by purist audiophiles. MP3s and other compression formats are comparatively quite "lossy" – there is loss of information as the signal is compressed – meaning that the mapping relation between playback and original is homomorphic, not isomorphic. Lossy compression formats are almost certainly a better metaphor for animal sensory systems than the near isomorphism of vinyl recordings; see Akins (1996). However, vinyl records are a useful metaphor because they record signals in an analog format, rather than a digital one, although we probably do not have a clear understanding of how much analog vs digital processing there is in neuronal assemblies.

[21] In other words, Sellars accepts with the cognitivists that the mind is comprised of computational states that function as representations of the environment, but he also insists, along with the anti-representationalist and 4E proponents, that cognition is necessarily embodied and embedded. See Huebner (2018) and Sachs (2018).

environment constructed as the animal sensed and interacted with that environment, he did not propose any underlying mechanism. Sellars, by contrast, uses cybernetics to propose an underlying computational basis to thought driven by sensorimotor feedback loops.[22]

Unlike a more "Cartesian" form of cognitivism, Sellars underscores the importance of the physical *habitus* of the cognitive system matters – the kinds of maps it will construct is inseparable from its iterated feedback loops with ambient environments.[23] We can therefore understand picturing as follows: feedback loop driven updating of nonconceptual representational states functionally embedded in a computational information processing system that, as a dissipative structure, continually exchanges causal flows of energy-matter with its ambient environment.

The thought experiment of the BBK robot thus puts a cybernetic spin on purposive behaviorism: the robot's purposive behavior can be explained from the perspective of the electrical engineer in terms of feedback loops between two systems – the ambient environment and the robot – informationally coupled through transducers and effectors. The upshot of the thought experiment is that the cognitive friction with the environment that both rationalists and empiricists sought to explain with something Given – whether "the *illuminatio* of Augustine" or "the *data* of the positivists" (Sellars 1963b: 356) – can be explained entirely by adopting the scientific image of mind: cybernetic behaviorism.

Cybernetic behaviorism is crucial for understanding Sellars's argument for why semantic terms such as "means", "refers to," and "is about" do not designate a relation between mind and

[22] However, it was not until the 1980s that Sellars finally applied the insights from cybernetics *directly* to biological systems, and not just as an analogy with them. It was at this time that he developed what he came to call "animal representational systems." In this regard Sellars was influenced by the cognitive revolution.

[23] The exact relevance of this account to contemporary debates between cognitivism and 4E cognition depends in part on whether the relation between the computational states of the robot and its body and environment is one of coupling or constitution; see Rowlands (2010). Further exploration of this point is beyond the scope of this paper.

world.[24] These terms belong to the manifest image of intentionality: they are the product of millennia of philosophical reflection on the world of everyday life and experience, and they have valuable roles to play in the elucidation of discourse. But although semantic terms have a 'surface grammar' of designating a mind-world relation, taking them at face value inevitably leads to positing intensional entities: meanings, propositions, *Sinne*, *ιδέα*. Sellars's strategy for securing a nominalistic, materialistic metaphysics does not (*pace* McDowell) require him to deny that intentionality is a mind-world relation; rather, his strategy is to argue that the manifest image conception of intentionality is now replaceable by a scientific image of intentionality.[25] In the scientific image of intentionality, we retain the manifest image commitment to the idea that intentionality is a mind-world relation. The crucial difference is the exact nature of the mind-world relation is not based on a conceptual explication of semantic vocabulary but rather on a causal explanation of cybernetic mechanisms.

3. Cybernetics, Community, and Personhood

Based on this admittedly quick and crude sketch of what I have been called cybernetic behaviorism, I shall now develop what I take to be a Sellarsian solution to long-standing problems in the philosophy of mind – chiefly, the nature of *content*, *intentionality*, or *meaning*, which is (Sellars thinks) a problem at the very heart of what it means to be a thinking thing. Though Sellars inherits much from German Idealism and American pragmatism with regard to

[24] McDowell (1998) ascribes to Sellars the position that intentionality is not a mind-world relation. For an incisive criticism of this interpretation, see Shapiro (2011).
[25] However, there is another sense of intentionality, "the language of individual and community intentions", which persists in the scientific image. What is replaced by picturing is the sense of intentionality that involves the world-directedness of thought. I would like to thank Willem deVries for pressing me to be clearer on this point.

the indispensable role of membership of a community in our self-conception as rational thinkers and agents, he also reworks this inheritance using cybernetic behaviorism.

