

H U**M** A**N** A **M** E**N**T**E**

Between Two Images

The Manifest and Scientific Conceptions
of the Human Being,
50 Years On

EDITED BY
Carlo Gabbani

ISSUE **21** - MAY 2012



Edizioni ETS

EDITORIAL MANAGER: **DUCCIO MANETTI** - UNIVERSITY OF FLORENCE
EXECUTIVE DIRECTOR: **SILVANO ZIPOLI CAIANI** - UNIVERSITY OF MILAN
VICE DIRECTOR: **MARCO FENICI** - UNIVERSITY OF SIENA

Editorial
Board

INTERNATIONAL EDITORIAL BOARD

JOHN BELL - UNIVERSITY OF WESTERN ONTARIO
GIOVANNI BONIOLO - INSTITUTE OF MOLECULAR ONCOLOGY FOUNDATION
MARIA LUISA DALLA CHIARA - UNIVERSITY OF FLORENCE
DIMITRI D'ANDREA - UNIVERSITY OF FLORENCE
BERNARDINO FANTINI - UNIVERSITÉ DE GENÈVE
LUCIANO FLORIDI - UNIVERSITY OF OXFORD
MASSIMO INGUSCIO - EUROPEAN LABORATORY FOR NON-LINEAR SPECTROSCOPY
GEORGE LAKOFF - UNIVERSITY OF CALIFORNIA, BERKELEY
PAOLO PARRINI - UNIVERSITY OF FLORENCE
ALBERTO PERUZZI - UNIVERSITY OF FLORENCE
JEAN PETITOT - CREA, CENTRE DE RECHERCHE EN ÉPISTÉMOLOGIE APPLIQUÉE
CORRADO SINIGAGLIA - UNIVERSITY OF MILAN
BAS C. VAN FRAASSEN - SAN FRANCISCO STATE UNIVERSITY

CONSULTING EDITORS

CARLO GABBANI - UNIVERSITY OF FLORENCE
ROBERTA LANFREDINI - UNIVERSITY OF FLORENCE
MARCO SALUCCI - UNIVERSITY OF FLORENCE
ELENA ACUTI - UNIVERSITY OF FLORENCE
MATTEO BORRI - UNIVERSITÉ DE GENÈVE
ROBERTO CIUNI - UNIVERSITY OF DELFT

**SCILLA BELLUCCI, LAURA BERITELLI, RICCARDO FURI,
ALICE GIULIANI, STEFANO LICCIOLI, UMBERTO MAIONCHI**

Editorial
Staff

TABLE OF CONTENTS

INTRODUCTION

Carlo Gabbani <i>Between Two Images? An Introduction</i>	V
---	---

PAPERS

Willem A. de Vries <i>Ontology and the Completeness of Sellars's Two Images</i>	1
P. Kyle Stanford <i>The Eyes Don't Have It: Fracturing the Scientific and Manifest Images</i>	19
David Hodgson <i>Identifying and Reconciling Two Images of "Man"</i>	45
Steven F. Savitt <i>Of Time and the Two Images</i>	57
Keith Lehrer <i>The Unity of the Manifest and Scientific Image by Self-Representation</i>	69
Giacomo Turbanti <i>Self-Normativity and the Realist Stance in Semantics</i>	83
Jay L. Garfield <i>Sellarsian Synopsis: Integrating the Images</i>	103
Owen Flanagan, Stephen Martin <i>Science and the Modest Image of Epistemology</i>	123
James R. O'Shea <i>Prospects for a Stereoscopic Vision of our Thinking Nature: On Sellars, Brandom, and Millikan</i>	149
Diego Marconi <i>Quine and Wittgenstein on the Science/Philosophy Divide</i>	173

COMMENTARIES

Massimo Marraffa <i>The Bounds of Sense</i> by Peter F. Strawson	191
--	-----

Raffaella Campaner
Scientific Explanation and the Causal Structure of the World
by Wesley Salmon 197

Gabriele De Anna
*Putnam on the Fact/Value Dichotomy and the scientific conception
of the world* 205

BOOK REVIEWS

Luca Corti
*Empiricism, Perceptual Knowledge, Normativity, and Realism:
Essays on Wilfrid Sellars*
by Willem A. de Vries (Ed.) 213

Between Two Images? An Introduction

Carlo Gabbani *
c.gabbani@tiscali.it

1. A Tale of Two Images

The relationship between common-sense representations of man and the world and scientific representations of them were widely debated in XXth century culture. This, of course, largely depends on the increasing and systematic development of scientific-experimental knowledge which now ranges over a huge amount of phenomena.

What makes this issue especially awkward is the fact that these two accounts do not seem to harmonize or be easily integrable in a unitary conception. Rather they convey two very different, and seemingly opposite, worlds: Eddington's more than famous "two tables" have become the icon of this diversity. If this is true and relevant when we deal with *objects*, it is even more true and relevant when *subjects* are concerned, when the scientific-experimental methods which were created for the study of nature from an "objective" point of view are then applied to the study of the conscious minds of persons, for instance.

This is not simply a theoretical issue; the way we describe and explain the world and man have a deep influence on the kind of person we eventually become. The understanding of the world and ourselves in fact plays an essential role in the shaping of our identity, and, as Arnold Gehlen once wrote (in a passage that in a certain measure anticipates Sellars): «there is a living being, one of whose most significant characteristics is the need for self-explanation, for which an "image", an interpretative formula, is necessary» (Gehlen, 1940/1988, p. 4).

Fifty years ago Wilfrid Sellars's essay *Philosophy and the Scientific Image of Man* was first published.¹ It is a classic analysis of this problem and

* University of Florence, Italy.

¹ Sellars (1962). The essay appeared as the second chapter in *Frontiers of Science and Philosophy*, the first volume of the *University of Pittsburgh Series in the Philosophy of Science*. It was then published (PSIM), the following year, in Sellars's *Science, Perception and Reality* (1963). The early

articulated very influential categories for dealing with it, while at the same time offering a much discussed interpretation and controversial answer to it.² Sellars's essay «consists of two lectures given at the University of Pittsburgh in December, 1960, as part of a series of lectures in the history and philosophy of science by various contributors» (1963, p. vii).³

Sellars's essay begins with a famous definition of philosophy that has been endorsed by many philosophers (see for instance: Putnam, 2012, ch. I) and that essentially contributes to explaining why philosophy is fully entitled to deal with the problem of the relationship between scientific and non-scientific representations of man and the world: «The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term» (PSIM, p. 1).

It is common knowledge that, according to Sellars, philosophers are confronted today «by *two* pictures of essentially the same order of complexity, each of which purports to be a complete picture of man-in-the-world, and which, after separate scrutiny, he must fuse into one vision» (PSIM, p. 4). The first picture is the «'manifest' image of man-in-the-world» (PSIM, §II), I will point out here three of the features with which Sellars characterizes it:

draft of this text and the handwritten notes concerning it are now at the Pittsburgh Archives: Wilfrid S. Sellars Papers, 1899-1990, ASP.1991.01, Archives of Scientific Philosophy, Special Collections Department, University of Pittsburgh. Collection Inventory: Section: Lectures; Box: 16; Folder: 03–07 “Philosophy and the Scientific Image of Man,” c. 1962. They are also freely accessible on-line: <<http://www.library.pitt.edu/libraries/special/asp/sellars.html>> (accessed on: May 13th 2012). In *Frontiers of Science and Philosophy* Sellars's essay was preceded (p. 35) by an excerpt from Ernst Cassirer's *An Essay on Man*.

² The essay has been translated into many languages. Some examples: Spanish translation by V. Sánchez de Zavala: *La Filosofía y la Imagen Científica del Hombre*, in: J. Muguerza (ed.), *La Concepción Analítica de la Filosofía*, Alianza Editorial, Madrid 1981, pp. 645–691; French translation by Y. Bouchard and D. Boucher: *La philosophie et l'image scientifique de l'homme*, in: D. Fiset-P. Poirier (eds.), *Philosophie de l'esprit. Tome I: Psychologie du sens commun et sciences de l'esprit*, Vrin, Paris 2002, pp. 55–115; Italian translation by A. Gatti: *La filosofia e l'immagine scientifica del mondo*, Armando Editore, Roma 2007.

³ The other contributors of this series were: C. G. Hempel, M. Scriven, E. Caspari, A. Grünbaum, P. K. Feyerabend, and E. Nagel (all their essays, except Nagel's one, were published along with Sellars's paper in *Frontiers of Science and Philosophy*). This Annual Lecture Series was a first step towards the establishment of the (now famous) *Center for Philosophy of Science* of the University of Pittsburgh (cf.: <http://www.pitt.edu/~pittcntr/About/history/history_1.htm> – accessed on: May 13th 2012). The University of Pittsburgh was then to be Sellars's own University from 1963 to his retirement.

- «the conceptual framework which I am calling the manifest image is, in an appropriate sense, itself a scientific image. [...] There is, however, one type of scientific reasoning which it, by stipulation, does *not* include, namely that which involves the postulation of imperceptible entities, and principles pertaining to them, to explain the behaviour of perceptible things», the “manifest image”, therefore, «limits itself to what correlational techniques can tell us about perceptible and introspectible events» (PSIM, pp. 7 and 19);
- «primary objects of the manifest image are *persons*», the manifest image, indeed, is seen as a «refinement of the ‘original’ image of man-in-the-world» (intended, in turn, as «a framework in which *all* the ‘objects’ are persons»), through a «gradual ‘de-personalization’ of objects other than persons» (PSIM, p. 10; cf. p. 12);
- «man is that being which conceives of itself in terms of the manifest image. To the extent that the manifest does not survive in the synoptic view, to that extent man himself would not survive» (PSIM, p. 18).

The second picture is the «scientific image of man-in-the-world» (PSIM, §IV) and also in this case it is worth underlining three of the main aspects of it:

- it obviously contrasts with the manifest image because it «postulates imperceptible objects and events for the purpose of explaining correlations among perceptibles» (PSIM, p. 19);
- it has to be interpreted, according to Sellars, in a non-instrumentalist way: «systems of imperceptible particles» introduced by the “scientific image” are *not* considered as «‘abstract’ or ‘symbolic’ ways» of representing manifest objects (see: PSIM, p. 26 and 32);
- «although the image is *methodologically* dependent on the world of sophisticated common sense, and in this sense does not stand on its own feet, yet it purports to be a *complete* image, i.e., to define a framework which could be the whole *truth* about that which belong to the image [...] the scientific image presents itself as a *rival* image» (PSIM, p. 20).

The aim of philosophy is to overcome this fragmentary situation, reaching the «eye on the whole» through a «fusion», a «synoptic vision», a «stereoscopic vision», a kind of «binocularity» (on these metaphors see also Stanford, *infra*, §2).

Sellars notoriously confers a kind of primacy, *as far as describing and*

explaining are concerned, to the scientific image: from this point of view, an ideal, accomplished scientific image will thus be the best inventory of what really exists in our world, and the best explanation of how it works.

The manifest image, on the contrary, does not adequately describe reality. But neither is it entirely wrong and massively nonreferential. Indeed, the main aspects, features, and phenomena that belong to the manifest image must be accounted for in the scientific image with scientifically well-grounded *successor concepts* of those applied in the manifest image. There are intentional states or «raw feels» as the manifest image tells us, but their ultimate nature is not such as it is represented within that image. So, the scientific image must provide us with a scientific conception of them which is not a kind of elimination.

Regarding the intentional states (and the real «mind-body problem») Sellars (PSIM, §VI) seems confident that the pathway towards an identification of conceptual states with objects of the scientific image is open: «if thoughts are items which are conceived in terms of the roles they play, then there is no barrier *in principle* to the identification of conceptual thinking with neurophysiological process» (PSIM, p. 34; but on this point, see deVries's essay in this volume).

On the contrary, as far as raw feels (and the so-called «sensorium-body problem») are concerned, things are more difficult from Sellars's perspective, because it seems impossible to him to reconcile the «ultimate homogeneity» of the qualitative states with the «particulate» foundation of the present scientific image. A scientific account of raw feels is not impossible, but it requires a conceptual refoundation of the scientific image – a very demanding project that Sellars himself will endorse in the last period of his career.

But within the framework of the manifest image we do not simply describe and explain things, we also find in that image «categories pertaining to man as a *person* who finds himself confronted by standards (ethical, logical, etc.)» (PSIM, p. 38; cf. p. 6). The conceptual framework of our normative notions, standards, rules and intentions, in the various fields of our discourse and rational practice as members of a group, is not something that, from Sellars's perspective, can be reduced to our scientific conceptual framework; rather, it should be «joined» (PSIM, p. 40) to the scientific image in order to reach a synoptic view – and «from this point of view, the irreducibility of the personal is the irreducibility of the ought to is» (PSIM, p. 39). One might then see

Sellars as a naturalist without a “naturalized epistemology”⁴, or, as O’Shea (2007) put it, as the proponent of a «naturalism with a normative turn».⁵

2. The Conceptual Framework of Sellars’s Analysis: Objections and Presuppositions

It is probably useful to recollect some objections that have been raised against the very categories or the very structure of Sellars’s account. Here, I will only present three main critical aspects of Sellars’s analysis: the first concerns the very concept of an “image”, the second the scientific image, the third the manifest one. I do not want to worry about whether they are effective or inappropriate towards Sellars’s account, I will only say that certain formulations of them may find an immediate answer in Sellars’s original text, while other formulations and arguments seem to represent *at least* serious objections to his vision and should be discussed further.

2.1. Imaginary Images?

The first point immediately concerns Sellars’s use of the term “image”. While image metaphors have been widely used in 20th century culture in order to refer to the scientific conception of the world, the adequateness of this choice has been criticized in many ways.

In a certain way, this kind of criticism already existed before Sellars’s analysis was conceived. Moritz Schlick, for instance, wrote in 1925:

[T]he expression ‘world picture’ is itself not the best one to use; it would be preferable to say ‘world concept’. For in philosophy the world ‘picture’ is better confined to the intuitively representable, whereas the physical representation of the world, although conceptual, is entirely non-intuitive. (1925/1974, §32, p. 294)⁶

⁴ The project of a «naturalized epistemology» may have many forms and meanings. I suppose that Sellars’s philosophy is incompatible with strong forms of naturalized epistemology such as those characterized by Susan Haack as «scientific» naturalism, which «make epistemology an enterprise internal to the natural sciences» and may assume a «reformist» or a «revolutionary» aspect (1993, ch. VI, p. 119).

⁵ This is the subtitle of his book on Sellars’s philosophy, where the discussion on the two images, introduced in chap. I, is the basis of the structure of the entire volume. The analysis of PSIM is the starting point also for deVries (2005, ch. I).

⁶ And Herbert Feigl writes: «I would prefer to contrast the manifest image with the scientific

Sellars does not ignore that science is not a matter of pictures which can easily be imagined, but, rather, one of conceptual accounts and he explicitly says: «I'm using 'image' as a metaphor for conception» (PSIM, p. 5). Indeed, he is certainly not the kind of philosopher who confuses what has a conceptual status with what doesn't.⁷ Nevertheless, it is important to analyse the role and influence of this metaphor in Sellars's essay, and even more to discuss the idea of two main conceptual matrixes which shape our view of the world and ourselves.⁸

2.2. The Scientific Image: One, Many, Practically None

Firstly, there is a deep problem concerning the idea of *an* image, and especially the idea (even if metaphorical) of *a* scientific image. It is worth remarking that the "image" Sellars speaks about is explicitly intended as an «idealization» and as an «ideal type» in Weber's sense.

Nevertheless, many philosophers have remarked that such a thing as *the* scientific image does not exist (and will probably never exist); rather, what we have are many different and partial scientific theories, concerning different types of phenomena, which operate on different levels and use different instruments, procedures and conceptual frameworks. Therefore, the idea of a unified scientific image would resemble a kind of neopositivistic relic within Sellars's philosophy. From this point of view, it has also been asserted that we should not look for one, single attitude towards all scientific theories and, accordingly, we should not expect there to be one, single solution concerning the relationship between science and common-sense perspectives (see Stanford's essay in this volume).

In fact, Sellars himself addresses a similar problem in his essay, speaking of

conception of the world. More strongly than ever before, I am convinced that it is primarily the concept of the "physical" that requires reinterpretation and reconstruction. The imagery that is so helpful heuristically and didactically is not and cannot be part of the cognitive meaning of physical concepts and hypotheses» (1967, p. 142).

⁷ In EPM (§47) Sellars speaks of «the classical tradition» on thought and he writes of «a number of confusions, perhaps the most important of which was the idea that thoughts belong in the same general category as sensations, images, tickles, itches, etc.».

⁸ From this point of view it is worth remarking that the title of this Issue expresses the theme around which it is built and the categories that characterise Sellars's analysis, but their adequacy and adoption is not presupposed in any way. At the same time, as deVries remarked, «this distinction has now taken on a life of its own, although the terms are not always used in accordance with Sellars's original intention» (2005, p. 9).

«the constitution of *the* scientific image from the several scientific images of which it is the supposed integration» (PSIM, p. 21). It is unlikely that Sellars's strategy will be convincing for those who endorse this line of criticism, but his point is that it is certainly possible to acknowledge a multiplicity of experimental sciences and their methodological differences: he does not want «to equate the sciences, for as sciences they have different procedures and connect their theoretical entities via different instruments to intersubjectively accessible features of the manifest world» (p. 21). But this kind of pluralism, from his point of view, would not preclude the possibility of an ontological identification of the different objects the various sciences speak about and the reduction of the higher level objects to aggregates of basic objects («the objects of biochemical discourse can be equated with complex patterns of the objects of theoretical physics» (p. 21); concerning the reasons for this identification, see also below the “principle of reducibility”).

The issue seems to me still open: if Sellars's proper view on the unifying of «*some* of the ‘partial’ images into one image» (p. 21) seems at least highly problematic and dependent on very controversial metaphysical underpinnings, its failure does not directly imply the impossibility to make sense of a *weaker* idea of a scientific image, intended as one general, common matrix that generates some *structural commonalities* among the different sciences or scientific theories, and contributes to determining their mutual relationships. Otherwise, why do we designate all those sciences and theories with *one* adjective such as “scientific”, “experimental”, “naturalistic” and the like? From this perspective, the existence of an ideal conceptual framework and matrix that they all belong to could be what would allow us to speak of a scientific image as an ideal type (but in a sense which is no longer that of Sellars, because the unification depends on epistemological commonalities, and not on ontological identifications).

The other problem that, in any case, still remains completely overt is that of interpreting the status, significance and ontological import of this “scientific image”. This obviously depends mainly on the status of scientific realism, but, even for those who adopt scientific realism, there are relevant problems to face here, both on a diachronic and on a synchronic level. On a diachronic level, a relevant aspect is the fact that each generation sees «the Scientific Image inherited from the older generation as open, vague, ambiguous in the light of our new understanding (that is: in the light of alternatives not previously conceived)» (van Fraassen, 1999, p. 36). But also on a synchronic level, a

problem of interpretation of the very ontological commitments implied by accepted scientific theories seems to arise:

How can we ask our fundamental physical theories to tell us about what there is in the world when each of those theories is subject to multiple interpretations, interpretations that often radically disagree with one another about what kind of a world the fundamental theory is really describing? (Sklar, 2001, p. 47)

2.3. The *Truth* About the Manifest Image

As we have seen, Sellars maintains that «the conceptual framework which I am calling the manifest image is, in an appropriate sense, itself a scientific image» (PSIM, p. 7). This seems to confer a theoretical status to the manifest image and its terms. For this reason (and especially because of the analysis proposed in EPM) Sellars is generally also regarded as one of the forerunners of the “theory view” on common-sense psychology, a fundamental component of the manifest image. On this view, common-sense psychology constitutes a body of knowledge with theoretical status and explanatory aims, and the referents of its “mentalist” terms are our mental states individuated on the basis of the functional roles they play. This common-sense psychology is therefore a kind of naive theory called “folk psychology” and can be confronted with science.

This point is extremely relevant also because various forms of reductionism or eliminativism overtly influenced by Sellars (see below §3) are essentially based on the thesis that our mentalistic and manifest image is a (science-like) theory, and for this very reason can also be confronted with/reduced to/preserved within/eliminated by succeeding scientific theories. At the same time, the theoretical status of the manifest image seems to represent a complex and problematic issue and deVries (2006, §3), for instance, has emphasized also several «disanalogies» between theoretical concepts and folk psychological concepts in Sellars’s thought.⁹ Here, it is worth remarking above all that the “scientific” status of our common-sense psychology does *not* imply an *overall* «homogeneity» (cf.: Elton, 2003, pp. 103–105) between itself and the scientific image. The manifest image, indeed, also includes those aspects concerning evaluations, norms, and standards that are in principle outside the

⁹ deVries (2006) has argued that for several reasons «There is [...] no sense of “theory” in which Sellars would have conceded that folk psychology is an eliminable theory, not even as a consequence of his claim that the manifest image is ultimately to be superseded by the scientific in matters ontological» (p. 67). On this issue cf. also: Garfield (1988) and (1989). See also footnote 18.

range of the scientific image. And this dishomogeneity is exactly the reason why we should “*join*” «the conceptual framework of persons» (PSIM, p. 40) to the scientific image. Thus, it seems that not even an ideal scientific image could substitute or eliminate *those* concepts; and, at the same time, they seem to constitute a very relevant component of our common-sense psychology.