The germinal seed of a Sellarsian account can be found in what he says about the conditions under which it would make sense to talk about the meaning of a machine state of the BBK robot. The states of the BBK robot, which picture its environment, are said to signify – to have meaning, to have content – insofar as we can *coordinate* our signifying behavior with its picturing behavior. Just as one can utter the English sentence "'*grun*' means *green*" to convey to an English speaker what the German speaker means by the German word "*grun*", we can also construct translation manuals for the BBK robot. We ascribe semantic content to the BBK robot to the extent that a translation manual can be constructed.[26]

To construct a translation manual, we need to be able to successfully notice what in our shared environment that the robot is responding to, classify the picturing states of the robot that are individuated by virtue of their causal role in the robot's sensorimotor feedback loops, and compare those states with our concepts as we use them in our social practices.[27] This triadic process – between us, the robot, and the environment – can be usefully conceptualized, following Davidson (1990; see also Davidson 1992), as a process of "triangulation". At the heart of Sellars's theory of meaning or conceptual content is what I will therefore call *triangulated cybernetic behaviorism*. The function of the ascription of semantic content is to facilitate triangulated cybernetic behavior: to construct a coordination device whereby we can say of our

[26] Though I believe the Quinean term "translation manual" is not inappropriate in discussing Sellars, there are two crucial and relevant differences: the Sellarsian translation manual is not constructed through stimulus-response pairs and it does not neglect the role of internal information processing.

[27] This is not to insist that all of the robot's states picture; if it has constructed cognitive maps that include abstract or theoretical terms, then those states are not themselves picturing, though they are required for picturing.

own linguistic behavior that it is similar enough to other linguistic behavior that the conditions for successful cooperation have been established.

But what, *in rerum natura*, is linguistic behavior? The Sellarsian answer is that linguistic behavior just *is* triangulated cybernetic behavior: when we ascribe semantic content to any utterance or inscription – even our own – is that it can be coordinated with other utterances or inscriptions that are functionally integrated into the sensorimotor feedback loops of other cybernetic systems, where the criteria of coordination lie in successful cooperation. How a cybernetic system interacts with its environment depends on how it models that environment, which means that cybernetic systems can cooperate only to the extent their models are sufficiently consistent that the actions guided by those models do not generate conflict. It is important to keep distinct the role of semantic attribution and the role of predictions and explanations of behavior. If you attribute to me the belief that geese are ducks, you are both making a claim about how I picture waterfowl and hence how I will engage with them and also making a claim that my picturing is incompatible with picturing based on sound scientific taxonomy.[28]

What we need at this point is an account of how we are to understand the relation between the manifest image of intentionality as embedded in our folk psychology and the scientific image of intentionality as explicated in cybernetic behaviorism, including triangulated cybernetic behavior. Though Sellars returns to this problem throughout his work, I want to focus on how he thinks about it at the same time as he is developing his theory of meaning in conjunction with cognitive/cybernetic behaviorism.

[28] Thanks to Willem deVries for the "Carl believes geese are ducks" example.

Sellars's claim that normative statements are logically irreducible to natural statements *and yet causally reducible* has provoked a good deal of commentary. As I see it, the crux of the argument depends on how Sellars understands meaning as functional classification. To say that mind is "logically irreducible" to body is to say *only* that the class of analytic truths does not include statements that stipulate an identity relation between statements made in folk psychological discourse and statements made in a suitably scientific discourse – cognitive behaviorism augmented by cybernetics. There is no equivalence of intension between statements made from within the intentional stance and statements made from within the cybernetic stance; no statement relating those statements could be true "by meaning alone".[29]