Nevertheless, *as far as describing and explaining are concerned*, it is also possible to compare the manifest and the scientific image, as the former also actually includes truth-apt descriptions and explanations and, thus, constitutes a corpus of empirical beliefs. Ultimately, however, in the dimensions of description and explanation, it is likely, from a Sellarsian point of view, that the manifest image (folk psychology included) turns out to be a false image, and will not survive – «*there is truth and error with respect to it even though the image itself might have to be rejected, in the last analysis, as false*» (PSIM, p. 14).

It is then clear that the status of the manifest image, and especially that of common-sense psychology within it, requires further reflections and many questions still appear to be open, both on an exegetical and on a theoretical level. For instance: in what sense and to what extent can the manifest image really be considered theoretical? What is the relationship between the explanatory and the normative roles of their categories? And, if common-sense psychology were also a theoretical component of the manifest image, what kind of theory is it? A scientific theory, or not? Must it be interpreted in a functionalist way, or not? Should we rather understand the theoretical status of folk-psychology in a weak and broad sense: simply being a truth-apt propositional corpus? Again: what is the impact of “simulationism” on the debate concerning realism and antirealism in common-sense psychology? And, when we discuss the theoretical status of folk psychology, are we adopting an *internalist*, or an *externalist* reading of it (see: Ravenscroft-Stich, 1994)?

It is worth briefly remarking also on the (explicit) assumptions that lie behind Sellars’s interpretation of the relationship between the two images, which make his diagnosis and answer to the threat of a «clash» plausible. I would like to underline at least three main relevant and controversial options:

- 1) *scientific realism*. «As I see it, to have a good reason for holding a theory is *ipso facto* to have a good reason for holding that the entities postulated by the theory exist» (PHM, p. 91). That is to say that scientific theories are capable of truth value and if we regard them as

good theories and accept them, we should also believe in them as *approximately* true, maintaining that the (so-called) non-observation terms really refer to imperceptible entities that exist in our world. Sellars, therefore, refuses an instrumentalist or empiricist interpretation of the ontological commitments of accepted scientific theories. Scientific objects do not only exist, but, according to Sellars, at least *in principle* (that is, considering an ideal, accomplished scientific image), «it is ‘scientific objects’, rather than metaphysical unknowables, which are the true things-in-themselves» (Sellars, 1968, ch. V, §79, p. 143; while, if these Kantian categories are used, «the world of common sense is a ‘phenomenal’ world»¹⁰).

- 2) *the principle of reducibility*: this principle «makes impossible the view that groups of particles can have properties which are not ‘reducible to’ the properties and relations of the members of the group» (PSIM, p. 35; cf.: pp. 21 and 27; in PSIM this principle is «accepted without argument»); this ontological assumption implies that to consist of micro-physical particles also means to be entirely reducible to these basic constituents. *Strong* forms of *emergentism* are, accordingly, excluded within the scientific image (but on Sellars’s own emergentism, see below). At the same time, a minimal kind of *intrascientific* pluralism is allowed *as far as* procedures, methodologies etc. are concerned. This principle also contributes, from Sellars’s perspective, to making it extremely difficult to describe and explain sensible qualities in our *present* scientific categorial framework (where they are candidates for identification with groups of particles).

¹⁰ «If [...] we replace the static conception of Divine Truth with a Peircean conception of truth as the ‘ideal outcome of scientific inquiry’, the gulf between appearances and things-in-themselves, though a genuine one, can in principle be bridged» (1968, ch. II, §51, p. 50); but on the “phenomenality” of the manifest image see: deVries (2005, pp. 157–161, and 269–271). On this point John McDowell has noted: «Sellars reads Kant as a scientific realist *manqué*; in Sellars’s view, had Kant only been sophisticated about the possibilities for scientific concept-formation, he would have cast the objects of the scientific image in the role of things in themselves. But for Kant, objects as they appear in the scientific image would be just another case of objects as they appear, with a transcendental background for that conception just as necessary here as anywhere. Sellars’s attempt to be responsive to Kantian transcendental concerns goes astray in his idea that an appeal to science could do the transcendental job; here Sellars’s scientism is seriously damaging» (2009a, p. 42, n. 30).

- 3) “*scientia mensura*”: «[S]peaking as a philosopher, I am quite prepared to say that the common sense world of physical objects in Space and Time is unreal – that is, that there are no such things. Or, to put it less paradoxically, that in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not» (EPM, IX, §41).¹¹ Sellars’s famous paraphrase of Protagoras’s *dictum* shows that he not only endorses scientific realism but he also sees no limits, *in principle*, for science; that is to say that there are no phenomena, events or entities that are *in principle* outside the scope of scientific inquiry. From this perspective, something in principle irreducible to any future scientific description of the world would not actually be irreducible, but nonexistent. O’Shea has thus rightly spoken of an «omnivorous scientific image» (2007, p. 3). It is extremely relevant to note, as Willem deVries remarks in this volume, that this kind of *primacy* of the scientific image and its «explanatory adequacy» (to use David Lewis’ expression) does not only entail that it holds «an adequate ontology of basic objects» but something more, that is that such a framework «must contain (or be able to construct) all the predicates necessary to describing and explaining the world. In Quine’s usage, the framework must be ideologically complete as well». At the same time, Sellars is not saying that our *present* science is such an adequate *mensura*. On the contrary, its present categorial arrangement in Sellars’s view is inadequate to account for the intrinsic characters of sensible qualities and their «ultimate homogeneity». For this reason a different, «non particulate foundation» of the scientific image is required, one based on «‘absolute processes’, [...] subjectless (or objectless) events» (Sellars, 1981, §50, p. 48; see also pp. 85–87; cf. on this point: deVries, 2005, ch. 8). Besides, his scientism does not seem to imply the reducibility of the conceptual framework, within which we describe and explain the *living* world, to the conceptual framework that would have been enough for a *non living* world (what Sellars labels «*physical₂*»).¹² Therefore, we should not expect, for

¹¹ Cf. deVries and Triplett (2000, pp. 108–116). «Note that the key terms here are ‘describing’ and ‘explaining’ [...] for Sellars there is more to say and do than to describe and explain» (Bernstein, 1966, p. 120).

¹² Sellars proposes two different and complementary characterizations of the “physical”. “Physical” according to the first sense of the term (*physical₁*) «also includes the entities and attributes required

instance, to construct the scientific image using only those primitives that would have been sufficient for a mindless world. This is the (*weak*) form of emergentism, «the emergence form of the identity approach», defended by Sellars since his early writings: «Emergence is one form taken by a negative answer to the question: ‘Could a world which includes minds be described with the same primitive predicates (and laws) as a mindless universe?’» (1949, pp. 453–454; cf.: Rosenberg, 1982, p. 334). Finally, it is worth remarking that the thesis according to which science is «the measure of all things» – as well as that of its descriptive and explanatory «primacy» (PSIM, p. 32) – does not seem to be science, but *philosophy*: it is indeed the synoptic view of the philosopher which may express and justify this judgment.¹³

Therefore, it seems evident to me that the entire analysis of the relationships between the two images rests on Sellars’s very articulated and controversial philosophy of science, and every advance in our understanding of them depends on a deeper insight into these theses, or on a radical challenge to them. Most importantly, *and problematically*, these theses seem to configure, together, a «stance» (van Fraassen) according to which (our) science is ideally regarded as a «neutral» (Marsonet, 2000, pp. 26–27), perspectiveless view «from nowhere» (Nagel, 1986) on the furniture of our world – a view that one should allow to determine what is ultimately *real* (cf.: deVries, 2005, pp. 278–279).

There is another extremely relevant and controversial aspect of this analysis and of Sellars’s philosophy in general that I want to mention (and which is also related to the former). I refer to the general relationship Sellars draws between the descriptive/explanatory moment and the normative moment. This is a very complex point and I would simply like to briefly discuss two aspects of it.

for the scientific description and explanation of the behaviour of living organisms (provided only that these entities do not have the irreducible intentionality...). “Physical” according to the second sense of the term (*physical₂*) includes «objects and attributes which are necessary to and sufficient for the scientific description and explanation of the behaviour of non living matter, or which are definable in terms of such items and attributes» and according to it «sense impressions and their counterparts in an ideal neurophysiology would not be ‘*physical₂*’ (but they are “*physical₁*”; 1971, pp. 401–402, see also: 1981, pp. 85–87; cf. deVries (2005, pp. 225–226 and 235).

¹³ This could raise a question concerning *where* Sellars’s analysis is located. As van Fraassen noted: «In telling his story of those images, Sellars was [...] speaking from a perspective *located neither in the Manifest Image nor in the Scientific Image*» (1999, p. 42).

First, Sellars seems to envisage a very sharp distinction between description/explanation, on the one hand, and normative judgments based on standards, on the other. But this can be regarded as problematic. The idea that to evaluate is conceived as «something more» (PSIM, p. 39) than describing and explaining, something *over and above* descriptions and explanations, or that begins *after* them, seems to configure a kind of *dichotomy* not too distant from that between facts and values last criticized by Putnam (see De Anna's Commentary in this volume). If the «language of norms» is for Sellars «a mode of discourse which presupposes, but is irreducible to the “language of fact”» (1952, p. 516), couldn't a converse presupposition regarding *our* language of fact also be asserted? In other words, Sellars's idea that «the identity and individuation conditions of a scientific object should be resolutely non-normative, purely factual» or that things regarded «as physical objects» are «all capable of value-free description» (deVries, 2005, pp. 273 and 275, but see the entire ch. X) is worth examining and discussing further, as well as the consequences concerning our metaphysical judgments of reality and our ontological commitments which Sellars draws from it.¹⁴

Secondly, and most importantly, it is no less problematic to maintain that a fully naturalized description and explanation of man and the world would be possible, while at the same time maintaining that it would remain in principle impossible to logically reduce our normative notions, discourses and practices to the scientific framework. This *equilibrium* between a fully naturalized ontological (and «ideological») image of man within the space of causes, and a «space of reasons» (and «persons»), which is a normative standpoint that could never be logically and conceptually naturalized and reduced, is indeed regarded by some scholars as one of the main merits of Sellars's philosophy. On the other hand, there are those who regard this equilibrium as highly problematic and ultimately unsatisfactory, or at least as open to further developments. But, here the interpretations may diverge. For some of them, this means that, moving from a fully naturalistic image of man, it will be legitimate to also look for a naturalized normative space of reasons, or, at least (and much more in the spirit of Sellars's thought), for a full-fledged causal-naturalistic account of our inferential and normative practices (even if our intentional concepts still remain in principle irreducible). From this

¹⁴ In his contribution to this volume, for instance, deVries himself claims that no language can «be purely descriptive, independent of all normative, prescriptive or practical elements» and draws some conclusions concerning our ontological commitments from this fact.

perspective, the naturalistic side of Sellars's philosophy should be *extended* or *deepened*. For others (and I would put myself among them), on the contrary, the idea that our normative-inferential practices cannot be naturalized and reduced to the scientific conceptual framework presupposes (and is intertwined with) the thesis that also the *description* and *explanation* of persons and the world *cannot* be fully naturalized, and thus also presupposes a fundamental incompleteness of the scientific image even on this very level. From this perspective, the naturalistic/scientistic side of Sellars's philosophy had already gone *too far*.¹⁵

3. A Plural Heritage: The Many Legacies of Sellars's Philosophy

As is characteristic of many great philosophers, Sellars's philosophy has considerably influenced, and continues to influence, many diverse thinkers all over the world who use his insights to support various lines of argument. Recently, like for Hegel, a distinction between "right-wing" and "left-wing" Sellarsians has been proposed (especially focusing on the adoption or rejection, respectively, of his "*scientia mensura*" claim).

John McDowell, for instance (see: 1999, 2009), especially puts value on Sellars's account of knowledge and the intentional states as belonging to a «space of reasons» (as opposed, for instance, to «the space of placement in nature», or «the space of subsumption under [...] natural law»), where they are regarded «in the light of norms of justification». In his interpretation, this may secure a «special irreducibility» to the epistemic and normative concepts (very broadly conceived), saving epistemology and philosophy of mind from the pervasive (and typically modern) risk of a «naturalistic fallacy»: «When Sellars warns of a naturalistic fallacy, he is implying that the structure of the space of reasons is *sui generis*, by comparison with the kind of structure that the natural sciences find in nature» (1999, p. 260).¹⁶ This is regarded by him as one of the

¹⁵ Besides, this may also have led Sellars astray in examining the relationships between *reasons* and *causes* and the changes that a fully naturalized causal account of man would likely also have on the categories and status of our evaluations and judgments concerning persons, as well as on the space of reasons they are framed within.

¹⁶ From this point of view, a passage of EPM such as the following is obviously very relevant (§5): «Now the idea that epistemic facts can be analysed without remainder – even 'in principle' – into non-epistemic facts, whether phenomenal or behavioral, public or private, with no matter how lavish a sprinkling of subjunctives and hypotheticals is, I believe, a radical mistake – a mistake of a piece with the so-called 'naturalistic fallacy' in ethics».

great philosophical lessons of Sellars's thought. At the same time, to appreciate and develop this lesson may require, in certain cases, to take distance from what McDowell regards as Sellars's infelicitous *scientism* (see for instance footnote 10, above).

But it is very easy to find Sellars's influence on the opposite side, where a kind of "neurophilosophy" is dreamt. Paul Churchland (who wrote a dissertation on "Persons and P-Predicates" under Sellars in 1969) overtly embraces scientific realism (1979, §1), and he finds himself especially indebted to Sellars for the theoretical account of folk-psychology, which is essential for his idea of an elimination, or imperfect (not Nagelian) reduction of it by science.¹⁷ But at the same time he overtly rejects a conviction by Sellars that is regarded by many others as evidence for Sellars's equilibrium: the fact that he maintains folk psychology, inadequate as it is, as being massively referential, and assumes that the substitution of the manifest image by the scientific image in explaining our world is in general a matter of adequate «successor concepts», and not of elimination.¹⁸ As Churchland himself wrote recently in an interesting footnote:

«The reader will here [i.e., in the exposition of his interpretation] recognize Wilfrid Sellars' well-known account of the origins and nature of our Folk Psychology, as outlined in the closing sections of his classic paper, "Empiricism and the Philosophy of Mind," chap. 3 of *Science, Perception, and Reality* (London: Routledge, 1963). Ironically (from our present perspective), Sellars was blissfully convinced that Folk Psychology was an *accurate* portrayal of our inner cognitive activities. (I recall finding it advisable to down-play my own nascent eliminativism during my dissertation defense, a meeting chaired by that worthy philosopher.) But Sellars' conviction on this point notwithstanding, Folk Psychology had invited systematic scepticism long before the present, and for reasons above and beyond the recent flourishing of cognitive neurobiology.

¹⁷ «The first explicit portrayal of our collective self-conception as importantly theory-like appears in a landmark paper by Wilfrid Sellars [the reference is to *Empiricism and the philosophy of mind*]. [...] The bare possibility of a wholesale rejection of F[olk] P[sychology] is of course a simple consequence of FP's speculative theoretical status» (Churchland, 1994, pp. 308 and 310).

¹⁸ From a Sellarsian point of view (as we have seen) the manifest framework is not simply reducible to an explanatory theory and, even as far as descriptions and explanations are concerned, the scientific image *must* in general account for the main features of the manifest image, thus, as deVries and Triplett have remarked «though he [Sellars] maintains that the manifest image is inadequate and will have to be rejected as a whole, he also thinks that it poses a major constraint on what an adequate scientific image could be» (2000, p. 114). Cf. Garfield (1988, ch. 2 and 6); Marsonet (2000, ch. 1.3); deVries (2006, pp. 65–67). See also footnote 9.

See, for example, my “Eliminative Materialism and the Propositional Attitudes,” *Journal of Philosophy* 78, no. 2 (1981), now twenty years old. (Churchland, 2000, p. 294, n. 4; cf.: 1979, pp. 4–5 and 91, n.1).

But Sellars’s lesson has also been relevant for those who refused the very aspects of scientific realism and scientism dear to Churchland. Bas van Fraassen was a graduate student at the University of Pittsburgh in the first few years that Sellars taught there and during the ‘70s he debated scientific realism at length with Sellars himself.¹⁹ The *locus classicus* of his “constructive empiricism” is a famous book whose title explicitly depends on the essay we are discussing: *The Scientific Image*.²⁰ Subsequently, however, he also developed an analytical critique and “deconstruction” of the very categories of Sellars’s analysis – a deconstruction which in some way was implicitly present in his constructive empiricism from the beginning.²¹ And constructive empiricism probably also represents the most interesting and systematic attempt to escape from Sellars’s interpretation of science and the principles listed above (scientific realism, “*scientia mensura*” etc.): but obviously to deepen and reconceive the relationship between scientific theories and common-sense points of view from an «empirical stance» would require a monographic volume on its own.

4. This Issue

This issue of *Humana.Mente* aims to present:

- theoretical and original contributions on the problem of the encounter or «clash» between the two, broadly conceived, images and on the very idea of scientific and manifest “images”. From this perspective, any

¹⁹ On van Fraassen’s confrontation with Sellars’s philosophy, see van Fraassen (1975, 1977), a response to Sellars (1977, part I; 1999); and for his general approach to science: 1980, 2002. See also: Gutting (1982); O’Shea (2007, pp. 38–41).

²⁰ «The title of this book is a phrase of Wilfrid Sellars’s, who contrasts the scientific image of the world with the manifest image, the way the world appears in human observation. While I would deny the suggestion of dichotomy, the phrase seemed apt» (van Fraassen, 1980, p. vii).

²¹ «*there are no such things as the Manifest and the Scientific Image at all*. Is that possible? Yes, in fact I can think of some very good reasons for that conclusion. If you agree to them, you may even find some reason to generalize this skeptical conclusion to all those – what shall I call them? – world-pictures, conceptual frames, worlds (as in “the world of science”, “the world of the physicist”, “the Ptolemaic world”) which have so easily and smoothly crept into our discourse» (1999, p. 38).

aspect of the relationship between the scientific and common-sense/non-scientific understanding of man is relevant.

- critical re-examinations of Sellars’s elaboration of this topic, and an analysis of his relevant texts;
- overviews of contemporary debates on this topic, as well as on the related topic of the relationship between philosophy and the sciences.

I would like to say that many essays in this issue give a contribution to more than one of these three aims. At the same time, many theoretical questions outlined above are analysed in depth in these papers, along with many others.

In his very insightful paper (“Ontology and the Completeness of Sellars’s Two Images”) Willem deVries focuses on the problematic character of Sellars’s account of intentionality within the scientific image: this may sound surprising because ‘raw feels’ are rather regarded by Sellars as the very “hard problem” (at least for our present scientific image), while the «identification of conceptual thinking with neurophysiological process» (PSIM, p. 34) seems to Sellars relatively smooth, in principle. But, as deVries persuasively argues, things are much more problematic. Firstly, as he recollects, «intentionality is irreducible in the sense that we cannot define in any of the vocabularies of the natural sciences concepts equivalent to the concepts of intentionality». Now, if the scientific image has to give us a complete description and explanation of a world involving intentional phenomena, then a problem arises concerning the possibility of providing, *within* the scientific image, adequate «successor concepts» of those basic normative concepts, as well as adequate resources for the essentially first-person «expressive use of intention-talk». Otherwise, it seems that we should always add at least extra predicates to the scientific image. The problem posed by intentionality (and persons) seems, then, no less relevant and not too distant from that of *sensa*. Having clearly stated this problem, deVries tries to sketch the outlines of a Sellarsian response to it, i.e., that of embedding the conceptual framework of persons within the «more encompassing, impersonal framework» of the scientific image. But he eventually finds it unsatisfying and concludes that the scientific image cannot either be completely separated from the manifest one, or eliminate it, or stand alone.

P. Kyle Stanford (“The Eyes Don’t Have It: Fracturing the Scientific and Manifest Images”) analytically reconstructs the central argument of Sellars’ essay while highlighting the main (visual) metaphors around which it is built.

He argues that those metaphors are in part responsible for the ultimate failure of the argument, insofar as they lead Sellars to neglect the possibility that there may be a variety of heterogeneous ways in which various elements of the two images are related: some parts of the scientific image might represent mere conceptual tools or instruments, for example (even if it is implausible to so regard the entire scientific image). He goes on to suggest, however, that Sellars arrives at the manifest and scientific images themselves through a process of idealization, abstraction and construction in which we should decline to follow him: the idea that we face exactly two fundamental images which must be reconciled with one another ignores crucial differences between the various points of contrast.

In his very interesting theoretical contribution “Identifying and Reconciling Two Images of ‘Man’” David Hodgson proposes (i) to re-articulate the distinction between two different kinds of images of man and the world, using a criterion based on the presence (or absence) within them of subjective components (instead of the one based on the presence/absence of imperceptible entities): on the one hand, we will have a «subjective folk-psychological image», and on the other an «objective scientific image».

Secondly, (ii) he articulates and defends an approach to the reconciliation of the two images which is different from Sellars’s approach, because, as he argues, «representation of reality requires both the subjective [...] image and the objective scientific image», while the latter alone cannot in principle provide us with a complete *description* and *explanation* of man and the world.

In his paper “Of Time and the Two Images” Steven Savitt discusses the relationship between the scientific and the manifest image from the point of view of the account they give of time. Sellars does not deal with this topic in PSIM, but it obviously has an enormous relevance for the appreciation of each image and of their mutual connection. Sellars confronted himself with the difficulties concerning time in a previous essay, *Time and the World Order* (1958). Savitt clearly presents the relevance of the topic, as well as Sellars’s attempt to locate the manifest or “folk” time within the framework of Special Theory of Relativity: a project, he argues, that is not the best option available for Sellars and should be eventually regarded as an unsuccessful one.