But if folk psychological statements and cybernetic behavioristic statements are *not* intensionally equivalent or synonymous, that nevertheless leaves open the possibility of co-extension. And this is not a possibility that Sellars rejects, though his acceptance of it takes a curious form: he says that folk psychological statements and cybernetic behavioristic statements "convey the same information." This should give us pause, because in 1953 the very idea of "information" as something that could be "conveyed" was just beginning to coalesce; Shannon's probabilistic definition of "information" was only published in 1948, five years earlier. John O. Wisdom's "The Hypothesis of Cybernetics" appeared in 1951, and that was one of the first philosophers to take up cybernetics.[30]

The relation between intentionality and cybernetic behavior is, however, slightly more complex than this suggests. From one perspective – that of the scientific image under

[29] Sellars never accepted Quine's critique of the analytic/synthetic distinction or the implications of that critique for ontological commitment, because Sellars – unlike Quine and, for that matter, Carnap – never thought that we should reject intensional semantics, though he was sympathetic to Morton White's view that the analytic and the synthetic had become an untenable dualism. Sellars's response to White is to repair the distinction, not reject it.

[30] John O. Wisdom, a philosopher of psychology, is not the same person as the ordinary language philosopher John Wisdom, though they were cousins.

construction – cybernetic behavior is the scientific image of thought. But if that were the end of the matter, what would become of the classical conception of intentionality as semantic content that is about the world in which it used? Sellars's answer to this question depends not just on his analysis of normativity but also on the role of that analysis in his understanding of community.

Sellars accepts and develops the classical German Idealist emphasis on the ineliminable normativity of rational thought and action: intentionality and normativity are logically interdependent. Cybernetic behaviorism cannot suffice for the scientific image of intentionality unless it can somehow accommodate this classical emphasis on normativity – and indeed, not just on normativity *simpliciter* but on the close tie between normativity and sociality. The rules of criticism that govern language-entry transitions (perceptions), formal and material inferences, and language-exit transitions (volitions) are, for us rational animals, interlocked with rules of conduct whereby we hold each other accountable for what we claim to perceive, think, and do.

To the extent that cybernetics could perhaps explain the rules of criticism or ought-to-be rules that govern the lives of non-rational animals, it would be in a weak or analogical sense – since those animals are (*ex hypothesi*) incapable of regarding themselves as governed by rules of criticism, it is we who regard them as being so governed, with the rules themselves being a consequence of how past natural selection has shaped the ways in which those animals occupy their niches (however plastic). In these cases, what we describe as rules of criticism are explained in terms of feedback across brain-body-environment causal loops, where the issuing of the rules functions as negative feedback to prevent behaviors that deviate too much from the rules.

Even if something like this were made plausible, Sellars would certainly accept that the story for us rational animals cannot be quite that simple, and that is because what distinguishes us *qua*

rational animals is not just that we can regard ourselves as governed by rules of criticism but also that we regard ourselves as being so governed by virtue of the interlocking relationship between rules of criticism and rules of conduct. When a little brown bat fails to capture a fleeing mosquito, it has done something that ought not be the case about what bats do – it has, in a broad sense, made a mistake – but it has not transgressed against Chiropteran social practices, for there are none. By contrast, when a person looks at an alligator and calls it a crocodile, they have not used the words correctly and are susceptible to correction from others.

This line of thought suggests that the scientific image of mind based on cybernetic behaviorism will be incomplete unless it can somehow accommodate not only rules of criticism but also rules of conduct. Without an account of normativity, the scientific image would be radically incomplete – it would not be a scientific image *of mind*. Thus, what we require here is an account that yokes together what Sellars says about the ineliminably normative dimension of human thought and action, based as it is on the philosophical clarification and elucidation of the manifest image, with what cybernetic behaviorism says about the scientific image of cognition.

In these terms, what are we to say about the ineliminable role of rules or norms in our linguistic and non-linguistic social practices? If the ascription of semantic content is to convey that the success (or failure) of triangulated cybernetic behavior, then the utterance or gestures of norms or rules that underpin meaning ascription are the behaviors that bring about that coordination. Rules or norms are ineliminable because rationality – or at least the human form of rationality -- is necessarily social.[31] Social life does not require perfect or ideal cooperation – at least not to a degree that would eliminate all conflict – but it requires that cooperation be, if not

[31] However, a correct Sellarsian reading of this point may require separating "the human" as normative concept from *Homo sapiens* as a biological concept; see Wolfendale (2019).

optimal, at least satisficing enough of the time for social life to be reproduced from one generation to the next.

With this element in place, we can finally draw out the following implication for a Sellarsian theory of personhood. In the concluding paragraphs of PSIM, Sellars remarks that to say of something – whether "a featherless biped or a dolphin or a Martian" (Sellars 1963c: 39) – that it is a person is to say that it is a member of one's community. It is to say that the naturalistic basis of community is triangulated cybernetic behaviorism: personhood is the status of a cybernetic system that actualizes a capacity to triangulate its behavior with other cybernetic systems that can also actualize their capacities for triangulated behavior. Triangulated cybernetic behavior is realized via the interlocking relation between rules of conduct and rules of criticism, such that it can say (or think) "I am one of you".

4. Sellars, Cognitivism, and Enactivism

A careful examination of the importance of cybernetics for Sellars's philosophy of mind has substantial implications for how we should assess his thought in light of contemporary cognitive science. This is because cybernetics is the ancestor of both cognitivism, with its emphasis on cognition as rule-governed manipulation of symbolic representations, and enactivism, with its emphasis on non-representational dynamic coupling between biologically autonomous systems and their environments. For this reason, I want to briefly explore how cybernetics came to influence both cognitivism and enactivism before indicating a place in Sellars's thinking where he could have re-oriented his ideas in a more enactivist direction than he actually did.

The rift between cognitivism and enactivism can be traced, according to Froese (2010), to the emergence of the split between computer science and second-order cybernetics. The decisive

issue turned on how cyberneticians responded to Ashby's demonstration that seemingly intelligent complex behavior could emerge from purely mechanistic assemblages. This called into question the presumptive realism at the heart of cybernetics as an objective science: if we ourselves are just machines turning 'noise' into 'meaning', then what are the rational credentials of any 'output' from the electrochemical computer called the brain?

This Ashbyian crisis, as Froese calls it, elicited two responses from the cybernetic community. The first response, which became computer science and cognitivism, remained committed to a realist epistemology and sought to implement it mechanistically by treating cognition as the mechanistic manipulation of symbolic representations. Consequently, there was no need to ground symbolic representations on anything more basic or mechanistic: the manipulation of symbols *was* cognition. The second response, which became second-order cybernetics and enactivism (among other paradigms), rejected the realism that defined cybernetics and instead embraced a constructivist epistemology. As Heinz von Foerester came to put it, the job of the brain is to compute an effective model of reality.

That the enactive and autopoietic approaches to cognition developed out of second-order cybernetics is relatively well-known (Froese 2010, Froese 2011). The basic idea of second-order cybernetics, in von Foerster's terms, turns on a shift from "observed systems" to "observing systems". In first-order cybernetics we are describing circular causality – recursion or feedback loops – in systems that we have built: we study them as objective components of the material universe. In second-order cybernetics we are describing circular causality in the systems that *we ourselves are.* But because we are persons, members of communities structured by relations (and asymmetries) of recognizing and being recognized, second-order cybernetics had to relinquish the commitment to pure objectivity: in becoming part of the conceptual structure in which we

experience and understand ourselves and others, it became necessary for the first-personal and second-personal perspectives to enrich the cybernetic vocabulary.

I want to now suggest a Sellarsian argument in support of second-order cybernetics. Briefly put, Sellars's criticism of the Given is best understood as rejecting the idea that how we experience the world can be decisively and clearly demarcated and protected against changes in our conceptual structure as a result of new discoveries in the empirical and formal sciences.[32] Despite his appreciation for phenomenology, Sellars rejects the Husserlian idea that the life-world can or should be defended against incursions by the sciences. Rather, for Sellars, the goal of joining the scientific and manifest images requires *incorporating* the sciences *into* the life-world – and this is precisely what second-order cybernetics does.