Keith Lehrer (“The Unity of the Manifest and Scientific Image by Self-Representation”) deepens here his long-term philosophical elaboration on representation and self-representation, and its analogies with Sellars’s view. In

so doing, he analyses the distinction between the two images and their respective conceptual frameworks, identifying the problematic feature of the distinction in Sellars's theory of inner episodes as theoretical entities. Sellars's account of our non-inferential knowledge is a very relevant aspect of his philosophy and in Lehrer's analysis it also becomes a key for interpreting the relationship between the manifest image and the scientific image, and the transition from the first to the second. In particular, according to Lehrer, self-representation and reflexive self-description may «provide an arch of representation connection between the conceptual framework of the manifest image and the scientific image».

Giacomo Turbanti ("Normativity and the Realist Stance in Semantics") starts a series of papers concerning the debate on normativity (and its relationship with naturalism). He especially deals with the normativity of meaning and semantic notions (like *reference* and *truth*). In so doing, he defends the compatibility of « a realist stance in semantics and a non-reductive account of the normativity of meaning», arguing that skepticism is not triggered by the normativity of meaning and that the rejection of the "Myth of the Given" is compatible with realism in semantics. Then, in a Sellarsian spirit, Turbanti aims at clarifying « how the normative analysis of linguistic roles may fit into the explanation of linguistic behavior provided by formal semantics». He sees this problem as a «particular instance» of the problem of fusing the two images by joining the normative vocabulary of shared intentions of a community of rational agents to the scientific image.

Jay Garfield ("Sellarsian Synopsis: Integrating the Images") interestingly draws our attention to a third 'image' in Sellars's essay which could be very relevant, but has been neglected: the «*original*» image. According to Sellars, this original image is, as we have seen, «a framework in which all the "objects" are persons» (PSIM, p. 10). Garfield underlines that this image is not only a relic from our past, but an expression of our «innate fundamental propensities to attribute intentionality». This capacity would contribute to the «ontogenesis of communities» and the birth of communities would, in turn, contribute to the birth of a space of norms and reasons which would eventually lead to the birth of science. From this perspective, «The original, from the standpoint of the scientific, hence explains the manifest», that is, that a naturalistic reconstruction of the capacities that generate the original image, and of the transition from one to another of the three images, might not only explain their consistence, but also their mutual «entailing» and how «naturalizing the

normative» also means «norming nature».

Owen Flanagan and Stephen Martin (“Science and the Modest Image of Epistemology”) also deal with the problem of epistemic normativity and justification within Sellars’s naturalistic framework, and the problem of a “naturalized” epistemology. What about truth, reasons and standards “in a world of causes”? And what about the role that they have within science, in order to make scientific claims genuine episodes of knowledge? But the point we are faced with is also «whether an ability, a sort of freedom to decide how to act and what to believe, respectively, remains available to us in light of information we have about how the world – including most relevantly, the mind – works». Flanagan and Martin underline the relevance of an accurate phenomenology of our epistemic and reasons-based practices as the basis for a reconciliation of them with the scientific image, and then, in the spirit of Dewey, propose a “compatibilist” and naturalistic account of these practices, arguing that the scientific image does not undermine the common-sense conceptual framework of reasoning and responsibility (“responsibility” with an ‘a’ being the ability to produce differential responses to future circumstances based on feedback about past successes and failures).

James O’Shea, with his great expertise in Sellars’s philosophy, in his paper (“Prospects for a Synoptic Vision of our Thinking Nature: On Sellars, Brandom, and Millikan”) analyses the crucial problem of providing a philosophical account of «norm-governed conceptual thinking within the natural world». That is: how is it possible, in a Sellarsian spirit, to “stereoscopically” combine an inferential, normative account of thought and a naturalist image of the world? O’Shea examines this task discussing the (very different) philosophical points of view of Robert Brandom and Ruth Millikan. He especially deals with two questions, namely:

(1) How is it possible to account for animal representations and, more generally, for a causal-naturalistic notion of representation within a normative framework? The issue seems extremely relevant to O’Shea also because he maintains that in Sellars’s thought the naturalistic dimension of representation is neither underrated, nor effaced by the normative one;

(2) Is it possible to develop a naturalistic attitude even towards our conceptual activities themselves (at the same time preserving the irreducibility of normative terms)?

Philosophy and the Scientific Image of Man also outlines Sellars’s view on

the nature of philosophy and on the relationship between philosophy and science. His way of conceiving this relationship has been an influential one, but other very relevant models were present in the same period within analytic philosophy. Diego Marconi's essay ("Quine and Wittgenstein on the Science/Philosophy Divide") clearly analyses and contrasts Wittgenstein's and Quine's perspective on this very subject. He especially focuses on the issue of *continuity* or *discontinuity* between philosophy and science. After analysing Quine's arguments *for* "continuism", he shows why they are not conclusive from a Wittgensteinian point of view, and which arguments Wittgenstein may advance *against* this continuity. From Marconi's reconstruction it also emerges that, even if Wittgenstein's rejection of continuity between philosophy and science can be regarded as a constant feature of his thought, in the very late period of his career, new questions and new perspectives emerge concerning the relevance of scientific facts to philosophy.

In the section devoted to the Commentaries, Massimo Marraffa, Raffaella Campaner and Gabriele De Anna analyse and discuss three books, published after *Philosophy and the Scientific Image of Man*, which constitute very relevant contributions on key aspects concerning the relationship about the two images: from transcendental arguments (Strawson), to the theory of causation and explanation (Salmon), to the dichotomy or distinction between facts and values (Putnam).

The volume ends with Luca Corti's Review of a recent recollection of studies on Sellars's philosophy (*Empiricism, Perceptual Knowledge, Normativity, and Realism: Essays on Wilfrid Sellars*, edited by Willem de Vries) – another testimony of the interest in Sellars's philosophy, which was also eventually confirmed by the foundation in 2012 of the Wilfrid Sellars Society (WSS).²²

I am extremely grateful to all the philosophers who have generously contributed to this Issue, in spite of their numerous commitments, and also to all those who responded to our call for papers. Many thanks to Silvano Zipoli Caiani, Executive Director of *Humana.Mente*, for giving me the possibility to freely plan and put together this monographic volume, and thanks also for all the support I received from him and Marco Fenici. The work of referees has

²² See: <<http://wss.categorymistake.com/wp/>> (accessed on May 13th 2012). For many materials concerning Sellars's thought, as well as his texts, see also the web-site "Problems from Wilfrid Sellars": <<http://www.ditext.com/sellars/sellars.html>> (accessed on May 13th 2012).

been attentive and precious. Last, but very much not least, I'm very grateful to Alice Giuliani for her expertise and willingness to do the final *mise en page* of the entire volume.²³

REFERENCES

- Bernstein, R.J. (1966). Sellars' Vision of Man-in-the-Universe, I. *The Review of Metaphysics*, 20(1), 113–143.
- Churchland, P.M. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
- Churchland, P.M. (1994). Folk Psychology. In S. Guttenplan (Ed.), *A Companion to the philosophy of mind*. Oxford: Blackwell, 308–316.
- Churchland, P.M. (2000). Rules, Know-How, and the Future of Moral Cognition. *Canadian Journal of Philosophy (Supplementary Volume)*, 26, 291–306.
- deVries, W.A. (2005). *Wilfrid Sellars*. Chesham: Acumen.
- deVries, W.A. (2006). Folk Psychology, Theories and the Sellarsian Roots. In M.P. Woolf & M. Nn. Lance (Eds.), *The Self-Correcting Enterprise. Essays on Wilfrid Sellars*. Amsterdam-New York: Rodopi, 53–84.
- deVries, W.A., & Triplett, T. (2000). *Knowledge, Mind, and the Given. Reading Wilfrid Sellars's "Empiricism and the Philosophy of Mind"*. Indianapolis-Cambridge: Hackett.
- Elton, M. (2003). *Daniel Dennett. Reconciling Science and Our Self-Conception*. Cambridge-Malden, MA: Polity Press.
- Feigl, H. (1967). Postscript after Ten Years. In H. Feigl, *The "Mental" and the "Physical": the essay with a Postscript after Ten Years*. Minneapolis: University of Minnesota Press, 133–170.
- Garfield, J. (1988). *Belief in Psychology*. Cambridge, MA: MIT Press.

²³ A special thanks to Willem deVries and Alexandra Shakespear for reading a previous version of this Introduction and improving its content and style. Obviously, I am the only one responsible for the final result.

- Garfield, J. (1989). The Myth of Jones and the Mirror of Nature: Reflections on Introspection. *Philosophy and Phenomenological Research*, 50(1), 1–26.
- Gehlen, A. (1940/1988). *Man: His Nature and Place in the World*. New York, NY: Columbia University Press. First German ed. by Klostermann, Frankfurt am Main 1940.
- Gutting, G. (1982). Scientific Realism vs. Constructive Empiricism: A Dialogue. *The Monist*, 65(3), 336–349. Reprinted in P.M. Churchland & C.A. Hooker (Eds.), *Images of Science. Essays on Empiricism and Realism with a Reply by Bas C. van Fraassen*. Chicago: The University of Chicago Press, 118–131.
- Haack, S. (1993). *Evidence and Inquiry. Towards Reconstruction in Epistemology*. Oxford: Blackwell.
- Marsonet, M. (2000). *I limiti del realismo. Filosofia, scienza e senso comune*. Milano: Franco Angeli.
- McDowell, J. (1999). *Naturalism in the Philosophy of Mind*. Now in J. McDowell (2009), *The Engaged Intellect. Philosophical Essays*. Cambridge: Harvard University Press, ch. 15, 257–275.
- McDowell, J. (2009). The Constitutive Ideal of Rationality: Davidson and Sellars. In J. McDowell, *Having the World in View*. Cambridge, MA: Harvard University Press, ch. 11, 207–220.
- McDowell, J. (2009a). The Logical Form of Intuition. In J. McDowell, *Having the World in View. Essays on Kant, Hegel and Sellars*. Cambridge, MA: Harvard University Press, ch. 2, 23–43.
- Nagel, T. (1986). *The View from Nowhere*. New York-Oxford: Oxford University Press.
- O’Shea, J. (2007). *Wilfrid Sellars. Naturalism With a Normative Turn*. Cambridge-Malden, MA: Polity Press.
- Putnam, H. (2012). *Philosophy in an Age of Science. Physics, Mathematics, and Skepticism*. Ed. by M. De Caro & D. Macarthur. Cambridge-London: Harvard University Press.

- Ravenscroft, I., & Stich, S. (1994). *What Is Folk Psychology?* Now in S. Stich (1996), *Deconstructing the Mind*. New York-Oxford: Oxford University Press, ch. 3, 115–135.
- Rosenberg, J.F. (1982). The Place of Colour in the Scheme of Things: A Roadmap to Sellars's Carus Lectures. *The Monist*, 65(3), 315–335.
- Schlick, M. (1925/1974). *Allgemeine Erkenntnislehre*. Berlin: Springer. Engl. transl.: *General Theory of Knowledge*, Open Court, La Salle.
- Sellars, W. (1949). Realism and the New Way of Words. In H. Feigl & W. Sellars (Eds.), *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts, 424–456. First pub. in *Philosophy and Phenomenological Research*, 8(1948), 601–634.
- Sellars, W. (1952). Obligation and Motivation. In W. Sellars & J. Hospers (Eds.), *Readings in Ethical Theory*. New York: Appleton-Century-Crofts, 511–517.
- Sellars, W. (1962). Philosophy and the scientific image of man. In R.G. Colodny (Ed.), *Frontiers of Science and Philosophy*. Pittsburgh, PA: University of Pittsburgh Press, 35–78.
- Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge & Kegan Paul.
- (EPM) Sellars, W. (1963). Empiricism and the Philosophy of Mind. In W. Sellars, *Science, Perception and Reality*. New York: Humanities Press, 127–196. First ed. by University of Minnesota Press, Minneapolis, 1956.
- (PSIM) Sellars, W. (1963). Philosophy and the scientific image of man. In W. Sellars, *Science, perception, and reality*. London: Routledge & Kegan Paul, 1–40.
- (PHM) Sellars, W. (1963). Phenomenalism. In W. Sellars, *Science, perception, and reality*. London: Routledge & Kegan Paul, 60–105.
- Sellars, W. (1968). *Science and Metaphysics: Variations on Kantian Themes*. London: Routledge & Kegan Paul. Reissued by Ridgeview Publishing Company, Atascadero, CA, 1992.

- Sellars, W. (1971). Science, Sense Impressions, and Sensa: A Reply to Cornman. *Review of Metaphysics*, 24, 391–447.
- Sellars, W. (1977). Is Scientific Realism Tenable?. In F. Suppe & P. Asquith (Eds.), *PSA 1976, Volume II (Philosophy of Science Association, East Lansing)*, 307–334.
- Sellars, W. (1981). Foundations for a Metaphysics of Pure Process. *The Monist*, 64(1), 3–90.
- Sklar, L. (2001). Naturalism and the Interpretation of Theories. *Proceedings and Addresses of the American Philosophical Association*, 75(2), 43–58.
- van Fraassen, B.C. (1975). Wilfrid Sellars on Scientific Realism. *Dialogue*, 14(04), 606–616.
- van Fraassen, B.C. (1977). On the Radical Incompleteness of the Manifest Image (Comments on Sellars). In F. Suppe & P. Asquith (Eds.), *PSA 1976, Volume II (Philosophy of Science Association, East Lansing)*, 335–343.
- van Fraassen, B.C. (1980). *The Scientific Image*. Oxford: Oxford University Press.
- van Fraassen, B.C. (1999). *The Manifest Image and the Scientific Image*. In D. Aerts (Ed.), *Einstein Meets Magritte: The White Book – An Interdisciplinary Reflection*. Dordrecht: Kluwer, 29–52.
- van Fraassen, B.C. (2002). *The Empirical Stance*. New Haven: Yale University Press.

Ontology and the Completeness of Sellars's Two Images^{*}

Willem A. de Vries[†]
willem.devries@unh.edu

ABSTRACT

Sellars claims completeness for both the “manifest” and the “scientific images” in a way that tempts one to assume that they are independent of each other, while, in fact, they must share at least one common element: the language of individual and community intentions. I argue that this significantly muddies the waters concerning his claim of ontological primacy for the scientific image, though not in favor of the ontological primacy of the manifest image. The lesson I draw is that we need to re-assess the aims of ontology.

I

There is an apparent puzzle in Sellars's characterization of the relation between the “manifest” and “scientific images”. Sellars clearly gives ontological priority to what he calls the “scientific image of man in the world”. He could not make this plainer than he does in his *scientia mensura*: «in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not» (EPM, §41; SPR, p. 172; KMG, p. 253).¹ The “scientific image” is something we are only partially in possession of; it is still in the process of formation. We have some conception of what that image might turn out to be, but the items we take to be fundamental in the image have been changing steadily for the past century. We have a better understanding, however, of *how* that image will come to be, for

^{*} I thank Joseph Rouse, Charlotte Witt, Mark Okrent, and Lauren Ashwell for very helpful comments on an earlier draft of this paper.

[†] University of New Hampshire.

¹ I cite Sellars's work using the standard abbreviations of the titles. See the bibliography for details.

we have a firmer grip on at least some of the proper methods of scientific theorizing than we do on the final results of such theorizing. Even so, it is clear that the scientific image will radically revise the received ontology and ideology of the contrasting manifest image.

The manifest image is a refinement of the conceptual scheme that evolution and the forces of natural history caused us to develop. It is the conceptual scheme in terms of which we came to be able to confront ourselves as persons, reflective agents subject to and active within a world. It is, in fact, the conceptual scheme the development of which *made* us persons.

The ontology implicit in the manifest image has been given different interpretations within *philosophia perennis*, the line of thought within the tradition that endorses the manifest image as real. Sellars himself thinks that what he calls «the Aristotelian-Strawsonian reconstruction» (MP, par. 60, p. 252) best captures the central structures of the manifest image.² This is an ontology of persons and things, but the emphasis is on *persons*. As Sellars construes the manifest image, persons are not thought of as *things* spiffed up with some extra properties and capacities beyond the run-of-the-mill capacities of normal things. Rather, *person* was the original category of the incipient image, and (mere) *things* were thought of as “truncated persons” with a vastly pared down complement of capacities. Part of the force of this thought, surely, is that, in the received tradition that is the manifest image, persons normally can (and do) become things by the subtraction or loss of certain properties or capacities – we normally call this “death” – but it is abnormal and unnatural for a thing to become a person by the addition of certain properties or capacities. Infusing personhood in a mere thing or assemblage of things traditionally requires some kind of supernatural intervention. There is, of course, a natural process by which persons are generated, but it is not a matter of assembling impersonal things. That was the dominant view for most of human history. That it is no longer simply obvious is a testament to the power of science to change what is manifest.

Aristotelian-Strawsonian persons are *unities*. Of course, they have complexes of complex properties, but they are not *teams* like Cartesian

² Sellars is very aware of the distinction between Aristotle and Aristotelians and especially «contemporary Aristotelians (such as Strawson)» (SM, VI, par. 54, pp. 170–171). He thinks that the ordinary-language philosophers were reviving an essentially Aristotelian framework, adapting it to contemporary circumstances.

persons. It is the same person who thinks who runs. More deeply, they are *basic individuals* in the manifest image (cf., SK, I §29, p. 303).

A person, then, according to the Aristotelian analysis, is a single individual which does not have subordinate individuals as its parts. Its unity is not that of a system (MP, p. 222; KTM, p. 284).

The “parts” we might think make up a person – arms, legs, nose, etc. – are merely *potential* parts; when made *actual* parts, that is, when separated from each other, they are no longer actual arms, legs, etc. Again, an important aspect of the idea is that we cannot just “build” a person out of parts; prior to assembly the parts aren’t of the right kind, and spatial assembly is not the right form of unity. Science has given us new or revised conceptions of what might count as the parts of person, such things as molecules or atoms, but just how assemblages of such things might stand to the unities that are persons has been of much debate.

Sellars proposes that the manifest image we are all raised into is under challenge from the scientific image, which intends to replace it. The manifest image has arisen more or less willy-nilly in the course of human development, so its organization is not patent to the unreflective eye, though Sellars thinks it is far from incoherent. The manifest image, however, raises a number of questions that it cannot give answers to.

Indeed, the development of the sciences puts ever-increasing pressure on the manifest image to reconcile scientific discovery with manifest truth. The scientific image, in contrast, is a rationally constructed and increasingly explicitly formulated framework that postulates new kinds of entities in order to perform its explanatory task and is responsible to the world via rigorously conducted programs of empirical research. We build and refine it with conscious, rational intent. We can also project a Peircean ideal in which the sciences finally settle into theories that are up to any challenge, a time when further revisions to science are only extensions or applications of a stable set of theories.

According to Sellars, both the manifest and the scientific images purport to be *complete*, «i.e., to define a framework which could be the *whole truth* about that which belongs to the image» (PSIM, par. 56; SPR: p. 20; ISR, p. 388). This characterization leaves still pretty vague what it means for a conceptual framework to be “complete”. There are two problems with it. First, the final clause, stipulating that completeness means being the whole truth «about that

which belongs to the image», threatens emptiness unless there is some non-circular specification of what “belongs to the image”. If what belongs to the image is only what it talks about, then completeness is too easy. If, for instance, we exclude from science such problematic cases as, say, the 4 M’s identified by Huw Price – Morality, Modality, Meaning, and the Mental – it seems relatively easy for science to be “complete”. Sellars includes the clause about what belongs to a framework because

[T]he conception of *the* scientific or postulational image is an idealization in the sense that it is a conception of an integration of a manifold of images, each of which is the application to man of a framework of concepts which have a certain autonomy. (PSIM, par. 55; SPR, p. 20; ISR, p. 388).

But the scientific and manifest images themselves are supposed to be *total* and all-inclusive images of *the world*. As far as I can see, a truly *complete* framework could only be one that could be the whole truth, *period*. This requires two dimensions of elaboration, however.

First, a conceptual framework is incomplete if, in order to explain some phenomenon, new kinds of objects must be added to its ontology (PSIM, par. 105; SPR, p.36; ISR, p. 405). Sellars must mean that a framework is incomplete if the task of description and explanation requires it to add new kinds of *basic* objects; a framework that can *construct* from its prior resources all the object-kinds it could need to describe and explain the world is not incomplete, just not yet fully elaborated. Yet containing a complete set of basic objects cannot, I think, be sufficient for completeness.

Second, the requirement that a complete framework be capable of being the «*whole truth*» requires more than having an adequate ontology of basic objects, for truths go beyond objects: truths and their alter egos, facts, have propositional structure, especially predicative structure, so to be capable of the whole truth, a framework must contain (or be able to construct) all the predicates necessary to describing and explaining the world. In Quine’s usage, the framework must be *ideologically complete* as well. The complete conceptual framework must not just encompass all that there is, but all that can be said about it.

Sellars devoted a significant amount of his work to arguing that the scientific framework, in order to legitimately claim completeness, would have to expand to include *sensa* – emergent, basic objects required to enable us to explain why color predicates have the logical grammar they do. But this is not

my concern here. Surprisingly, the status of intentionality in Sellars's conception of the scientific image is murkier in some respects than that of sensation. This is surprising because Sellars developed a theory of intentionality that seems calculated to so construe intentional phenomena as to make them compatible with developments in the sciences.

Now if thoughts are items which are conceived in terms of the roles they play, then there is no barrier in principle to the identification of conceptual thinking with neurophysiological process. There would be no "qualitative" remainder to be accounted for. The identification, curiously enough, would be even more straightforward than the identification of the physical things in the manifest image with complex systems of physical particles. And in this key, if not decisive, respect, the respect in which both images are concerned with conceptual thinking (which is the distinctive trait of man), *the manifest and scientific images could merge without clash in the synoptic view* (PSIM, par. 96; SPR: p. 34; ISR: p. 402).