In other words, by incorporating cybernetic concepts into how he understood himself and others, von Foerster executed a Sellarsian strategy for incorporating scientific concepts into the world of everyday life. As Sellars puts it, "by construing the actions we intend to do and the circumstances in which we intend to do them in scientific terms, we *directly* relate the world as conceived by scientific theory to our purposes, and make it *our* world and no longer an alien appendage" (Sellars 1963c: 40). But we could not do this if the Given were not a Myth. For if the Given were not a Myth, there would be a clearly discernible stratum of our experience that would be unrevisable, come what may any changes elsewhere in our conceptual structure; it would not be possible to observe oneself or others as cybernetic systems.

This point can also be framed in terms of "the Myth of Jones" at the concluding sections of "Empiricism and the Philosophy of Mind" (Sellars 1963d: 183-196). There, Sellars imagines a group of people – "our Rylean ancestors" – who lack the concepts of *thought* and of *sensation*.

[32] See O'Shea (2021) as to why a commitment to the Given is a commitment to holding that cognitive experience has a categorial structure that is unrevisable, come what may.

The concepts are invented by a mythical "Jones" who one day, puzzling over certain behaviors – that people act as if they had been talking to themselves but without saying anything aloud, or that people act as if they are seeing or hearing things that no one else sees or hears – comes up with the concepts of thought and of sensation by analogy. People have thoughts that are like overt verbal episodes, except that no one can hear or see them; and they are sensations that are like the sensible qualities of physical things, except that they only exist for that person, in their 'consciousness'. Over time, Jones teaches these innovations to others, so that what began as a theoretical posit becomes part of our non-inferential awareness of self and others. Transposing this lesson from thoughts and sensations to cybernetics, we can say that Heinz von Foerster was the 'Genius Jones' of cybernetics.[33]

Thus far I have only suggested that the transition from first order to second order cybernetics is one that makes sense philosophically in light of the criticism of the Myth of the Given. But I think that we also need to take careful notice of two further considerations: how the concepts of second-order cybernetics become transformed as a consequence of being incorporated into the lifeworld, of first-person and second-person linguistic performances, and how this transformation affects our reading of Sellars himself.

On the first point: first-order cybernetics depended essentially on abstracting away from the differences between organisms and machines, in order to produce abstract concepts like "information" and "feedback". Neglecting the material reality of physics, chemistry, and biology was necessary for producing the abstract models that the cyberneticians analyzed and debated. For example, when McCulloch and Pitts (1943) demonstrated that recurrent networks of neuron-like elements can realize Boolean functions, they explicitly introduced the simplifying

[33] I put it this way largely for rhetorical effect – I do not intend to slight the contributions of Maturana, Pask, Beer, Varela, Bateson, Mead, Thompson, and many others.

assumption that each neuron either does or not fire. They constructed an abstract model, a digital neuron. The fact that their model did not take into account biological reality, where neurons are modulating their activity in ways that are much more "analog" than "digital", was not among their concerns. Likewise, Wiener deliberately chose as the subtitle of his manifesto "control and communication in the animal and in the machine" – with the assumption that there is no difference that makes a difference between animals (including us) and machines (including computers).

This simplifying abstraction, however important for allowing the conceptual and empirical breakthroughs of first order cybernetics, could not be maintained when cybernetics became incorporated into how we understand ourselves: the abstract had to become concrete. It is for this reason that second order cybernetics rather quickly evolved into autopoiesis theory: a formal model about the specific kinds of organizational features that a complex system must have in order to be described as "alive". Autopoiesis theory and related approaches, such as those of Robert Rosen, Stuart Kauffman, Alvaro Moreno, Matteo Mossio, and Ezequiel di Paolo, have generated a rich and sophisticated way of understanding why organisms are *not* machines and cognition is *not* computation – *contra* both first-order cybernetics and the cognitivist research program that it also gave rise to.