But it turns out that things are not quite so simple. The first thing to notice is that Sellars maintains that intentionality is irreducible in the sense that we cannot define in any of the vocabularies of the natural sciences concepts equivalent to the concepts of intentionality. The language of intentionality is introduced as an autonomous explanatory vocabulary tied, of course, to the vocabulary of empirical behavior, but not reducible to that language. The autonomy of mentalistic discourse surely commits us to a new ideology, a new set of basic predicates, above and beyond what can be constructed in the vocabularies of the natural sciences. What we get from the sciences can be the *whole truth* about the world, including intentional phenomena, then, only if there is some way to construct, using proper scientific methodology, concepts in the scientific image that are legitimate successors to the concepts of intentionality present in the manifest image. That there is such a rigorous construction of successors to the concepts of intentionality is, I think, a clear commitment on Sellars's part, though it would have been nice had he spelled out in greater detail the metaphysics of the functionalist theories that would have to be involved. The only real alternative is some form of eliminativism, an alternative that some of his students adopted and some of his critics thought Sellars was committed to, but which never held any real attraction for Sellars.³

³ The students include the Churchlands; the critics Joseph Margolis.

The second thing to notice is that the concepts of intentionality, especially the concepts of agency, differ in some significant ways from the normal concepts of the natural sciences. In PSIM Sellars puts it this way:

To say that a certain person desired to do A, thought it his duty to do B but was forced to do C, is not to *describe* him as one might describe a scientific specimen. One does, indeed, describe him, but one does something more. And it is this something more which is the irreducible core of the framework of persons. (PSIM, par. 111; SPR, p. 39; ISR, p. 407)

Here the focus is explicitly on the language of agency, but the point is fundamentally the same as in Sellars's well-known dictum from EPM:

in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (EPM, §36; SPR, p. 169; KMG, p. 248)

In both epistemic and agential language something extra-descriptive is going on. In order to accommodate this important aspect of such phenomena, Sellars tells us, we must add to the purely descriptive/explanatory vocabulary of the sciences “the language of individual and community intentions”. He points to *intentions* here because the point is that epistemic and agential language – mentalistic language in general – is ineluctably *normative*; it always contains a prescriptive, action-oriented dimension and engages in direct or indirect assessment against normative standards. In Sellars's own theory, norms are grounded in the structure of intentions, particularly community intentions, so any truly complete image must contain the language of intentions.

We need to be a bit careful here, for the language of intentions can be used in two very different ways. We can use it *descriptively*, attributing to ourselves or others intentions that figure in third-person explanations of their behavior. But we can also use it *expressively*, which use is necessarily first-personal. For example, one might say, after committing some gaff, “I intend to set things right”, or even more directly, “I will make it right”. In its expressive use “I intend to set things right” is not a self-attribution of an intention to be used in the explanation of my behavior, but an expression of that intention, which, as an intention, also normally expresses itself in action. Other parties can attribute to me the intention to set things right, but they cannot *express* that intention in the requisite sense. There can be ways I might attribute such an intention to myself (de facto) in which I am blocked from *expressing* the

intention because I don't realize that the person to whom I have attributed the intention is *me*, the speaker-agent. John Perry and Hector Castañeda have brought to light the peculiar logic of such first-personal expressions, but Sellars was conscious of it, and I think he is committed to the notion that one does not fully command "the language of individual and community intentions" unless one grasps the expressive use of such language. Grasping such a use of language, given Sellars's approach to language, entails having such usage in one's own repertoire. Without an understanding of the expressive role of intentions, I doubt one can make good sense of the distinction between a norm as a statistical regularity in a population and a norm as an action-controlling prescription.

The expressive use of intention-talk, because it is essentially first-personal, seems to outstrip the resources made available through the natural sciences, yet it cannot be ignored or eliminated from our thought or language, and certainly not from scientific practice, which is as norm-rich an activity as one can imagine. That is why Sellars says that the ultimate image must be stereoscopic, joining the purely descriptive vocabulary of the sciences with "the language of individual and community intentions", which he believes grounds all normativity.

So here is another source of potential incompleteness in the scientific image: even if normativity does not require us to add new *objects* to the ontology of the scientific image, it does at very least require us to expand the ideology of the image by joining to it the ideology of intentions. Sellars tells us that the scientific image purports to be complete, but he then turns around almost immediately to put that claim in doubt, because the scientific image apparently needs supplementation with the language of intentions and agency.

There is a seemingly cheap and easy way out of this problem that is, however, unavailable to Sellars. One could deny that the scientific image is incomplete in the way I have just suggested by denying that the language of individual and community intentions and the normative language based on it contains any *truths*. If normative language is not ultimately truth-apt, then science could provide the *whole truth* even though it makes no mention of norms. But Sellars's own treatment of truth rules out this thoroughly non-cognitivist approach in two ways. First, in his view, truth itself is an essentially normative notion, so science cannot avoid norms unless it avoids all talk of truth. He views truth as *warranted assertibility*, where that means assertibility in accordance with the proper application of the ideal rules of the language.

That is certainly loaded with normativity. Second, Sellars clearly believes normative claims can be warrantably assertible. In his view, normative claims such as “one ought to treat all people with respect” are as truth-apt as empirical claims such as “Diamonds are the hardest substance known”. He cannot claim completeness for the scientific image by restricting the scope of the truth predicate.

So we have to ask the question: How are we supposed to think about joining to the descriptive resources of the scientific image the prescriptive resources generated by the language of intentions?

II

Can we consider joining the language of individual and community intentions – ultimately, the language of normativity – to an otherwise purely descriptive language of science to be ontologically conservative? It means adding to the ontology of the scientific image a kind of thing otherwise not countenanced in it: persons. This is the apparent puzzle for Sellars I mentioned at the beginning of this paper. He seems committed to both the completeness and the incompleteness of the Scientific image – and, not surprisingly, *persons* form the nub of the issue.

1. «A person can almost be defined as a being that has intentions» (PSIM, par. 114; SPR: p. 40; ISR: p. 408). (Let’s forget the “almost” for the purposes of this argument.)
2. In Sellars’s own analysis of intentions and intention-talk, there is a built-in first-person reference in all intentions.
3. While there are community intentions – “we” intentions – that can be the intentions of a system (such as a government), such intentions presuppose individual intentions in which the first-person singular reference is unanalyzable in terms of any further constitutive system.
4. First-person singular reference is ineliminable from language and conceptual thought.
5. The things referred to in first-person singular references are (surprise, surprise!) persons.
6. In the manifest image, the original home of the language/concepts of individual and community intentions, persons are basic individuals and person-level properties are not in general reducible to the properties of the parts of persons and their relations.

7. «[T]he scientific image of man turns out to be that of a complex physical system» (PSIM, par. 70; SPR, p. 25; ISR, p. 393).
8. Adding the language of individual and community intentions to the scientific image effectively adds a set of new basic individuals to that image, namely persons, possessors of a first-person perspective.

Sellars acknowledges the force of this argument when he grants that

if the human body is a system of particles, the body cannot be the subject of thinking and feeling, *unless thinking and feeling are capable of interpretation as complex interactions of physical particles*; unless, that is to say, the manifest framework of man as *one* being, a *person* capable of doing radically different kinds of things can be replaced without loss of descriptive and explanatory power by a postulational image in which he is a complex of physical particles, and all his activities a matter of the particles changing in state and relationship. (PSIM, par. 80; SPR, p. 29; ISR, p. 397)

Sellars has the problem of *sensa* primarily on his mind when he says this, and he does not believe that the sensory can be dealt with adequately, given the current stock of physical particles. In effect, I am raising the question: can intentions be dealt with adequately in an image according to which persons are complex physical systems and all a person's "activities [are] a matter of the particles changing in state and relationship"? Notice that it is a condition on "dealing adequately" with the phenomena of the manifest image that the concepts of the replacement scientific image not lose "descriptive and explanatory power". This can be difficult to judge, since the descriptive and explanatory resources of the scientific image are inevitably *different* from those of the manifest image. Presumably, in abandoning the myth of Demeter and Persephone in favor of the tilt of the earth's axis of rotation relative to its orbit around the sun, we have not lost descriptive or explanatory power with regard to the basic phenomenon: the change of seasons. Can the scientific image develop a recognizable successor to the manifest concept of a person that will enable it to do justice to the basic phenomena we employ the manifest concept to understand?

This argument mobilizes in a Sellarsian context some considerations that others have used to attack naturalism (and physicalism as well) based on the idea that the naturalistic picture of the world is thoroughly third-personal, or better, *impersonal*. In such attacks, the propriety of making first-person reference, our knowledge of first-personal facts, and even the existence of

persons are taken to be both beyond question and unable to be accommodated in the naturalistic picture of things.⁴

III

Sellars's basic line of response to such an objection is fairly clear. The concepts of personhood are essentially *functional* concepts, and as such, they do not commit us to any particular *ultimate* ontology. Ontologically, persons, their states and activities, will be dependent objects. This needs to be carefully reconciled with Sellars's conviction that «In the Manifest Image, a person is a basic individual» (SK, I, par. 29, p. 303), that «in the common sense framework of persons and physical objects [...], thoughts and sense impressions are adjectival to single logical subjects (as contrasted with sets of logical subjects)» (PHM, par. 94; SPR, p. 100; ISR, p. 344). The strategy for reconciliation is also fairly clear: distinct ideologies do not entail distinct ontologies.

The basic roadblock [to seeing a person as a physical system] is the unity of the person as the subject of conceptual activities. (PHM, par. 95; SPR, p. 100; ISR, p. 345)

But Kant, Sellars thinks, shows us a way to avoid taking the unity of a person to be ontologically ultimate and irreducible.

The heart of the matter is the fact that the irreducibility of the 'I' within the framework of first person discourse (and, indeed, of 'you' and 'he' as well) is compatible with the thesis that persons can (in principle) be exhaustively described in terms which involve no reference to such an irreducible logical subject. For the description will *mention* rather than *use* the framework to which these logical subjects belong. Kant saw that the transcendental unity of apperception is a form of experience rather than a disclosure of ultimate reality. If persons are "really" multiplicities of logical subjects, then unless these multiplicities used the conceptual framework of persons there would be no persons. But the idea that persons "really are" such multiplicities does not require that concepts pertaining to persons be *analysable into* concepts pertaining to sets of logical subjects. Persons may "really be" bundles, but the concept of a person is not the concept of a bundle (PHM, par. 95; SPR, pp.

⁴ The importance of the contrast between the first- and third-person perspectives was perhaps first made prominent (in recent times) in Thomas Nagel (1965, 1974, 1986). We can find the argument spelled out clearly in the work of Lynne Rudder Baker (1998, 2007, 2011).

100–101; ISR, p. 345).

The crucial move here is that the conceptual framework of persons and intentions gets embedded within a more encompassing, impersonal framework. In that larger framework, there is no commitment to the reality of persons or intentions as bottom-level individuals, but there is a commitment to the existence of states and relations of physical systems that are interpretable as perceptual responses to the world, as thoughts, and as actions. Such states and relations are possible only within a broader system of such states and relations, some of which are interpretable (by other such systems) as *employing* the concepts of personhood, agency, etc.

Notice that something's being *interpretable* as a perception, thought, or action is not, on this view, an "absolute" property it possesses aperspectivally; interpretability is always interpretability *to some (possible) interpreter*. Being an interpreter is also not an "absolute" property. The possibility (indeed, it turns out, the necessity) of self-interpretation complicates matters. There are persons only because there are complex systems that interact with themselves, with the world, and with other such systems in such ways that their states and activities realize or implement the conceptual framework of persons. Ultimately, however, that means that these systems are such as to be interpretable *by us*, for the conceptual framework of persons is *our* framework. We can distance ourselves from the framework of persons only in a kind of self-alienation that would change our fundamental nature.

The idea that the framework of persons, which is anchored in its use by certain beings with complex internal structures and external relations, can be embedded in a larger, aperspectival picture of the world that makes no mention itself of persons does not automatically dissolve the notion that persons are somehow basic individuals. It may be true that "persons can (in principle) be exhaustively described in terms which involve no reference to such an irreducible logical subject", but it is not clear just what this establishes. The activity of describing itself presupposes the framework of persons. Any particular person may be described "in terms which involve no reference to [persons as] irreducible logical subject[s]", but this cannot occur in circumstances in which there is *no reference at all* to persons as logical subjects, for then there would be no describers. As Sellars himself notes, «unless these multiplicities [that persons are now conceived to be] *used* the conceptual framework of persons there would be no persons» (my emphasis). Referring, describing, and explaining are, of course, *things people do*. Any talk

of conceptual frameworks ultimately has to be cashed out in terms of the conceptually-informed activities of the possessors of such a framework, so reference to persons as the possessors and utilizers of conceptual frameworks is implicit whenever such frameworks or framework-enabled activities are the topic. We can eliminate reference to persons on a piecemeal, even a wide-spread piecemeal basis, but we cannot eliminate all reference to persons without pulling the rug out from the very set-up of the problem. There can be no full-fledged conceptual framework that does not contain and, on its own terms, properly apply to real objects, the concept of a person or its equivalent.⁵

The conclusion I draw is that Sellars misleads us in some important ways in his contrast between the scientific and manifest images. His talk of the completeness of the scientific image and of “joining” the scientific image with the conceptual framework of persons pushes us to think that the scientific image is *separable* from the conceptual framework of persons and might stand alone, apart from that other framework. A careful reading of PSIM shows that this picture is too crass: Sellars acknowledges that the scientific image is not “complete” *until* it is enriched with the language of individual and community intentions.⁶ But even this acknowledgment is misleading, for my point is that any conceptual framework within which a scientific image of the world is possible must *already* be a conceptual framework containing persons and the language of individual and community intentions.

For his rhetorical purposes, it suits Sellars to emphasize the distinctness of the manifest and the scientific, but sometimes he seems to forget that it is his own considered view that «scientific discourse is but a continuation of a dimension of discourse which has been present in human discourse from the very beginning,» (EPM, §41; SPR, p. 172; KMG, p. 252). Sellars concludes from this that there is «a sense in which the scientific picture of the world *replaces* the common-sense picture» (EPM, §41; SPR, p. 172; KMG, p. 252). But this encourages the misunderstanding that science stands as a totality in opposition to the manifest image. Only in his follow-up appositive does Sellars get it right, for there he claims only that there is «a sense in which the scientific account of “what there is” *supersedes* the descriptive ontology of everyday life» (EPM, §41; SPR, p. 172; KMG, p. 252). The limitation to the “descriptive ontology” is both necessary and significant; we should not forget it.

⁵ The problem I raise here is related to the problem Bernard Williams poses for what he calls «the absolute conception of reality» on pp. 64–65 of Williams (1978).

⁶ See PSIM, par. 114; SPR: p. 40; ISR: p. 408.

There is, in my view, a subtle process/product confusion to be found in Sellars. On the one hand we have the scientific image as a conceptual framework-in-use, a norm-governed *framework* that enables certain kinds of activity, namely, theory construction, and without which that activity would not be possible. On the other hand we have the scientific image as a *product*, a detailed theoretical description of the structures of the world. Our current best examples of science as a *product*, as a theoretical description of the structures of the world, mostly make no mention of persons or norms. As a *process*, *methodology*, or framework for epistemic activity, science – current science and future science – makes full use of the concepts of personhood and of normative standards, for it includes proprieties governing experimental design, data-handling, and inference. It is an integral part of the practice of science to worry about the conduct of scientists, to apportion responsibility for creative ideas and blame for misconduct. The idea that some form of scientific image of the world is possible independently of the framework of persons, which could then be “joined” to it, has lost sight of the fact that science is primarily a human activity.

These considerations, it seems to me, cast doubt on the adequacy of the simple distinction between what is *methodologically* primary and what is *ontologically* primary that Sellars uses to regiment the relation of the manifest and the scientific images. The methodology itself has an implicit ontology. There is a subtle but, in the long run questionable, assumption that some form of description is possible that escapes all coloration from the constraints imposed by our social, subjective, finite, and ultimately practical nature. Sellars was bothered by this assumption and developed his notion of picturing as a (partial) response to it.⁷ But to my mind, Sellars did not take sufficiently into account the indispensability of the framework of intentions, persons as the subjects of intentions, and the norms that arise within communities of agents. The scientific image cannot eliminate or displace this aspect of the manifest image without pulling the rug out from under itself as well. That is, it cannot get rid of our *commitment* to the validity of the framework of persons without robbing us of the notion of validity itself. What it can do is put into a new light just what is going on in our having such a commitment.

Does this force us to retain the idea that persons are basic individuals, a separate and irreducible kind of item in our ultimate ontological catalog?

⁷ See the argument in deVries (2010).

Sellars's view of persons is really only partially Kantian. Brandom is right to emphasize that there is an equal measure of the Hegelian in it in so far as persons are not, in Sellars's view, unknowable things in themselves, but manifest realities constituted in part by relations of mutual recognition. Sellars devoted a great deal of effort to arguing that the manifest image conception of persons as sentient beings puts constraints on science that current science does not allow us to satisfy; this is one respect in which science will have to preserve and adapt to the manifest concept of persons. But the line of thinking I have focused on here, centering on the indispensability of a first-person perspective to any conceptual framework in use, exposes a different way in which science will have to preserve and accommodate a central feature of the manifest image.

IV

So where does this leave us? I have expressed dissatisfaction with Sellars's own description of the relation between the manifest and scientific images. What alternatives are there? We could reject the very idea of Sellars's distinction. Inevitably, this winds up as an endorsement of the primacy of the manifest image, for that is the framework that is already up and running, the framework within which scientific modes of thought have arisen. The scientific image cannot, it would then be thought, develop as a significantly independent conception of the world, for it could make sense only within the normative structures of the manifest image. If we accept a realistic reading of the manifest image, then we are forced to instrumentalize scientific claims that seem to conflict with manifest reality. The basicness of persons and person-level truths cannot then be threatened by developments in science. But I do not want to abandon Sellars's distinction, because it captures something important: the developments in science do challenge in fundamental ways the received conception of the world. Recent developments in science and medicine have forced us to rethink many of our beliefs about the origin and demise of persons, for instance.

We could hold on to Sellars's distinction, but revise our construal of the two images. We could, for instance, think of the scientific image primarily as a *framework for* human activity, rather than as a *product of* human activity.⁸ I

⁸ This is the thrust of recent developments in the philosophy of science. See, for instance, the work of Joseph Rouse.

think there is much to recommend this strategy, and I have moved in this direction here. Ultimately, however, we need a still more general re-assessment of the ontological enterprise.

As I mentioned earlier, Sellars himself is loose in his description of the relations between the two images but is on the right track when he talks of «the scientific account of ‘what there is’ *supersed[ing]* the descriptive ontology of everyday life» (EPM, §41; SPR, p. 172; KMG, p. 252). The phrase “descriptive ontology” is an odd one, for isn’t any ontology a description of “what there is”? Would the contrast to “descriptive ontology” be “prescriptive ontology”, where that is a statement of what there ought to be, or is it “descriptive ideology”, a listing, not of objects, but the possible descriptions of objects? Neither is correct. In consonance with the *scientia mensura*, “descriptive ontology” is the ontology implicit in the language we use to *describe and explain* the phenomena in the world. Since, according to Sellars, «persons can (in principle) be exhaustively described in terms which involve no reference to such an irreducible logical subject» (PHM, par. 95; SPR, p. 101; ISR, p. 345), persons, in Sellars’s view, will not be basic individuals in the “descriptive ontology” of the sciences.

Now, it has been argued (or at least claimed) elsewhere,⁹ there is sensible use for the notion of a prescriptive ontology here, where that is not a listing of what there ought to be, but is the ontology implicit in the language we use to deliberate about and act within the world. If, as I have claimed here, no language could be purely descriptive, independent of all normative, prescriptive, or practical elements, we are no less committed to our practical ontology than we are to our descriptive ontology. The descriptive ontology of the ideal future science may justifiably supersede the descriptive ontology of the commonsense framework in many regards, but we have no reason to believe that the practical ontology of future science will differ significantly from the practical ontology of the manifest image. As far as I can see, Sellars does not even try to give us such reason. His concern seems rather to show us how we could reasonably hold on to the language of practice (and therefore our prescriptive ontology) in the face of a radically revised descriptive ontology. This is where the Kantian/Hegelian story about functional unities, mutual recognition, and social practices comes into play. My point is that an exclusive focus on descriptive ontology seems too narrow.

⁹ See deVries (2005), chapter 10.

The manifest image, Sellars tells us, is phenomenal in the Kantian sense, – that is, not really real – but he says this because he is willing to privilege the descriptive over the practical in matters ontological. I am increasingly less inclined to try to isolate the descriptive from the practical in thought and language, much less put all of my ontological eggs in one of those baskets. I therefore join with those others who are currently rethinking the privilege of the descriptive.¹⁰ What is important is to understand the sorts and kinds we are committed to and the sorts and kinds of commitments we have to them. Thinking through Sellars's distinction helps us do that better.

REFERENCES

- Baker, L.R. (1998). The First-Person Perspective: A Test for Naturalism. *American Philosophical Quarterly*, 35(4), 327–348.
- Baker, L.R. (2007). Naturalism and the First-Person Perspective. In G. Gasser (Ed.), *How Successful is Naturalism?*. Frankfurt: Ontos Verlag, 203–226.
- Baker, L.R. (2011). Does Naturalism Rest on a Mistake?. *American Philosophical Quarterly*, 48(2), 161–173.
- deVries, W.A., & Triplett, T. (2000). *Knowledge, Mind, and the Given: A Reading of Sellars' "Empiricism and the Philosophy of Mind"*. Cambridge, MA: Hackett Publishing.
- deVries, W.A. (2005). *Wilfrid Sellars*. Chesham, Bucks, UK: Acumen Publishing, and Montreal and Kingston: McGill-Queen's University Press.
- deVries, W.A. (2010). Naturalism, the Autonomy of Reason, and Pictures. *International Journal of Philosophical Studies*, 18(3), 395–413.
- Kukla, R., & Lance, M. (2009). *"Yo!" and "Lo!": The Pragmatic Topography of the Space of Reasons*. Cambridge, MA: Harvard University Press.
- Nagel, T. (1965). Physicalism. *Philosophical Review*, 74(3), 339–356.

¹⁰ I have in mind here such philosophers as Kukla and Lance (2009) and Price (2011).

- Nagel, T. (1974). What is it like to be a bat?. *Philosophical Review*, 83(4), 435–450.
- Nagel, T. (1986). *The View From Nowhere*. Oxford: Oxford University Press.
- Price, H. (2011). *Naturalism without Mirrors*. Cambridge, MA: Harvard University Press.
- (ISR) Scharp, K., & Brandom, R.B. (Eds.) (2007). *In the Space of Reasons: Selected Essays of Wilfrid Sellars*. Cambridge, MA: Harvard University Press.
- (EPM) Sellars, W.S. (1956). Empiricism and the Philosophy of Mind. (Presented at the University of London in Special Lectures in Philosophy for 1956 under the title “The Myth of the Given: Three Lectures on Empiricism and the Philosophy of Mind”). In H. Feigl & M. Scriven (Eds.), *Minnesota Studies in the Philosophy of Science*, vol. I. Minneapolis: University of Minnesota Press, 253–329. Reprinted with additional footnotes in W. Sellars (1963), *Science, Perception and Reality*. New York: The Humanities Press, 127–196. Republished as a volume (1997) with an introduction by R. Rorty and a study guide by R. Brandom, Cambridge, MA.: Harvard University Press. Also reprinted in: deVries, W.A., & Triplett, T. (2000). *Knowledge, Mind, and the Given: A Reading of Sellars’ “Empiricism and the Philosophy of Mind”*. Cambridge, MA: Hackett Publishing.
- (PSIM) Sellars, W.S. (1962). Philosophy and the Scientific Image of Man. In R.G. Colodny (Ed.), *Frontiers of Science and Philosophy*, Pittsburgh, PA: University of Pittsburgh Press, 35–78. Reprinted in SPR and ISR.
- (SPR) Sellars, W.S. (1963a). *Science, Perception and Reality*. London: Routledge & Kegan Paul. Reissued by Ridgeview Publishing Company, Atascadero, CA, 1991.
- (PHM) Sellars, W.S. (1963b). *Phenomenalism*. In SPR; in ISR.
- (SM) Sellars, W.S. (1968). *Science and Metaphysics: Variations on Kantian Themes*. London: Routledge & Kegan Paul. Republished by Ridgeview Publishing Company, Atascadero, CA, 1992.

- (MP) Sellars, W.S. (1969). Metaphysics and Concept of a Person. In K. Lambert (Ed.), *The Logical Way of Doing Things*. New Haven: Yale University Press, 219–252.
- (SK) Sellars, W.S. (1975). The Structure of Knowledge. In H-N. Castañeda (Ed.), *Action, Knowledge and Reality: Studies in Honor of Wilfrid Sellars*. Indianapolis: Bobbs-Merrill, 295–347.
- (KTM) Sellars, W.S. (2002). *Kant's Transcendental Metaphysics: Sellars' Cassirer Lecture Notes and Other Essays*. Ed. and intr. by J.F. Sicha. Atascadero, CA: Ridgeview Publishing Company.
- Williams, B. (1978). *Descartes: The Project of Pure Enquiry*. Atlantic Highlands, NJ: Humanities Press.

The Eyes Don't Have It: Fracturing the Scientific and Manifest Images

P. Kyle Stanford *
Stanford@uci.edu

ABSTRACT

Wilfrid Sellars famously argued that we find ourselves simultaneously presented with the scientific and manifest images and that the primary aim of philosophy is to reconcile the competing conceptions of ourselves and our place in the world they offer. I first argue that Sellars' own attempts at such a reconciliation must be judged a failure. I then go on to point out that Sellars has invited us to join him in idealizing and constructing the manifest and scientific images by conflating a number of importantly distinct contrasts between heterogeneous forms of representation we employ and to argue that we are better off declining this invitation. Recognizing the important differences between these contrasts does not simply obviate the problems of integrating, connecting, and reconciling the various sorts of representations we have of various parts of the world and our own place within it, but it reveals as misguided the notion that there is just a single, fundamental problem of such reconciliation to be solved. It also suggests a potentially far more promising starting point for trying to satisfy the fundamental ambition Sellars attributes to philosophical inquiry itself.

* Department of Logic and Philosophy of Science – University of California, Irvine.

To be a naturalist is to see human beings as frail complexes of perishable tissue, and so part of the natural order. It is thus to refuse unexplained appeals to mind or spirit, and unexplained appeals to knowledge of a Platonic order of Forms or Norms; it is above all to refuse any appeal to a supernatural order. After that, the degrees of austerity that naturalism imposes can be variously interpreted: some philosophers are more relaxed than others about reconciling the world as we know it, “the manifest image”, with the world as science tells us it is, “the scientific image”. But we nearly all want to be naturalists and we all want a theory of ethics. So the problem is one of finding room for ethics, or of placing ethics within the disenchanted, non-ethical order which we inhabit, and of which we are a part.

(Simon Blackburn, *Ruling Passions*, pp. 48–49)

1. Introduction: The Instrumentalist’s Burden

Wilfrid Sellars’ famous essay *Philosophy and the Scientific Image of Man* introduced the suggestive terminology of the “scientific” and “manifest” images, along with some tantalizing thoughts on the relationship between them and the prospects for their reconciliation, and Anglophone philosophy has not been quite the same since. There seems little question that Sellars managed to put his finger on *some* kind of opposition between competing conceptions of ourselves and our place in the world that continues to provoke and fascinate us, but it can be hard to understand precisely what Sellars’ manifest and scientific images are supposed to be, and equally hard to understand what we are supposed to do with, for, or about them.

For a start, it might seem natural to expect the distinction to play a prominent role in our discussions of the status, role, and interpretation of scientific theories themselves. As the quotation above from Simon Blackburn reminds us, Sellars’ famous distinction is usually thought to contrast “the

world as we know it” from our experience with “the world as science tells us it is”. But just how to understand what it is that empirical science tells us about the world has been a contentious issue since at least the dawn of the modern scientific enterprise itself. Although the “scientific realist” view that our best scientific theories simply report how things stand in various otherwise inaccessible domains of nature remains widespread and extremely influential, a long-standing minority tradition has consistently raised challenges for any such realist view. Among other lines of concern, this contrarian minority sometimes points to the long historical record of empirically successful but ultimately abandoned scientific theories or to our repeated failure to even conceive of many scientifically serious theoretical alternatives also well-confirmed by the evidence available at any given time, and then asks why we should think that our own epistemic position is ultimately any different from that of our scientific predecessors.

Such opponents of scientific realism have sometimes suggested that instead of seeing even the most successful contemporary scientific theories as accurate descriptions of otherwise inaccessible domains of nature, we should understand them simply as conceptual tools or instruments that we can *use* to navigate our practical circumstances with lesser or greater (sometimes even astounding) degrees of success. But any such “instrumentalist” proposal faces quite a serious problem, for the intuitively appealing distinction between merely using a theory to navigate the world successfully and simply believing what it says becomes considerably murkier under closer inspection. First, one of the things we might use a theory to do is get information (and thus form beliefs) about remote and inaccessible parts or aspects of the world. Howard Stein elegantly makes this point against the claim that quantum mechanics is “merely” an instrument for predicting experimental outcomes, paraphrasing Eugene Wigner’s remark that one also «uses quantum theory, for example, to calculate the density of aluminum» (1989, p. 49). Moreover, even simply making use of a theory to predict and intervene in the world around us seems to require that we believe at least *some* of what it says about the world – that about a quarter of the next generation really will exhibit the mutant phenotype, say, or that the boiling point of a pure solvent really will rise as we add more of a (non-volatile) solute to it. Nor can we say that to make use of a theory is simply to believe the claims it makes about so-called “observables”, for what our theories say about observable parts of nature is thoroughly suffused with the terminology, conceptual apparatus, and implicit assumptions the theory

deploys in order to say anything at all. If we consider a specific claim about nature made in terms of some rejected past theory, such as “heating the red calx of mercury generates dephlogisticated air”, we are not so much inclined to insist that (all) such claims were *false* as that “calx of mercury” and “dephlogisticated air” have turned out not to be the most useful conceptual categories with which to engage this part of nature. Likewise, if one or more of our own theories are ultimately discovered to be fundamentally mistaken, we will not want to say that all or even most of their claims about mutant phenotypes, pure solvents, tectonic plates, distant nebulae, or other perfectly observable entities were false so much as that “mutant phenotypes”, “pure solvents”, “nebulae” or “tectonic plates” have turned out not to be the most productive and powerful way(s) to think about phenomena in these domains after all. Even if we retain and repurpose the *names* used for entities in rejected scientific theories, as we have done in the case of, say, “atom” and “planet” but not “germ-plasm” or “ether”, our conceptions or descriptions of the corresponding entities will be radically revised if our theories have been. Thus, thinking of the instrumental use of a theory as a matter of simply believing whatever that theory says about “observable” entities and processes will not allow us to answer this challenge for instrumentalism. We are still awaiting a fully satisfying articulation of the idea that our theories might simply be useful conceptual instruments *rather than* accurate descriptions of how things stand in the natural world.

It is in connection with such efforts to put meat on the bones of the instrumentalist idea that we might make use of one or more of our best scientific theories without simply believing what they say that Sellars’ distinction between the manifest and scientific images seems to promise potentially invaluable assistance. It is at least facially plausible to suppose that it is open to us to withhold our credence from the image of the world offered to us in scientific theorizing while nonetheless using it to predict, intervene, and otherwise usefully guide our pragmatic engagement with the image of the world that arises in our experience.¹ In what follows, I will suggest that Sellars’

¹ Perhaps surprisingly, it seems that this suggestion was *not* intended by the title of Bas van Fraassen’s extremely influential book *The Scientific Image*, notwithstanding Sellars’ evident influence on that work and on van Fraassen’s thinking more generally. In his Preface, van Fraassen writes “The title of this book is a phrase of Wilfrid Sellars’s, who contrasts the scientific image of the world with the manifest image, the way the world appears in human observation. While I would deny the suggestion of a dichotomy, the phrase seemed apt” (1980 vii). For instrumentalists seeking to distinguish the image of some part of the world given to us by one or more of our best scientific

development of the distinction cannot actually satisfy the instrumentalist's burden in the way that this natural suggestion envisions, and indeed that his own efforts to reconcile the two images he describes must ultimately be judged a failure. But I will go on to argue that this failure comes about for instructive reasons that should lead us to rethink the distinction between the manifest and scientific images that he seeks to draw, what any convincing form of instrumentalism about scientific theories would look like, and even the broader ambitions that Sellars holds out for philosophy as a whole. The place to begin, however, is surely with Sellars' own account of how the need to draw the distinction between the manifest and scientific images arises in the first place.²

2. The Aim of Philosophy and Sellars' Ocular Turn

Philosophy and the Scientific Image of Man (all Sellars quotations are from this work) opens with one of the most famous one-liners in recent philosophical history: «The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term» (p. 37). This claim is no mere rhetorical flourish, but instead announces the subject of the paper's main line of argument: the correct conception of what philosophy distinctively is and does. But Sellars goes on to elaborate this aim by making use of two recurring metaphors that are almost strikingly inequivalent. In the earliest stages of Sellars' discussion, the role of philosophy or philosophical activity is analogized to that of learning to "know one's way around" with respect to the world's wide variety of heterogeneous inhabitants, «not only "cabbages and kings", but numbers and duties, possibilities and finger snaps, aesthetic experience and death» (p. 37), of forming a sense of how the "bailiwick" of each special discipline or department of knowledge «fits into the countryside as a whole» (p. 38), and «com[ing] to know one's way around in the highway system as a whole» (p. 39). Almost immediately, however, the cultivation of such know-how is identified

theories from the image that appears in our experience, the dichotomy would seem to be the whole point.

² I make no effort in what follows to reconcile what Sellars says in *Philosophy and the Scientific Image of Man* with what he says elsewhere. Sellars is a complex figure and the various claims he makes in different texts are notoriously difficult to reconcile with one another. As will become clear, my primary concern is not with the exegesis of Sellars in any case, but with a fundamental lesson that I think can be gleaned from his attempt to introduce and delineate the manifest and scientific images in this seminal paper.

with “seeing all things together” and having one’s “eye on the whole”, initiating what I will ultimately suggest is a regrettable visual or ocular turn in Sellars’ metaphorical conception of the central task of philosophy from which it never recovers.³

This transformation is cemented by Sellars’ determined critique of the view that philosophical activity is a kind of “analysis”, a notion he supplements with the further metaphor of attempting to bring a picture into focus. Even so supplemented, Sellars insists, the implied contrast of “analysis” with “synthesis” remains extremely misleading, for at least two reasons. First, it suggests that scientists are confused about the subject matters of their own disciplines until philosophy helps to clarify them, and second, it obscures the fact that «the unity of the reflective vision [of ourselves and our place in the world] is a task rather than an initial datum» (p. 40). The scientific and manifest images enter the story as part of Sellars’ effort to elucidate this latter claim, for it is the difficulty of uniting or fusing these two images into a single coherent conception of ourselves and our place in the world that makes the unity of any such “reflective vision” an achievement which must be earned, rather than an “initial datum” we get for free (as implied by the language of analysis or the metaphor of bringing a picture into focus).

The challenge of achieving such a unified reflective vision arises, Sellars suggests, because we find ourselves simultaneously presented with «two pictures of essentially the same order of complexity, each of which purports to be a complete picture of man-in-the-world» (p. 40). Neither seems to require or even permit supplementation from the other, or from resources outside the image itself; they are, he later suggests, «two whole ways of seeing the sum of things» (p. 55). Because we find ourselves presented simultaneously with these competing and ostensibly complete scientific and manifest images of the world and our own place in it, he suggests, understanding how things hang together will involve much more than simply combining the two images into a panorama, as if they were generated by pointing a periscope in different directions: instead we will need to *reconcile* or *integrate* the two images, neither of which

³ In the essay’s second paragraph, Sellars makes a point of explicitly characterizing such “knowing one’s way around” as a form of “knowing how” rather than “knowing that” (p. 37). This also seems a curious fit with Sellars’ later turn to the ocular metaphor of seeing the task of philosophy as that of fusing two images into a stereoscopic vision, though he does insist that knowing how “at the level of characteristically human activity” presupposes a great deal of “knowledge that”.

seems even to leave room for the other, into a single coherent conception of ourselves and our place in the world around us.

But having thus rejected the language of analysis, Sellars nonetheless proceeds to replace the metaphor of bringing a picture into focus with an equally visual or ocular conception of the fundamental task of philosophy which he proceeds to treat as simply identical to the cultivation of practical know-how with which he began: our task, he repeatedly insists, is to “fuse” the separate manifest and scientific images “into one vision” (p. 41) and to achieve a «stereoscopic vision, where two differing perspectives on a landscape are fused into one coherent experience» (p. 40)

It is by no means obvious that these two metaphors amount to the same thing. To be sure, the earlier philosophical imperative is to “know our way around” not in the unreflective way that the centipede knew how to walk *before* someone inquired how it managed to do so, but in «that reflective way which means that no intellectual holds are barred» (p. 37). But even such reflective and thoughtful knowing one’s way around does not seem equivalent to having a simultaneous synoptic view of the surrounding territory: using a wide variety of tools and tricks to navigate an area (e.g., following the coastline by sight in a sailing vessel, following running water in order to consistently move downhill, maintaining your orientation towards a faraway object in order to travel in a single direction) and knowing how and even why each one works is just not the same thing as trying to fuse two different maps or pictures or other representations of an entire area into a single synoptic representation.⁴ Nonetheless, Sellars’ subsequent discussion proceeds to treat the latter metaphor as equivalent to the one with which he began, insisting that

The philosopher, then, is confronted by two conceptions, equally public, equally nonarbitrary, of man-in-the-world, and he cannot shirk the attempt to see how they fall together into one stereoscopic view. (p. 41)

⁴ Sellars uses a similar alternation of metaphors to characterize the specialist’s sense of how the subject matter of her own department of knowledge fits into the “intellectual landscape” as a whole (e.g., pp. 38–40, *passim*.) as well as the philosopher’s concern to know her way around that landscape and/or achieve a synoptic view of it. It is not always clear when Sellars means to refer to this philosophical ambition rather than that of knowing one’s way around the various entities and relations actually making up the world itself (or indeed, whether he thinks these are genuinely distinct), but our primary concern will be with the latter in any case.

By this point Sellars' ocular or visual turn is complete. He never returns from the metaphor of stereoscopic vision to that of "knowing one's way around" the world and its inhabitants with which he began.⁵

3. Surveying the Images

What precisely, then, are these scientific and manifest images that prove so difficult to integrate or reconcile into a single "stereoscopic" view? We should not imagine that Sellars is proposing an atavistic picture on which the manifest image is somehow composed of sense data or isolated phenomenal experiences or some such and the scientific image represents the conceptual superstructure we use to navigate and anticipate such sense data. Such a picture would be implausibly attributed to Sellars (himself a famous critic of sense-data theories of knowledge and the "myth of the given" more generally) in any case, but he makes a point of rejecting any such view as an inadequate description of the manifest image itself: «whether or not the world as we encounter it in perception and self-awareness is ultimately real,» he insists, «it is surely incorrect [...] to say as some philosophers have said that the physical objects of the encountered world are "complexes of sensations"» (p. 51). Even more importantly, such an attribution would abjectly fail to make sense of Sellars' insistence that the manifest image no less than the scientific «purports to be [...] the *whole* truth about that which belongs to the image» (p. 57). Even if there were ultimately some sense to be made of the notion of a collection of raw sense data or atheoretical perceptual experiences, such a collection would not even purport to be and would not present itself phenomenologically as a complete picture of the world and our place within it.⁶

⁵ Sellars later seems to suggest at one point (but only one, p. 55) that such "knowing one's way around" has been the task only of the "perennial tradition" in philosophy (which is in turn identified with exploration and development of the manifest image). If we take this suggestion seriously, however, it implausibly entails that Sellars' famous proposed aim for philosophy "to see how things in the broadest sense of the term hang together in the broadest sense of the term" also applies only to the perennial tradition and/or to philosophical activity as conducted within the manifest image (cf. p. 37). However that may be, I will ultimately suggest that the idea of coming to "know one's way around" offers a better guide for our efforts to integrate and reconcile the manifest and scientific images than any of the ocular or visual metaphors Sellars offers in its place, whether or not Sellars is himself switching horses at just this point.

⁶ Nor does it seem that such a collection could be continually "refined" or "sophisticated" in the way Sellars describes below, so as to itself constitute a kind of "scientific image".

It might instead seem obvious from the language of “scientific” and “manifest” images alone that Sellars means to contrast the conception of man and his place in the world that we get from sophisticated scientific inquiry with the conception that appears instead in common sense, or that existed before the rise of modern science, or that is embodied in various “folk” theories concerning these matters. However, while Sellars does sometimes characterize the manifest image as “sophisticated common sense” (e.g., p. 57), he also clearly and emphatically denies that the contrast he has in mind «is that between a pre-scientific, uncritical, naïve conception of man-in-the-world, and a reflective, disciplined, critical – in short a scientific – conception,» nor is the manifest image that found in «an historical and bygone stage in the development of man’s conception of the world and his place in it» (pp. 42–43). Instead, he says, «what I mean by the manifest image is a refinement or sophistication of what might be called the “original” image...» (p. 43). This sophisticated manifest image is “disciplined and critical”, and it «makes use of those aspects of scientific method which might be lumped together under the heading “correlational induction”». Perhaps surprisingly, then, it is not the *scientific* character of the scientific image which distinguishes it from the manifest image, for «the manifest image is, in an appropriate sense, itself a scientific image» (p. 43).

Nonetheless, Sellars goes on to note that

[t]here is... one type of scientific reasoning which it, by stipulation, does *not* include, namely that which involves the postulation of imperceptible entities, and principles pertaining to them, to explain the behavior of perceptible things. (p. 43)

He even goes so far as to claim that the scientific image «might be better called the “postulational” or “theoretical” image» (p. 43).⁷ And Sellars later repeats that this is the fundamental difference between the two images even as he is careful to remind us that this is not at all the same as the difference between a scientific and unscientific conception:

the contrast I have in mind is not that between an *unscientific* conception of man-in-the-world and a *scientific* one, but between that conception which limits itself to what correlational techniques can tell us about perceptible and introspectible events and that which postulates imperceptible objects and

⁷ He then goes on to say «[b]ut, I believe, it will not be too misleading if I continue, for the most part, to use the former term» (p. 43). I am not sure history has borne out his confidence on this point.

events for the purpose of explaining correlations among perceptibles. (p. 56)

Sellars' view seems to be that our efforts to find correlations between perceptible and introspectible aspects of our experience has produced an increasingly sophisticated and refined version of the "original" image of ourselves and our place in the world, and this is to be usefully contrasted with a competing image generated from practices of scientific theorizing or postulation of "imperceptible objects and events". We cannot, however, think of the manifest image as encapsulating our knowledge of perceptible entities or phenomena and the scientific image as encapsulating our knowledge of a completely distinct or disjoint realm of imperceptible or unobservable entities without again doing violence to Sellars' insistence that each image claims to provide a *complete* picture of the world and our place within it. A more promising alternative is suggested by his description of the two images as «different *perspectives* on a landscape that are fused into a single coherent experience» (p. 40, my emphasis). This description does not, of course, suggest a division of labor in which different observers contemplate disjoint parts or regions of a single landscape (say, the perceptible and imperceptible regions), but instead one in which two observers have distinct views of *one and the same set of objects* due to something like a difference in those observers' respective vantage points.