On the second point: this reconstruction of the transition from first order cybernetics to second order cybernetics also matters for Sellars, because as I have argued here, his own philosophy of mind is deeply indebted to first order cybernetics. Sellars is (perhaps) the first computational functionalist in philosophy of mind, because of how he incorporates cybernetics into his understanding of cognitive systems. But, somewhat ironically, Sellars himself does not take von Foerster's step of incorporating cybernetics directly into the life-world: a step that

Sellars should have taken given the larger shape of his thought, and perhaps one that he would have taken if this inconsistency had been pointed out to him.

The fact that Sellars did not take this step has had further repercussions for post-Sellarsian philosophy – that is, philosophy that takes itself to be building upon Sellars's considerable achievements. Because Sellars himself did not take the von Foerster step of incorporating cybernetics into his experience of himself and the world, he did not question the realist epistemology that first order cybernetics, like all modern objective thought, took for granted. For Sellars, the goal of science is to construct testable models of the fully determinate regularities that exist in a fully mind-independent sense, and which the philosopher can use to tell us which aspects of the phenomenal world are truly mind-independent and which are not. These commitments also shape how post-Sellarsian philosophers, such as Daniel Dennett, Paul Churchland, and Robert Brandom, took up the legacy (in very different ways) of scientific realism and computational functionalism. These philosophers belong to the legacy of first-order cybernetics and cognitivism because they continue Sellars's refusal to take the step towards second-order cybernetics, even though Sellars himself could have done so – and arguably should have. Sellars's cybernetic behaviorism is grounded in his appropriation of first-order cybernetics, but in light of the rift between cognitivism and enactivism, it is questionable whether first-order cybernetics was a dialectically stable position. As I see it, Sellars's overarching project should have led him to adopt second-order cybernetics – though without necessary abandoning his commitment to representationalism, despite the anti-representationalism that has become definitive of enactivism.[34]

[34] A different route from Sellars to contemporary autonomy or enactive theories could be traced by considering Sellars's own project of "naturalizing Kant" in light of the Third Critique account of teleological judgment; see Weber and Varela (2007).

5. Conclusion

It has become something of a commonplace that Sellars's complicated image of humanity in the universe belongs to the tradition of philosophers who aspired to naturalize Kant. The extent to which Sellars succeeded in doing so has been obscured by confusion about what he meant by "picturing" – or even if we need it at all. I have tried to show that Sellars's concept of picturing had been difficult to understand due to ignorance of his historical context. With the proper context in place, we can see that Sellars drew upon Tolman's cognitive behaviorism and Wiener's cybernetics to transform Wittgensteinian picturing into a cybernetic-behavioral concept. In doing so, Sellars conceptualized picturing in terms of feedback loops as being described and built by the first generation of cyberneticists. It is because Sellars's use of cybernetics was not even fully appreciated in Sellars's own time, and has been completely forgotten since, Sellars's distinct version of naturalizing Kant has not yet received the full treatment that it merits.[35]

Works Cited

Akins, Kathleen, 1996, "Of Sensory Systems and the 'Aboutness' of Mental States" in *Journal of Philosophy*, 93,7: 337-372.

Baars, Bernard, 1986, *The Cognitive Revolution in Psychology*, New York, The Guilford Press.

Davidson, Donald, 1990, "Epistemology Externalized" in Davidson 2001.

Davidson, Donald, 1992, "The Second Person" in Davidson 2001.

[35] I would like to thank Willem de Vries, Bryce Huebner, and Kyril Popatov for their detailed and encouraging criticisms of previous versions of this paper, and comments from Evan Thompson on §4. I would also like to thank Steve Levine for our many conversations about Sellars over the years.

Davidson, Donald, 2001, *Subjective, Intersubjective, Objective*, Oxford, Oxford University Press.

Friedman, Michael, 2001, *Dynamics of Reason*, Stanford, Center for the Study of Language and Information.

Froese, Tom, 2010, "From cybernetics to second-order cybernetics: A comparative analysis of their central ideas," in *Constructivist Foundations*, 5,2: 75-85.

Froese, Tom, 2011, "From Second Order Cybernetics to Enactive Cognitive Science: Varela's Turn from Epistemology to Phenomenology" in *Systems Research and Behavioral Science*, 28, 6: 631-645.