On this conception, most of the world's inhabitants can be located within both the manifest and the scientific images, but the two images conceive of them quite differently and tell us very different sorts of things about them: they represent different aspects or dimensions of what we know about (and how we know about) cabbages, kings, finger snaps, and all the rest. On such a view, cabbages belong neither to the scientific image nor to the manifest image, but our conception of them includes aspects that belong to the scientific image (such as their evolutionary history, their role in an economy or an ecosystem, their chemical composition or nutritional value for human beings) and aspects that belong instead to the manifest image (the way they look and smell; how you might use one to prop open a door; the extent to which eating the one in your hand would alleviate your hunger). In fact, it would seem that these identifications of entities in the scientific and manifest images are the *products* of our efforts to fit the two images together, the enterprise to which Sellars seeks to contribute.

But in addition to capturing distinct aspects of a common set of entities, each image would also seem to include entities that simply do not appear in the

other at all. In the case of the scientific image, these will be entities or aspects of entities that the image itself tells us are imperceptible even as it offers us theoretical resources for their detection and/or for learning much else about them: contemporary cosmology and perceptual psychology together tell us that human beings are unable to perceive dark matter, for example, even as we use our perceptual contact with other physical features of the universe to calculate (assuming the fundamental truth of the further theories upon which we rely to do so) how much there is and where it resides. More generally, when the scientific image itself tells us that particular entities or aspects of entities that it posits are sufficiently removed in some way (e.g., causally or temporally) from the perceptual apparatus of human beings, we regard those entities as “mere posits” or “imperceptible” even when we see ourselves as having sophisticated procedures for detecting their presence, magnitude, frequency, etc. in particular contexts by means of similarly hypothesized causal influences on or contact with features of our own perceptual states. By the scientific image’s own lights, Cherenkov radiation is perceptible by human beings but neutrinos are not.

In addition, however, Sellars’ conception of the manifest image famously includes such items as duties, intentions, justifications, and other occupants of what he elsewhere calls the “conceptual space of reasons” that do not presently seem to find counterparts in the scientific image.⁸ This is in part why Sellars argues that «there is an important sense in which the primary objects of the manifest image are *persons*» (p. 46), and that «what the objects of this framework, primarily *are* and *do*, is what persons are and do» (p. 48). He describes the manifest image as the subject of the “perennial” project of self-understanding with which most of the philosophical tradition has characteristically been concerned, and he insists that the contemporary manifest image has been formed by successive modifications of the “original image” of human self conception, «the modification consisting of a gradual pruning of the implications of saying with respect to what *we* would call an inanimate object, that it *did* something» (p. 49). Whether or not Sellars is right to think that the contemporary manifest image has indeed been generated in whole or in part by such a process of gradual depersonalization – in which

⁸ Given his articulation of the scientific and manifest images, of course, this would seem to mean not that intentions, duties, justifications, and linguistic meanings have no place in a scientific conception of ourselves and our place in the world, but rather that Sellars thinks we do not make sense of them by positing imperceptible entities.

«[n]ature became the locus of “truncated persons”» (p. 49) –, it seems clear that the manifest image is intended to include entities and phenomena that simply do not arise in the putatively complete description offered by the scientific image, including the sorts of normatively loaded phenomena of which we seem (at least at present) to have only a subjective apprehension.⁹ And as we will see, aspects of the manifest image that do not appear at all in the scientific image are among the most serious challenges for Sellars’ own efforts to “fuse” the two images into a single “stereoscopic view”.

4. Irreconcilable Differences?

How does Sellars himself seek to reconcile the two images he has described? Although he will ultimately embrace what he calls the «primacy of the scientific image» (p. 69), Sellars explicitly denies that we should simply appeal to the scientific image to explain what we can in the manifest image and dismiss whatever remains as error and illusion. Instead, as he notes, «the very fact that I use the analogy of stereoscopic vision implies that as I see it the manifest image is not overwhelmed in the synthesis» (p. 45), and his explicit intention is to identify «the respective contributions of these two [images] to the unified vision of man-in-the-world which is the aim of philosophy» (p. 55). Sellars goes on to claim that the «major stresses and strains involved in any attempt at a synoptic view» are made evident by considering the three fundamental possibilities that emerged in connection with the early modern attempt to «construe physical things, in a manner already adumbrated by Greek atomism, as systems of imperceptible particles, lacking the perceptible qualities of manifest nature» and with Descartes’ own efforts to synthesize this scientific view with the existing manifest image (p. 62). The remainder of Sellars’ discussion is largely devoted to navigating among the three possibilities: rejecting the first, studiously ignoring the second, and trying desperately to rehabilitate the third against what appear to be insuperable obstacles.

The first possibility Sellars considers is that manifest objects are simply identical with systems of imperceptible particles, «in that simple sense in which a forest is identical with a number of trees» (p. 62). He quickly and firmly rejects this possibility, however, on the ground that manifest objects have

⁹ Note that this alone implies that the scientific image is not formed simply by *adding* methods of hypothetical postulation to resources of the existing manifest image – if it were, the scientific image would contain everything that the manifest image contains and more besides.

properties which cannot be regarded as simply a matter of the imperceptible particles making them up themselves having particular properties and/or being related in particular ways: a pink ice cube in the manifest image, for example, presents itself to us as pink «through and through, as a pink continuum, all the regions of which, however small, are pink» rather than such a property being «made up of imperceptible qualities in the way in which being a ladder is made up of being cylindrical (the rungs), rectangular (the frame), wooden, etc.» (p. 63). That is, although a manifest object need not be (and typically is not) homogeneous in its color, it is “ultimately homogeneous” with respect to the property *of being colored* in a way that cannot be recapitulated by or identified with properties of its (imperceptible particulate) constituents or relations among them. The thought here seems to be that objects in the scientific image do not (or do not uniformly) have the right sorts of properties to serve as compositional parts of the *sorts* of objects we actually encounter in the manifest image.

This recognition, Sellars suggests, is what motivates the third possibility: that manifest objects «are “appearances” to human minds of a reality which is constituted by systems of imperceptible particles» (p. 63). Although he identifies serious challenges for this alternative (see below), it is the view he will ultimately seek to defend. But Sellars’ treatment of the second line of thought regarding the possible relationship between the manifest and scientific images is exceedingly curious. This second line of thought is the (broadly instrumentalist) notion that «[m]anifest objects are what really exist, systems of imperceptible particles being “abstract” or “symbolic” ways of representing them» (pp. 62–63). Sellars notes that this option «merits serious consideration, and has been defended by able philosophers,» but then proceeds to gently set it aside and discuss only the first and third options instead (p. 63f). Even after the serious challenges he encounters in his defense of the third option lead him back around to posing this instrumentalist possibility a second time, he offers no argument against it but again simply sets it aside and proceeds to try to rehabilitate the third option instead (p. 69f).

We will ultimately return to Sellars’ curious failure to engage this second, “instrumentalist” option, but let us first consider his efforts to overcome the challenges for the third option, the view that manifest objects «are “appearances” to human minds of a reality which is constituted by systems of imperceptible particles» (p. 63). The central problem with thus construing objects and processes in the manifest image, he argues, lies in making sense of

the qualitative character of such appearances in cases like sensation and conceptual thought. Even if we can identify «complex states of the brain, which, obeying purely physical laws, would resemble and differ from one another in a way which corresponded to the resemblances and differences between the conscious states with which they were correlated» (p. 67) he argues, the following challenge remains:

We have pulled perceptible qualities out of the physical environment and put them into sensations. If we now say that all there really is to sensation is a complex interaction of cerebral particles, then we have taken them out of our world picture altogether. We will have made it unintelligible how things could even *appear* to be colored. (p. 67)

In a similar fashion, «the claim that “thoughts, etc., are complex neurophysiological processes” leaves us with the problems of accounting for the *introspectible qualities* of thoughts» (p. 68). It is this challenge, he suggests, which underlies the persistent temptations of both dualism and the instrumentalist conception of the scientific image that he has set aside without argument.

Sellars' attempt to rehabilitate the third option begins by insisting that the problems posed by the need to account for the status of sensation and of conceptual thought are in fact quite different, and call for different kinds of solutions. In the case of conceptual thought, he suggests, the demand arises from «the mistake of supposing that in self-awareness conceptual thinking presents itself to us in a qualitative guise» (p. 69). We mistakenly suppose that thoughts are like sensations in this way, Sellars suggests, because the two are actually alike in a different but closely related way: our knowledge of sensations and of our thoughts are both non-inferential in character. After suggesting that we conceive of our own thought largely by analogy to overt expression in speech, he goes on to insist that all we properly (and non-inferentially) know in the case of our thoughts is that «*something analogous to and properly expressed by* [a given sentence] *is going on in me*» (p. 70; cf. also p. 61). That is, Sellars thinks we have mistaken the non-inferential character of our knowledge of our own thoughts for a qualitative character directly available to introspection; in fact we have no qualitative experience of our own thoughts, and conceptual thought remains eligible for identification with one or more complex physical processes. As he (somewhat confusingly) puts the point, he has sought to make room for «the possibility that the inner state conceived in

terms of this analogy [to overt speech] is *in its qualitative character* a neurophysiological process» (p. 70).

This bare possibility is supplemented, however, with the further suggestion that the identifying features of thought do not concern any intrinsic quality at all, but instead the *role* that such thoughts play in a larger system (like that of speech, computation, and neurophysiological organization). Much as a pattern of such relationships and relative roles (rather than any intrinsic characteristic(s)) are what make a particular movement of pieces (no matter how it is physically implemented) a “castling” in a game of chess, Sellars suggests,

our concept of “what thoughts are” might, like our concept of a castling is in chess, be abstract in the sense that it does not concern itself with the *intrinsic* character of thoughts, *save as items which can occur in patterns of relationships which are analogous to the way in which sentences are related to one another and to the contexts in which they are used.* (p. 71)

Moreover, «if thoughts are items which are conceived in terms of the roles they play, then there is no barrier *in principle* to the identification of conceptual thinking with neurophysiological processes. There would be no “qualitative” remainder to be accounted for» (p. 71). Putting together the denial that conceptual thought involves a distinctive qualitative character available to introspection with the positive claim that it is not *any* intrinsic feature or character but rather the respective roles they play that constitute thoughts as such, Sellars thus seeks to fuse these parts of the manifest and scientific images by making room for thoughts to simply be identified with complex neurophysiological states of the brain.

The case is quite different, Sellars argues, when it comes to sensation. While in the case of thought the analogy to a publicly observable item (speech) «concerns the *role* and hence leaves open the possibility that thoughts are radically different *in their intrinsic character* from the verbal behavior by analogy with which they are conceived,» by contrast «in the case of sensations, the analogy concerns the *quality* itself» (p. 72). Sensations *are* constituted by their intrinsic, qualitative characters, and nothing in the intrinsic character of a neurophysiological process will count as sufficiently analogous to make the sort of identification available in the case of thought even moderately plausible. The crucial qualitative feature we encounter in sensation is what Sellars earlier called their “ultimate homogeneity” – the fact that a pink ice cube is pink throughout all of its constituent parts, no matter how small – and it is this

feature which prevents us, he insists, from *identifying* (rather than merely correlating) sensations with corresponding neurophysiological processes or with any other part of our present scientific image. Lest we despair, however, Sellars reminds us that «the scientific image is not yet complete; we have not yet penetrated all the secrets of nature» (p. 74). Perhaps we will find that particles are not in fact “the primitive entities of the scientific image” and that «*when it comes to an adequate understanding of the relation of sensory consciousness to neurophysiological process, we must penetrate to the nonparticulate foundation of the particulate image*» (pp. 74–75).¹⁰

It is hard to know what to make of this seemingly desperate appeal to the explanatory capabilities of some possible future physics in which the “ultimate homogeneity” Sellars finds characteristic of sensations is also exhibited (somehow!) by the imagined nonparticulate constituents of physical particles or in which the properties of such constituents diverge radically from those of familiar physical entities in some other way that (somehow!) makes their actual identification with sensations seem plausible. Indeed, it seems difficult to regard this hopeful invocation of an imaginary physics so different from our own as anything more than whistling past the graveyard or a pious counsel of desperation: Sellars appears to have painted himself into a corner from which there seems little hope of any dignified escape.

5. Learning From Failure: An Instructive Diagnosis?

This might seem the natural point in the tale for instrumentalism to sweep in and save the day, shaming Sellars’ abortive earlier treatment of it and rescuing us from his somewhat sheepish appeal to an imaginary future physics. But such a triumphant rescue is simply not in the cards for reasons that turn out to be important and revealing.

Return for a moment to Sellars’ earlier consideration of what I called the broadly instrumentalist option that «[m]anifest objects are what really exist, systems of imperceptible particles being “abstract” or “symbolic” ways of

¹⁰ In the final section (pp. 75–78) Sellars goes on to sketch a program for extending his attempted reconciliation to incorporate human action under various kinds of *standards* (ethical, logical, etc.) by construing action under any description employing such standards as incorporating an assertion of (actual or potential) membership in a community and being subject to the collective intentions of that community. This seems a plausible enough line to take regarding such standards (sharing at least the spirit of some existing noncognitivist views in ethics), but our immediate concerns can be served without considering it.

representing them» (pp. 62–63). We can perhaps begin to see why Sellars finds this line of thought so implausible or distasteful when he later expands this brief description of the instrumentalist option: the possibility he invites us to consider is that

[R]eality is the world of the manifest image, and that all the postulated entities of the scientific image are “symbolic tools” which function (something like the distance-measuring devices which are rolled around on maps) to help us find our way around in the world, but do not themselves describe actual objects and processes. (p. 69)

As Sellars quite rightly notes,

On this view, the theoretical counterparts of *all* features of the manifest image would be *equally* unreal, and that philosophical conception of man-in-the-world would be correct which endorsed the manifest image and located the scientific image within it as a conceptual tool used by manifest man in his capacity as a scientist. (p. 69)

This view is indeed implausible and Sellars is perhaps right to simply dismiss it outright, but his own italics inadvertently reveal where the real source of the implausibility lies: the most unpromising aspect of this proposal is its presumption that we must either be instrumentalists about *all* “theoretical counterparts” of features of the manifest image or none of them. In other words, what Sellars (sensibly) sets aside is the view that the *entire* scientific image is nothing but a massive calculational tool we use to mediate our engagement with the “real” world we encounter in the manifest image.

This is indeed an unappealing prospect, for a wide variety of reasons that certainly includes the consequence Sellars points to above: it requires us to reject *all* posited or hypothesized entities or aspects of the world as unreal. Even those who offer the most cogent and pressing reasons for doubting that our best scientific theories are simply accurate reports of how things stand in otherwise inaccessible natural domains can usually identify some constituents of the scientific image, even some “imperceptible” theoretical posits (paramecia? dinosaurs? continental plates?), whose description in the scientific image they think of as something more than merely cogs in a gigantic calculational tool for moving from some observable states to others. But there is no reason we must choose between being instrumentalists about every hypothesized or imperceptible entity and/or aspect of nature or being instrumentalists about none of them. Indeed, even those with strongly realist

sensibilities are often tempted to adopt some variety of such an instrumentalist position concerning particular theories. In the case of a theory like quantum mechanics, for example, our theoretical descriptions of fundamental physical entities as “simultaneously particle-like and wave-like” or as occupying “superpositions” of classical states like having determinate positions or momenta seem likely to reflect as much about the constraints or limits on the ways we are able to conceptualize the natural world as they do about the fundamental constitution of that world itself. There is simply no ground for insisting that we must choose between adopting an instrumentalist attitude towards every hypothesized or imperceptible entity and/or aspect of nature or towards none of them.¹¹

This realization suggests that Sellars has approached his inquiry with a fundamental presupposition that deserves to be questioned, namely that there is just *one* monolithic and homogeneous way in which the scientific and manifest images are related to one another, and thus some single way in which each component part of each image is related to some corresponding part (if such there be) of the other. But to recognize this presumption is to see why we have little reason to embrace it. Few scientific instrumentalists wish to regard *everything* in the scientific image as merely instrumental, few scientific realists want to extend their realism to absolutely everything in the scientific image, and few sensible people think upon serious reflection that the relation of every “theoretical”, “postulated”, or even “imperceptible” entity or aspect of nature to our manifest experience must be just the same as that of every other. The sensible prospect of being instrumentalists about *some* scientific posits or theoretical claims and not others, shared by many self-described realist as well as nonrealist philosophers of science, illustrates why we have every reason to doubt that “the” relationship between the manifest and scientific images has the sort of monolithic and homogeneous character that Sellars supposes it must or should. The first lesson to draw from Sellars’ failure, then, is not so much that he was too quick to dismiss instrumentalism as that it was a mistake

¹¹ Intriguingly, Sellars’ own description of a *nonparticulate* possible future physics is one on which particles «could be treated as singularities in a space-time continuum which could be conceptually “cut up” without significant loss – *in inorganic contexts, at least* – into interacting particles» and in which «for many purposes the central nervous system can be construed without loss as a complex system of physical particles» (pp. 74–75). This would seem to articulate (albeit vaguely) just the sort of instrumentalism regarding a particular scientific description of a particular domain of nature (here fundamental physical “particles”) whose broader possibility Sellars seems to ignore.

for him to go looking for just *one* relationship between the scientific and manifest images in the first place.

It may be that Sellars has been betrayed into this presumption by what I suggested at the outset was an unfortunate metaphorical transformation: the shift from the demand to “know our way around” the world and all its many heterogeneous inhabitants to that of “fusing” two images into a “stereoscopic vision”. The metaphor of fusing the manifest and scientific images into a single stereoscopic view almost inexorably suggests that there should be just a single monolithic relationship between the corresponding parts of the two images, for this is indeed how composite images are formed in stereoscopic vision. In effect, then, it obscures the possibility that the relationship between the manifest and scientific images might be heterogeneous in character, for that possibility finds no counterpart in the metaphor of stereoscopic vision. But as the prospect of selective instrumentalism about theoretical posits vividly illustrates, we have every reason to recognize the possibility (even the plausibility) of heterogeneity in the ways that different theoretical posits are related to the world we encounter in experience, perception, or common sense. We have substantial reasons to doubt that such theoretical posits as superpositions, gluons, and Newtonian gravitational forces, all seemingly good candidates for us to regard as mere “useful fictions”, are related to the manifest image in just the same way(s) as paramecia, dinosaurs, or continental plates, but the metaphor guiding Sellars’ inquiry leaves no room for a difference between them.

Once we have noticed this room for heterogeneity in the relationship that different elements in the scientific and manifest images bear to one another, however, our attention is inevitably drawn to a further respect in which Sellars’ picture of the situation seems insidiously oversimple. For part of what makes it so plausible to suppose that different elements in the scientific image bear distinct relationships to their counterparts in the manifest image is the fact that Sellars seems to have been making use of more than just *one* opposition between competing conceptions of ourselves and our place in the world all along. As we’ve seen, Sellars repeatedly emphasizes that the most fundamental difference between the manifest and scientific images is that the former is limited to sophisticated techniques of correlation while the latter admits postulational techniques as well. But at other points (such as the discussion of the introspectible characteristics of thoughts and sensations) the primary contrast seems to be between egocentric, imagistic, first-personal, and

perspectival ways of representing ourselves and our place in the world and more abstract, third-personal, aperspectival forms of such representation. Still elsewhere the central point of contrast is asserted to be between a refined modern descendant of the “original” (or “perennial” or commonsensical) conception of ourselves and our place in the world and that which arises only after the dawn of modern science. But few contemporary philosophers would be willing to follow Sellars in simply identifying the central terms on each side of these very different contrasts with one another. Those who accept Quine’s influential contention that the tables, chairs, and kittens familiar to us from common sense and everyday experience are no less theoretical posits introduced to systematize and explain our experiences than are quasars and electrons, for example, will vigorously resist the suggestion that the conception of ourselves and our place in the world offered by «sophisticated common sense» (p. 57) or «a refinement or sophistication of what might be called the “original” image» in terms of which man first encountered himself (pp. 42–43) or «the perennial philosophy of man-in-the-world» (p. 44) is purely correlational or does not traffic in theoretical postulation.¹² Nor will many contemporary philosophers be tempted by the idea that either the correlational/postulational contrast or that between a sophistication of the “original” image and that made possible by the rise of modern science lines up neatly with the contrast between the sorts of concrete, imagistic, egocentric, perspectival representations of parts of the world that we receive more-or-less automatically from our senses when we survey a scene and the more abstract, theoretical, nonegocentric and aperspectival representations we make a conscious effort to construct.¹³

To be sure, there are deep and important puzzles concerning how the sorts of representations on each side of each of these contrasts are to be integrated, related or connected to those found on the other, but it is simply not the case that the puzzles are the same in each case. That is, the challenges concerning how to connect or integrate the sorts of concrete, consciously mediated, egocentric, perspectival representations of parts of the world that are

¹² Of course, Sellars’ opposition explicitly described the theoretical posits of the scientific image as imperceptible, but this seems simply to leave no room at all for theoretical or postulated entities that are also perceptible.