Haugeland, John, 1998, "The Intentionality All-Stars" in *Having Thought: Essays in the Metaphysics of Mind*, Harvard University Press, Cambridge MA, 127-170.

Hook, Sidney, 1943, "The New Failure of Nerve" in *Partisan Review*, 10,1: 2-23.

Huebner, Bryce, 2018, "Picturing, Attending, and Signifying" in *Belgrade Philosophical Annual* 31: 7-40.

Kline, Ronald, 2017, *The Cybernetics Moment: Why We Call Our Age the Information Age.* Johns Hopkins University Press, Baltimore.

Kusch, Martin, 1995, *Psychologism; A Case Study in the Sociology of Philosophical Knowledge*, New York, Routledge.

McCulloch, Warren and Walter Pitts, 1943, "A Logical Calculus of the Ideas Immanent in Nervous Activity" in *The Bulletin of Mathematical Biophysics* 5: 115-133.

McDowell, John, 1998, "Intentionality is a Relation" in *Journal of Philosophy* 95, 9:471-491

Olen, Peter, 2016, *Wilfrid Sellars and the Foundations of Normativity*, New York, Palgrave.

O'Shea, James, 2021, "What *is* the Myth of Given?" in *Synthese* https://doi.org/10.1007/s11229-021-03258-6

Rosenblueth, Arturo, Norbert Wiener, and Julian Bigelow, "Behaviour, Purpose and Teleology" in *Philosophy of Science* 10: 18-24.

Rowlands, Mark, 2010, *The New Science of the Mind*, Cambridge, The MIT Press

Sachs, Carl, 2019, "In Defense of Picturing: Sellars's Philosophy of Mind and Cognitive Neuroscience" in *Phenomenology and the Cognitive Sciences*, 18,4: 669-689.

Sachs, Carl, 2020, "A Conceptual Genealogy of the Pittsburgh School: Between Kant and Hegel" in *The Cambridge History of Philosophy, 1945-2010.* Ed Kelly Becker and Iain Thompson, Cambridge, Cambridge University Press.

Sellars, Wilfrid. 1980. "Language, Rules, and Behavior" in *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, ed Jeffrey Sicha, Ridgeview Publishing Company.

Reprinted from 1950, "Language, Rules, and Behavior" in *John Dewey: Philosopher of Science and of Freedom*, ed. Sidney Hook. New York: Barnes and Noble.

----------, 1963a, "Being and Being Known" in *Science, Perception, and Reality*, Atascadero, Ridgeview Publishing Company, 41-59.

----------, 1963b, "Some Reflections on Language Games" in *Science, Perception, and Reality*, Atascadero, Ridgeview Publishing Company, 321-358.

----------, 1963c, "Philosophy and the Scientific Image of Man" in *Science, Perception, and Reality*, Atascadero, Ridgeview Publishing Company, 1-40.

----------, 1963d, "Empiricism and the Philosophy of Mind" in *Science, Perception, and Reality*, Atascadero, Ridgeview Publishing Company, 127-196.

Shapiro, Lionel, 2011, "Intentional Relations and the Sideways-On View: On McDowell's Critique of Sellars" in *European Journal of Philosophy*, 21,2: 300-319.

Tolman, Edward, 1932, *Purposive Behavior in Animals and Man*, New York, The Century Company.

Tolman, Edward, 1948, "Cognitive Maps in Rats and Men" in *Psychological Review* 55,4: 189-208.

Weber, Andreas and Francisco Varela, 2002, "Life After Kant: Natural purposes and the autopoietic foundations of biological individuality" in *Phenomenology and the Cognitive Sciences* 1: 97-125.

Wiener, Norbert, 1948, *Cybernetics, or Communication and Control in the Animal and in the Machine*, Cambridge, The MIT Press.

Wisdom, John, 1951, "The Hypothesis of Cybernetics" in *British Journal for the Philosophy of Science* 2,5:1-24.

Wittgenstein, Ludwig, 1974, *Tractatus Logico-Philosophicus*, trans David Pears and Brian McGuinness, New York, Routledge.

Wolfendale, Peter, 2019, "The Reformatting of *homo sapiens*" in *Angelaki* 21,1:55-66.