¹³ Of course, our brains do automatically integrate information from many different sources into a single egocentric perceptual representation of the world around us, but this integrated egocentric perceptual representation cannot hope to exhaust the many different roles and characteristic features Sellars attributes to the idealization he christens as the manifest (or “original”) image.

presented to us by our senses with more abstract, theoretical, nonegocentric and aperspectival representations we make a conscious effort to construct are simply not the same as those of relating or integrating representations relying only on correlations with those involving theoretical postulation as well, or those involving only perceptible objects and processes with those deploying imperceptible ones as well, or a refined modern descendant of the “original” (or “perennial” or commonsensical) conception of ourselves and our place in the world with that which arises after the dawn of modern theoretical natural science. But Sellars seems to slide freely between these contrasts in describing the opposition between the manifest and scientific images.

It is revealing, then, that Sellars readily concedes that the opposition between the manifest and scientific images does not even arise until after each of those images has been *constructed* by a self-conscious process of idealization and abstraction from the welter of heterogeneous forms of engagement and representation with which we encounter the world. Sellars repeatedly describes the manifest and scientific images as themselves “idealizations” (p. 41; cf. p. 43), “ideal constructs” (p. 56), and “poles to which philosophical reflection has been drawn” (p. 44). Specifically with regard to the contrast between correlational and postulational methods, he allows that the idea of a purely correlational scientific view is «both a historical and methodological fiction» (p. 43), because in reality our scientific worldview has been formed by a complex interplay in which both sorts of methods «have gone hand in hand» and been «dialectically related», with «postulational hypotheses, presupposing correlations to be explained and suggesting possible correlations to be investigated» (p. 43). Sellars allows that the manifest image he has constructed is simply “a useful fiction” formed by “abstracting correlational fruits from the conditions of their discovery”, and it is claimed to be “no *mere* fiction” only because it enables us to bring the contrast with the scientific image into sharper view (p. 43). But the same work of abstraction and construction is, if anything, even more evident in the case of the scientific image, even by Sellars’ own lights: the manifest image, we are told, allows us to

[D]efine a way of looking at the world, which, though disciplined and, in a limited sense, scientific, contrasts sharply with an image of man-in-the-world which is *implicit* in and can be *constructed* from the postulational *aspects* of contemporary scientific theory. (p. 43; my emphasis)

The scientific image is constructed in one further way as well, by the imagined integration of many different and partial theoretical pictures of different scientific domains or subject matters: «Thus the conception of *the* scientific or postulational image is an idealization in the sense that it is a conception of an integration of a manifold of images» (p. 56). Sellars is not much troubled by this dimension of the problem of constructing a single scientific image of the world, but our less reductionistic age has learned to be considerably more circumspect about simply assuming that the various sorts of representations found in different parts of theoretical science can be smoothly integrated with one another.

It seems, then, that we are not *actually* confronted with any fundamental opposition or contrast between the scientific and manifest images unless and until we ourselves have done quite a bit of idealizing, abstracting, and constructing in order to generate these two competing images (or more properly, simply the ideas of these two competing images) from the materials of experience and scientific representation. But this implies first, that the demand to reconcile these two images is not a task set for us by our encounter with ourselves and the world but rather one that Sellars suggests we set for ourselves by idealizing, abstracting, and constructing our way to these two images in the first place, and second, that it is simply an article of faith on Sellars' part that the contrasts involved in the many different oppositions between which he slides can all be made to line up neatly into just two fundamental "master" representations of the world and our place within it. In other words, Sellars *invites* us to follow him in idealizing, abstracting, and constructing our way to a contrast between these two master representations, and it is by no means clear that we should accept the invitation.

Indeed, Sellars has pulled off a sort of conjuring trick, for whenever we go looking for particular examples of representations contrasting in *one or another* of the various ways he characterizes the manifest and scientific images respectively, suitable candidates can nearly always be located, but this does not show that Sellars has managed to identify exactly two "pictures of the world" that must be stereoscopically "fused". The ease with which we can identify such candidates in particular cases of contrasting representations is perhaps well illustrated by Sellars' appeal to Eddington's famous "two tables" as representatives of the manifest and scientific images respectively (p. 73), but it is a substantial further step to suppose that the representations on each side of *each* of the various oppositions so easily illustrated jointly constitute a single

comprehensive “image” of the world and our place within it that competes with a comparable single image jointly constituted by the other side of each opposition. And that further supposition appears to be a substantive as well as an historical and methodological fiction: we are not in fact faced with just one problem of fitting together distinct and sometimes competing conceptions of ourselves and our place in the world, but with many different such problems.¹⁴

All this suggests in turn that the broad ambition Sellars holds out for philosophy as a whole, to “understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term”, may be much better served by his original metaphor of learning to “know one’s way around” to than by the alternative of stereoscopic vision to which he so quickly shifts following what I called a visual or ocular turn. That is, instead of ascribing to philosophy the job of first creating and then reconciling two fundamental competing images of the world and our own place within it, these reflections suggest that the broad ambition Sellars holds out for philosophy is actually better served by viewing its charge as that of learning to navigate among the *various* kinds of representations we encounter and deploy of cabbages, kings, numbers, duties, possibilities, finger snaps, aesthetic experience, and death, with no expectation of any single uniform or homogeneous relationship between those representations. In forming a sense of how the “bailiwick” of each special discipline or department of knowledge «fits into the countryside as a whole» (p. 38), philosophical inquiry might be better seen as investigating how the various and multiply heterogeneous *actual* ways in which we represent parts or aspects of the world and our own place within it fit together and are related or connected to one another, rather than first idealizing away from those actual representations so as to create two fictionalized master representations and then asking what single homogeneous relationship those two idealized or fictionalized representations could even possibly bear to one another. In light of the sort of heterogeneity among representations we have encountered, this seems a far more promising starting point for reflectively illuminating and even improving the distinctive sort of

¹⁴ I should perhaps acknowledge explicitly that I have not contributed anything here to what Sellars seems to regard as the most intractable of these problems: finding room for qualitative conscious experience in our theoretical scientific conception of ourselves. My point is rather that we will not gain traction on the many different problems of such reconciliation we face by insisting that they all arise in the course of trying to fuse two fictional “master” representations, or that there must be a single homogeneous relationship between those master representations.

“know-how” that allows us to navigate *between* and *among* the various sorts of representations of the world and our place in it that we use to make our way in that world.

Moreover, this recognition invites us to take a somewhat different view of the instrumentalist’s burden with which we began. No longer does it seem natural to ask the instrumentalist what it would be like to take the picture of the world given to us by theoretical science as simply a powerful cognitive tool or instrument for guiding our practical engagement with the world, for we have lately been reminded that there is no such picture of the world. Instead, we regard some *particular* scientific theory about some particular part or aspect of nature as such an instrument. When we do so, we use it to predict, intervene, and otherwise guide our pragmatic engagement with inhabitants of the world *as* those inhabitants can be understood in terms of *other* representations (including perhaps other theories and even other scientific theories) whose ontological and other descriptive commitments we do straightforwardly embrace. Indeed, this is just the instrumental use that even scientific realists make of Newtonian mechanics, illustrating that we are *all* instrumentalists in this sense, and those on opposite sides of the question of scientific realism have historically differed not in whether they take up instrumentalist commitments at all but in their view of just which theories are the ones towards which such an attitude should be adopted. Elsewhere (2006, Ch. 8) I have tried to give a somewhat more detailed account of what this form of instrumentalist view would look like, but the most important point for present purposes is that regarding a particular scientific theory simply as a useful cognitive device for navigating with respect to other representations of the world and our place in it will represent just *one* of the many different ways in which our heterogeneous (scientific and nonscientific) representations of various parts or aspects of the world can be systematically related, connected, or integrated with one another. Seen in this way, it seems quite right for Sellars to suggest that learning to reflectively navigate among those interconnected representations and to “know our way around” with respect to subjects of our representations as diverse as cabbages, kings, numbers, duties, and all the rest is indeed a special concern of systematic philosophical inquiry, and thus that at least *one* fundamental aim of philosophy (though surely not the only one) is indeed to «understand how things in the broadest sense of the term hang together in the broadest sense of the term» (p. 37).

REFERENCES

- Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. New York: Oxford University Press.
- Sellars, W. (1962). Philosophy and the Scientific Image of Man. In R.G. Colodny (Ed.), *Frontiers of Science and Philosophy*, Pittsburgh, PA: University of Pittsburgh Press, 35–78.
- Stanford, P.K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.
- Stein, H. (1989). Yes, but...Some Skeptical Remarks on Realism and Anti-Realism. *Dialectica*, 43(1–2), 47–65.
- Van Fraassen, B. (1980). *The Scientific Image*. New York: Oxford University Press.

Identifying and Reconciling Two Images of “Man”

David Hodgson *
raeda@tpg.com.au

ABSTRACT

Fifty years ago the philosopher Wilfred Sellars identified two images of “man”, which he called respectively the “manifest image” and the “scientific image”; and he considered whether and how these two images could be reconciled. In this paper, I will very briefly look at the distinction drawn by Sellars and at his suggestions for reconciliation of these images. I will suggest that a broad distinction as suggested by Sellars can indeed usefully be drawn, but that the distinction can be more helpfully characterised than it was by Sellars. I will argue that there are more ways of reconciling the two images than those proposed by Sellars. And I will elaborate on what I think are the most promising lines along which the reconciliation could take place.

Sellars’ Distinction and Proposed Reconciliation

In his article *Philosophy and the scientific image of man*, Sellars (1963) identified two broad conceptual frameworks in terms of which human beings conceive of themselves and their place in the world.

One he called the “manifest image”, being the framework in terms of which human beings first became aware of themselves, and in terms of which they ordinarily conceive of and explain themselves and their place in the world. According to Sellars, this framework is not necessarily naive or unsophisticated, but on the contrary could be and indeed has been the subject of highly rational and sophisticated elaboration. However, according to Sellars, this image wholly excludes «the postulation of imperceptible entities, and principles pertaining to them, to explain the behaviour of perceptible things» (1963, p. 7).

* Recently retired as a Judge of Appeal of the Supreme Court of NSW, Australia.

The other he called the “scientific image”, the defining feature of which, according to Sellars, is that it «postulates imperceptible objects and events for the purpose of explaining correlations among perceptibles» (1963, p. 19).

Sellars asserted that (although the objects of the manifest image include animals and things) «there is an important sense in which the primary objects of the manifest image are persons» – that is, objects «capable of the full range of personal activity» (1963, p. 12); and that although the scientific image is only in the process of coming into being, it is potentially one which will purport to give a complete description of the world and its processes, in which «the scientific image of man turns out to be that of a complex physical system» (Sellars, 1963, p. 25).

He then identified three “lines of thought” as to ways in which the two images might be reconciled:

1. Manifest objects are identical with systems of imperceptible particles in the sense in which a forest is identical with a number of trees.
2. Manifest objects are what really exist; systems of imperceptible particles being “abstract” or “symbolic” ways of representing them.
3. Manifest objects are “appearances” to human minds of a reality which is constituted by systems of imperceptible particles.

Sellars favoured the third alternative, but identified two particular difficulties with it.

One was that, according to this alternative, there must be minds to which the systems of particles have an appearance; and it may be questioned whether these minds can themselves be just systems of particles (presumably, appearing to themselves in self-awareness as minds). Sellars contended that, in relation to thinking, this objection can be met by the point that «the concept of a thought is the concept of an inner state analogous to speech» (1963, p. 33), that is, to overt conduct that could plausibly be constituted by the processes of systems of particles. However, he accepted that this answer would not suffice in relation to sensations; and proposed that the scientific image would need to be extended to embrace some «non-particulate foundation of the particulate image» (Sellars, 1963, p. 37), in which, presumably, sensory consciousness could have a place.

The other difficulty was

[T]here would remain the task of showing that categories pertaining to man as a person who finds himself confronted by standards (ethical, logical, etc.) which often conflict with his desires and impulses, and to which he may not conform, can be reconciled with the idea that man is what science says he is. (Sellars, 1963, p. 38)

Sellars proposed that the conceptual framework of persons was «a framework in which we think of one another as sharing the community intentions which provide the ambience of principles and standards [...] within which we live our own individual lives» (1963, p. 40) and that this framework did not need to be reconciled with the scientific image, but rather could be joined to it.

Reformulating the Distinction

Like Sellars, I think it is useful to draw a distinction between two conceptual frameworks in terms of which human beings conceive of themselves and their place in the world, one being that in which the concept of persons plays a central role, and the other being that constructed in accordance with the methods of the objective sciences; and to consider how these two “images” can be reconciled.

However, unlike Sellars I think the distinction is most helpfully drawn by reference to (1) the centrality in the first framework of persons considered as subjects who have conscious experiences (including visual and auditory experiences, thoughts and feelings), beliefs, desires and intentions, and who do things for reasons; and to (2) the total exclusion from the second framework of explanations in terms of purely subjective factors (as distinct from objectively verifiable reports, or other objectively verifiable indications, of subjective factors).

That is, I think the essence of the most helpful distinction lies in the importance of subjective factors in the first framework, and their exclusion from the second. This aspect of the distinction is masked in Sellars’ discussion, because of his inclusion of things (as well as subjects) in the first framework, and his exclusion from the second framework of those aspects of the objective sciences that do not postulate imperceptible objects and events for their explanations. I think these matters give rise to two disadvantages to the distinction drawn by Sellars:

- (1) It suggests a sharp distinction between those aspects of the objective sciences which postulate imperceptible objects and events, and those

aspects which do not do so, whereas in fact there is no such sharp distinction.

- (2) It assimilates the reconciliation of the manifest and scientific images of things to the reconciliation of the manifest and scientific images of persons, whereas in fact the two raise quite different problems.

The reconciliation of the manifest and scientific images of things is not without difficulty; but the main difficulty arises largely from the measurement problem of quantum mechanics, a matter that Sellars does not address and that I will not consider here. The reconciliation of the manifest and scientific images of persons involves the problem of accommodating subjectivity along with the objective processes that are the concern of the objective sciences. Today this is sometimes expressed in terms of the problem of reconciling folk psychology and neuroscience.

Approaches to Reconciliation

As noted above, Sellars proposed three alternative ways of reconciling the images of “man” that he had identified. What I have written so far suggests one problem with his proposal, namely that it assumes that the scientific image must be of things and persons as systems of imperceptible particles. This is an assumption made highly dubious by quantum mechanics; and as I have noted above, later in his article Sellars himself proposed that the scientific image would need to be extended to embrace some “non-particulate foundation of the particulate image”.

More importantly, Sellars entirely omits a fourth possible line of thought as to ways in which the two images might be reconciled, namely that both are partial images of reality, which is not fully represented by either image on its own.

This is a view that has not been given much consideration in contemporary discussion, because of assumptions that are generally made about the supervenience of the mental on the physical and the causal closure of the physical. It is widely assumed by philosophers that the mental supervenes on the physical, not only in the sense that there is no change in mental processes without a corresponding change in physical processes, but also in the sense that what mental processes occur depends entirely upon what physical processes occur; and accordingly that the physical is closed to affectation by any causal influence that is not physical. That is not to say that mental

processes cannot affect physical processes at all, but rather to say that they can do so only in virtue of their character as being themselves physical processes.

Now I do accept that there is no change in mental processes without a corresponding change in physical processes. But I say the view is open that it is no more correct to say that what mental processes occur depends upon what physical processes occur, than it is to say that what physical processes occur depends on what mental processes occur; and that it is better to assert correlation between the physical and the mental than to assert that one wholly depends on the other. One reason for this is that quantum mechanics strongly suggests that laws of nature do not uniquely determine how initial conditions change over time, but generally leave open spectra of possible outcomes. It thereby undermines an argument sometimes put that the physical world must be closed to non-physical affectation, because otherwise there would have to be some kind of mental force operating alongside the known physical forces. In fact the spectra of possibilities left open by quantum mechanics are all consistent with the operation of known physical forces, so that any selection between them would not require the application of any force.

All this means that the fourth possible line of reconciling two images of persons, which I would characterise as the subjective folk-psychological image and the objective scientific image, deserves close investigation. Indeed, my own view is that representation of reality requires the two images in combination, and cannot be achieved by one or other of the images on its own. Contrary to Sellars, I don't think the objective scientific image can provide a complete description of the world and its processes. I say there cannot be a complete description that excludes reference to subjective matters.

This last assertion could be taken in a weak sense or a strong sense.

It could be taken merely as asserting that reality does include subjective experiences, and as not asserting that those experiences make any difference to what happens that does not itself have a complete explanation in terms of objective physical processes. That approach would be consistent with the type of dualism associated with David Chalmers (1996).

Or it could be taken as asserting that subjective experiences do make a difference to what happens that does not itself have a complete explanation in terms of objective physical processes, so that the folk-psychological descriptions refer to a reality the functioning of which is not fully captured by the objective sciences. This is the view I support.

Supporting the Fourth Line of Reconciliation

I have recently published a book which sets out a systematic and cumulative argument for this general approach (Hodgson, 2012). I will here very briefly summarise some of the arguments.

Scientific explanations assume the exceptionless operation of laws of nature, in combination with circumstances on which the laws operate and with aspects of which the laws engage; so that whatever happens is determined by the engagement of laws of nature with circumstances, or else occurs randomly within probability parameters determined by engagement of laws of nature with circumstances. In neither case is there any room for an efficacious non-random input to what happens that is not itself determined by the engagement of laws of nature with circumstances. My contention is that there are powerful considerations in favour of the propositions (1) that conscious experiences do make an efficacious non-random input into what happens that is not determined by the engagement of laws of nature with circumstances, and (2) that to understand this input there needs to be reference to the subjective folk-psychological image of persons.

All intellectual endeavours presuppose that the persons engaged in them have the capacity to make reasonable decisions about what to believe and what to do. Of course everyone's thinking is fallible and subject to fallacies and biases; but unless we assume that we have the capacity to combat fallacies and biases, and to make reasonable albeit fallible decisions concerning whatever it is we are investigating, there would be no point in setting out on any investigation. This is true for scientific investigations as much as any other kind of investigation.

An important part of the capacity to make reasonable decisions consists in the capacity to engage in plausible reasoning, that is, reasoning in which the conclusions are not conclusively determined by overt application of rules for good reasoning (such as rules of logic or mathematics or probability, or any other kind of rule that could be incorporated into a computer program) to premises or data, but rather require the resolution of inconclusive reasons by exercise of reasonable albeit fallible judgment. The need for plausible reasoning is not avoided by resort to the scientific method, because plausible reasoning is needed for formulating hypotheses to be tested, for devising experiments to test them, and for determining which unrefuted hypotheses should be provisionally accepted. Arguments of Hilary Putnam (1981, pp.

174–200) and others have shown that plausible reasoning cannot be reduced to any kind of algorithmic process using discovered or invented rules for good reasoning.

However, it is of course possible that plausible reasoning might be achieved wholly by brain processes which unfold as determined by laws of nature and/or computational rules, and which produce reasonable decisions because the structures supporting these processes, and any computational system they instantiate, have been selected by millions of years of evolutionary trial and error. In terms of the three levels of cognitive processing originally identified by David Marr (1982), plausible reasoning at the top (overt) level could be supported by rule-determined computational processes at the middle (algorithmic) level and law-determined physical brain events at the bottom (implementational) level. On this approach, what appears to be plausible reasoning, resolving inconclusive reasons, is the overt expression of conclusive rule- and/or law-determined processes operating at lower levels, with no further efficacy in relation to the resolution of the inconclusive reasons being provided by the plausible reasoning at the top level.

Contrary to this approach, I say there are strong reasons to think that conscious experiences, operating at the top level of cognitive processing, have an input into decision-making that is neither random nor determined by rules of any kind.

Our brains do have a prodigious capacity for unconscious information-processing, but when we have an important decision to make, we generally cannot help addressing it consciously. Potential solutions to problems we address are thrown up by unconscious processes (for example, when we “sleep on” a problem), but we do not adopt those solutions without addressing them consciously. Our unconscious information-processing seems to be finely tuned to support conscious experiences, in which currently important information is presented simply and vividly, in the manner of an executive summary prepared for a decision-maker in business or government. Surely, evolution has selected the capacity to provide these executive summaries, just because they are useful in decision-making and contribute positively to it.

Another strong indication that conscious experiences contribute positively to decision-making is the fact that we have feelings like pain to motivate us. If there was no positive contribution to decision-making from conscious experiences, why would there be any more than unconscious computation and implementation of the course of action best suited to detecting and repairing

damage to ourselves and avoiding damage in the future? Pain would be a superfluity.

Then, if one accepts, as I think one should, that conscious experiences do make a positive contribution to decision-making, these questions arise: what is that contribution, and what is it about conscious experiences that enables them to make a contribution that is not made by unconscious processes. Any contribution that depends for its usefulness solely on the operation of evolution-selected computational rules, or on the law-governed operation of evolution-selected structures, would be a contribution that could be made automatically, without consciousness, at least unless consciousness somehow emerged as a by-product of such operation. No one has ever suggested any plausible explanation of how or why consciousness would be such a by-product.

I have a specific and straightforward suggestion as to what it is that consciousness can bring to decision-making that is not provided by unconscious information-processing: consciousness enables an organism to determine an apt response to circumstances facing it, which has regard, not only to features that can engage with laws of nature and/or computational rules, but also to whole combinations of features that are particular and perhaps unique to those circumstances and do not as wholes engage with any laws or rules.

The point here is that laws or rules engage with types or classes of things, or with variable quantities that can engage with mathematical rules. Generally, a conscious experience such as a visual experience comprehending many features of an observed scene, is not such as would, as a whole, engage with any law of nature or computational rule – although of course many of its constitutive features could do so. We do however grasp such experiences as gestalt wholes, and the question is whether this grasp of wholes, that we undoubtedly have, makes a contribution to decision-making.

That it does so appears most clearly, I think, in relation to aesthetic judgments, made by persons creating aesthetic works or by persons appraising them. Even a melody as simple as *The Man I Love* (and indeed each of many two or four bar chunks of that melody) is a unique whole that did not exist until the melody was created by George Gershwin. When Gershwin was composing it, no doubt possibilities for how it should proceed were thrown up by unconscious processes – but he must then have consciously appraised these possibilities in order to decide whether to adopt them or modify them or look

for other possibilities. In doing so, Gershwin must surely have responded to gestalts of the melody and/or chunks of it, which because they were unique and unprecedented could not have engaged with pre-existing rules of any kind; and his adoption of the melody in its final form could not have been wholly pre-determined by pre-existing circumstances and pre-existing laws or rules.

After this melody had been composed and heard by the composer or another, there could from this initial hearing be constituted, for the purpose of future cognitive processes of that person, computational rules capable of engaging with that melody as a type. But that could not be the case before the person’s first hearing of the melody; and I suggest that rules supporting apposite responses to such a gestalt would not be constituted unless the person had first consciously grasped and responded appositely to the gestalt.

Generally, I contend that if there is any merit or validity in aesthetic judgments, as I believe there is, there must be a contribution to those judgments from the appraiser’s grasp of unique wholes and their relationship to constituent features of the work in question; and what I say is that this contribution cannot be either merely random or precisely determined by laws of nature or computational rules. And I contend that what goes for aesthetic judgments also goes for plausible reasoning generally. In particular, I suggest that the grasp of gestalts of conscious experiences contributes to reasonable judgments as to:

- (1) what it is about what is experienced that is significant, thereby promoting reasonable generalisations, reasonable use of analogies, and reasonable inference generally;
- (2) whether information given by the senses is accurate information about something that is real;
- (3) whether something experienced relevantly or sufficiently approximates to an objective or ideal; and
- (4) generally, how inconclusive and incommensurable reasons are to be resolved.

This grasp of gestalts can thereby assist the understanding of areas of intellectual concern.

In my book I develop these arguments in some detail, and deal with objections to them; and I contend that they are consistent with and indeed cohere well with what science tells us about the world. So I say there is good

reason to think that subjective experiences do make a difference to what happens that does not itself have a complete explanation in terms of objective physical processes, so that the folk-psychological descriptions refer to a reality the functioning of which is not fully captured by the objective sciences.

Reconciling Science with Standards

It will be recalled that Sellars considered that, even if one can explain the existence of minds in terms of the scientific image,

[T]here would remain the task of showing that categories pertaining to man as a person who finds himself confronted by standards (ethical, logical, etc.) which often conflict with his desires and impulses, and to which he may not conform, can be reconciled with the idea that man is what science says he is. (Sellars, 1963, p. 38)

As mentioned earlier, Sellars proposed that the conceptual framework of persons was «a framework in which we think of one another as sharing the community intentions which provide the ambience of principles and standards [...] within which we live our own individual lives» (1963, p. 4); and that this framework did not need to be reconciled with the scientific image, but rather could be joined to it.

I suggest that Sellars' proposal does not do justice to the standards that he sees as confronting "man", and that standards are better accommodated by the approach I am advocating.

There is no doubt that the communities within which we live our lives are enormously important in shaping what we see as standards and principles confronting us and possibly conflicting with our desires and impulses. But to say, as Sellars does, that it is the community intentions that "provide the ambience of" these principles and standards, faces these difficulties:

- (1) It precludes the possibility of criticising and developing the principles and standards adopted by our communities, by reference to reasons that go beyond what is presently accepted by those communities.
- (2) It means that principles and standards have no more weight or bindingness on any person than is actually accorded to them by that person and/or is actually imposed on that person by the community.

These are difficulties that to some extent face any attempt to explain moral (or aesthetic) values as being no more than artefacts of human evolution and/or

culture; and of course the question whether or not values are no more than artefacts of human evolution and culture is an enormous and controversial question. However, what I say about Sellars’ proposal is that it not only assumes without argument that values are no more than artefacts of human evolution and culture, but also further limits them to being just artefacts of “community intentions”.

My proposal, that representation of reality requires both the subjective folk psychological image and the objective scientific image, does not require any assumption either that values are no more than artefacts of human evolution and culture, or that they are more than such artefacts. It leaves open the possibility that moral standards are binding each person, whether or not the person actually accepts them or actually has them imposed on him or her by the community. It also leaves open the possibility that moral and other standards may be supported or challenged on the basis of reasons that are not confined to community intentions (a possibility that is supported by my views about plausible reasoning and the contribution of conscious experiences to that reasoning). This I suggest is a further advantage of my proposal.

REFERENCES

- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Hodgson, D. (2012). *Rationality + Consciousness = Free Will*. New York: Oxford University Press.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Sellars, W. (1963). Philosophy and the Scientific Image of Man. In W. Sellars (1963), *Empiricism and the Philosophy of Mind*. London: Routledge & Kegan Paul Ltd, 1–40.

Of Time and the Two Images

Steven F. Savitt *

savitt@mail.ubc.ca

ABSTRACT

In this paper I argue that the clash of the Sellars' two images is particularly acute in the case of time. In *Time and the World Order* Sellars seems embarked on a quest to locate manifest time in Minkowski spacetime. I suggest that he should have argued for the replacement of manifest time with the local, path-dependent time of the "scientific image", just as he suggests that manifest objects must be replaced by their scientific counterparts.

1. Two Tables

Sir Arthur Eddington began the Introduction to his 1927 Gifford Lectures, *The Nature of the Physical World*, with a classic, but now little read, presentation of a philosophical puzzle:

I have settled down to the task of writing these lectures and have drawn up my chairs to my two tables. Two tables! Yes; there are duplicates of every object about me – two tables, two chairs, two pens....

One of them has been familiar to me from earliest years. It is a commonplace object of that environment which I call the world. How shall I describe it? It has extension; it is comparatively permanent; it is coloured; above all it is *substantial*. By substantial I do not merely mean that it does not collapse when I lean upon it; I mean that it is constituted of "substance" and by that word I am trying to convey to you some conception of its intrinsic nature. It is a *thing*....

Table No. 2 is my scientific table. [...] My scientific table is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed; but their combined bulk amounts to less than a billionth of the bulk of the table itself. Notwithstanding its strange construction it turns out to be an entirely efficient table. It supports my writing paper as satisfactorily as table No. 1; for when I lay the paper on it the little electric

* Department of Philosophy – University of British Columbia.

particles with their headlong speed keep on hitting the underside, so that the paper is maintained in shuttlecock fashion at a nearly steady level....

[T]here is a vast difference between my scientific table with its substance (if any) thinly scattered in specks in a region mostly empty and the table of everyday conception which we regard as the type of solid reality... (Eddington, 1928, p. ix)

The philosophical puzzle can be stated in a simple, almost Seussical, way. Which table is real, Table 1 or Table 2?

In grappling with this question, Wilfrid Sellars constructed a grand and imaginative metaphilosophy, presented in *Philosophy and the Scientific Image of Man* (Sellars, 1963).¹ There are, according to Sellars, two global conceptual frameworks or “images”. One, the “manifest image”, is the venerable worldview of common sense, augmented by non-postulational scientific reflection. Table 1 finds its home in the manifest image.

Table 2 is a denizen of the “scientific image”, an emerging, evolving, unified framework based on (but of course not restricted to) the postulational science that arose in the 19th century. This image differs radically from the manifest image. It paints a different picture of things (broadly construed), as Eddington so colorfully indicated above, and it tells a different story about time, as I will suggest below. But before getting to that topic, we should not neglect the question that is on the floor – which table? That is, which image?

When this question arose, Sellars noted that:

Three lines of thought seemed to be open: (1) Manifest objects are identical with systems of imperceptible particles in that simple sense in which a forest is identical with a number of trees. (2) Manifest objects are what really exist; systems of imperceptible particles being “abstract” or “symbolic” ways of representing them. (3) Manifest objects are “appearances” to human minds of a reality which is constituted by systems of imperceptible particles. Although (2) merits serious consideration, and has been defended by able philosophers, it is (1) and (3), particularly the latter, which I shall be primarily concerned to explore. (PSIM, p. 26)

Eddington seemed to accept *both* (3) and (1). «I need not tell you», he wrote, «that modern physics has by delicate test and remorseless logic assured me that

¹ I will refer to this essay as PSIM. PSIM was originally given as two lectures at the University of Pittsburgh in December, 1960. My page references will be to the reprinted version in *Science, Perception, and Reality* (1963).

my second scientific table is the only one which really is there...» (1928, p. xiv). Yet later on that page he imagines someone asking him:

You speak paradoxically of two worlds. Are they not really two aspects or two interpretations of one and the same world?

And he replies:

Yes, no doubt they are ultimately to be identified after some fashion.

Sellars rejects this second line of thought, his (1), which I heard him once characterize as “the kid-in-the candy-store approach”. At first sight, this rejection is mildly puzzling, for after a brief initial description of the two images Sellars says:

[T]he philosopher is confronted not by one complex many-dimensional picture, the unity of which, such as it is, he must come to appreciate; but by *two* pictures of essentially the same order of complexity, each of which purports to be a complete picture of man-in-the-world, and which, after separate scrutiny, he must fuse into one vision....

The Philosopher, then, is confronted by two conceptions, equally public, equally non-arbitrary, of man-in-the-world and he cannot shirk the attempt to see how they fall together in one stereoscopic view. (PSIM, pp. 4–5)

Would not a genuine *fusion* of the two images retain the objects of the manifest image as systems of “particles” of the scientific image? Tempting as this idea may be, Sellars thought there was a decisive objection to it. First, systems of objects had to conform to a certain principle, which I will call (R):

If an object is in a strict sense a system of objects, then every property of the object must consist in the fact that its constituents have such and such qualities and must stand in such and such relations or, roughly, every property of a system of objects consists of properties of, and relations between, its constituents. (PSIM, p. 27)

Sellars famously thought that a (manifest) pink ice cube, an object that is coloured pink through and through, a pink continuum, could not be a system of imperceptible “particles” of the scientific image in a way that satisfied (R).² (PSIM, p. 27) Fusion or no, when it comes to ontology, one of the two images must predominate, and it is clear to Sellars which one it has to be.

[S]peaking as a *philosopher*, I am quite prepared to say that the common sense

² To appreciate the complexity of this argument, the reader might well consult Hooker (1977).

world of physical objects in Space and Time is unreal – that is, that in the dimension of describing and explaining the world, science is the measure of all things, of what it is that it is, and of what is not that it is not. (PSIM, p. 173)

What I will suggest in the rest of my essay is that Sellars failed to carry his metaphilosophical ideal through for a central philosophical topic, time. From here on out my discussion will be more speculative and venture into areas that are less well-known, but are no less fundamental, than the familiar territory sketched above.

2. Two Times

The clash between the manifest view of time and the emerging scientific picture of time is, as I see it, stark. In the manifest image the present is a distinguished global hyperplane of simultaneous events. The passage of time is the successive occurrence of such presents.³ Contrast this with the following remark in Einstein (1949):

We now shall inquire into the insights of definite nature which physics owes to the special theory of relativity.

There is no such thing as simultaneity of distant events... (Einstein, 1949, p. 61)

If Einstein is correct, then the legs are simply cut out from under the pre-relativistic concept of time. If there is no distant simultaneity, there are no distinguished 3-dimensional, global sets of simultaneous events. If the passing of time is the successive occurrence of such sets of events, then there is no temporal passage as well.

Here, if anywhere, the two images clash. Here, if anywhere, there is need for philosophical attention. Can the two disparate images of time be fused? It is by no means easy to see how. If «science is the measure of all things, of what it is that it is, and of what is not that it is not» (Sellars, 1956, p. 173), should not commonsense time be consigned to the “what is not” bin and replaced by its emerging, austere scientific successor concept?⁴ This is the question I believe

³ I argue that this is indeed the folk or commonsense picture of time and passage and that it evades the usual *metaphysical* objections in Savitt (2002).

⁴ One might also wonder how this replacement could possibly be reconciled with our ostensible experience of time passing. I explored this question in Savitt (1994).

we should have in mind as we turn to Sellars' major treatment of time, *Time and the World Order* (1962, henceforth referred to as TWO).

Sellars acknowledges the centrality of time at the beginning of his essay.

I have "taken time seriously" since I cut my philosophical teeth on McTaggart's well-known paper on the unreality of time and the attempts of Broad and others to refute him. I soon discovered that the "problem of time" is rivaled only by the "mind-body problem" in the extent to which it inexorably brings into play all the major concerns of philosophy. Here, if anywhere, analysis without synopsis must be blind. (TWO, p. 527)

After indicating the range of topics Sellars thinks is involved in dealing with the "problem of time", he adds a warning:

As is implied by the dialectical character of the treatment, these topics make multiple appearances, and the "conclusions" of one section are often radically recast in another.

It will be all-too-easy, then, to mistake a provisional dialectical ploy for a final conclusion, so I offer my thoughts on TWO tentatively, even if I express them bluntly, hoping that those with a deeper grasp of Sellars' views will find it worthwhile to explore more deeply, and perhaps more accurately, the topics to follow.

3. Time 1

The first curious things to note is that, despite the fact that Sellars "takes time seriously", the problem of time make no appearance in PSIM. Conversely, the terms "manifest image" or "scientific image" do not appear at all in TWO, even though internal evidence indicates that TWO was written in 1958,⁵ just two years before the lectures of PSIM were given in 1960.

Although the terminology of PSIM does not appear in TWO, the paper can be read as structured in terms of its leading ideas. Sections 1–4 of TWO deal with what Sellars calls *the thing framework*, whereas sections 4–8 deal with *the 'event' framework*. Sellars consistently used single quotes to distinguish the 'events' of the second framework from the events in the first.

The basic entities of the thing framework are, naturally enough, things (or substances). The basic way we speak of these things is tensed. We speak about

⁵ See the first three numbered sentences in TWO, for instance.

what things are doing, have done, or will do. Given this bedrock tensed discourse about things, we can introduce “episodes”. For instance, following a schema developed in Reichenbach (1947, §48), one can introduce events (one kind of episode) by noting the equivalence of, for example,

(1) George VI was {is, will be} crowned

with

(1') The coronation of George VI took {is taking, will take} place.

Sellars insists that the thing language is basic, the event language derived, though Reichenbach, noting that there is an equivalence between the two sentences, seems willing to admit either form of discourse as equally basic (in the last paragraph of his 1947, §48). To assert that there are past or future episodes is merely to assert sentences like (1), which are basic to the tensed discourse of the thing framework. But to move from this simple, non-relational tensed discourse to sophisticated, metrical time, we need, according to Sellars, “such locutions as”

(2) Nero fiddled *while* Rome burned.⁶

Unfortunately, Sellars does not say directly what other locutions are like (2) in the relevant way. Since (2) seems to assert that two episodes are simultaneous, one might suppose (following a hint on p. 573 of TWO that will be quoted below) that the other locutions might look like

(2') Nero fiddled *before* Rome burned

and

(2'') Nero fiddled *after* Rome burned

yielding the traditional crop of statements expressing McTaggart’s B-relations. But Sellars may be *blocking* precisely this way of looking at (2) et al. when he says that «we must not equate statements involving temporal connectives such as “while” with statements formulating temporal relations between episodes» (TWO, p. 522).

Time, and times, are not built up from events or episodes, but rather they are «introduced as a metrical framework» (TWO, p. 552) in which events are then ordered. The framework is, perhaps, overlaid on events, though there is no discussion of sentences like

⁶ This sentence appears as numbered sentence (92) on p. 552 of TWO. Unfortunately, there is another sentence with the same number on p. 549.

(3) The sun is overhead here today 24 hours later than it was overhead here yesterday,

which might have clarified Sellars' intent.⁷ What is clear, though, is at the final step things and their activities find their places in the framework of time, so that we can move from statements like

(4) S was Φ_1

to

(4') S was Φ_1 at t

or

(4'') S was Φ_1 in 1957.

The picture that Sellars is working with in the framework of events, it would seem, looks something like this.

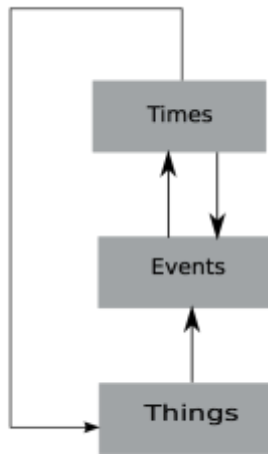


Diagram 1.

⁷ But he does later point out that «all metricizing in the framework of things is a matter of the location of *events*...» (TWO, p. 572). As I understand TWO, Sellars maintains that the distinctive contribution of the special theory of relativity is to expand the space of possible metricizings. When this is properly understood, he thinks that the apparent gulf between the two frameworks is bridged.

4. Time 2

In sections 5 through 8 of TWO Sellars discusses the framework of ‘events’, which is in essence the framework of the special theory of relativity (STR). In thinking of the philosophical aspects of STR, philosophers (and physicists too) often suppose that it requires that ‘events’ are ontologically basic and that the “real” relations between these events are the relations that are invariant under the Lorentz transformations; but a philosopher who adopts these views is, according to Sellars, «seriously confused» (TWO, p. 567).

Just as in the thing framework, relations between ‘events’ depend ultimately on our activities and the activities of things about us. We are urged not to lose sight of the fact that

[T]opologically characterized events, instead of being the concrete reality of the world process, are simply abstract features common to all metrical pictures of the world. The temptation to think of the continuum of events topologically conceived apart from specific metrics as the basic reality which includes these metrics as specific patterns of topological relationship is a mislocation of the fact that the metrical discourse about events is rooted in premetrical tensed discourse in which we talk about doing this or that *while (before, after)* other things do this or that in our immediate practical environment. (TWO, p. 573)

I am not quite sure what Sellars means by “metric” in the passage above and by “metrecizing” in the passage to be quoted soon, but I would like to hazard a guess, based on his various uses of the term. Since he uses the plural “metrics”, I don’t think he is referring (as one would assume nowadays) to the standard Minkowski metric (say, $h_{ab} = (-1,1,1,1)$). I suspect he is simply thinking of the various ways of coordinatizing Minkowski spacetime, using standard Einstein synchrony, relative to different “inertial observers”.

Given the well-known basic results of STR like the relativity of simultaneity, length contraction and time dilation, it is often argued that the “real” quantities in Minkowski spacetime are (or are defined in terms of) the invariant quantities like the invariant spacetime interval between points or events in the spacetime. A particularly elegant and influential expression of this point of view was to come a few years after TWO in Stein (1968).⁸ Sellars will have none of it.

⁸ Though I do not wish to imply that Stein ever makes much ado about the term “real”.

In what sense are space-time intervals “more real” than lapses of time and spatial distances? Only in the sense that the space-time interval between two events is an invariant quantity with respect to the Lorentz transformations, that is, with respect to all the metrecizings into a temporal order of spatially related events. To suppose that it is in any other sense “more real” is, as we shall see, analogous to supposing that events as standing in the earlier-later relations (with respect to a given metrecization) are “more real” than events as past, present, or future (in a given metrecization) because earlier than is invariant with respect to “the changing location of the ‘now’”.

On this issue I will not take sides. My own *laissez-faire* view of “is real”, expressed at some length in Savitt (2006), is that as long as one is clear as to what the term is meant to include and – most importantly, following Austin (1962, ch. 7) – to exclude, and as long as one uses the term consistently, a given item might wind up in the *real* box given one way of understanding “is real” or in the *unreal* box on another. What is far more important is the fact that Sellars wishes to minimize the difference between the thing and the event framework. Consider the following remark a bit later on in TWO:

If these considerations are sound, then the idea that, in an ‘event’ framework, events have a *timeless* existence in which they stand in objective *temporal* relations and constitute a system which includes the perspectival distinctions of pastness, presentness, and futurity as properties relative to points of view located within the system is a mistake. It is the mistake of assuming that a primary temporal picture of the world can be one which does not *use* but only *mentions* the term “now”. (TWO, p. 590)

The fact that there are many “nows” in Minkowski spacetime is less a problem for Sellars than the fact that there seem to be no “nows” therein when only the light-cone structure is considered. Put in terms of the diagram we offered above, the ‘event’ framework in Sellars’ view is just like the thing framework except that there will be more reciprocal pairs of arrows connecting the level of events with the level of times, the pictorial manifestation of the multiplicity of new “metrecizings” made available by STR.

This line of thought is reinforced, I believe, by a surprising suggestion that Sellars makes at the end of Part I of TWO⁹, when he writes:

The more one appreciates the systematic character of the difference between

⁹ Part II, *Determinism and Truth*, examines fatalism and three-valued logics, a topic I will neglect here *in toto*.

the framework of things and the framework of ‘events’, the more one comes to realize that the latter framework is *in the first instance* simply a reaxiomatization of the former, and differs from it only as a Euclidean geometry axiomatized with one set of primitives differs from one which has been axiomatized with another set of primitives. (TWO, p. 594)

While the hedge “in the first instance” is cashed out later in the paragraph by some speculation that matters might be different at «deeper levels of physical theory», the view is misguided even “in the first instance”. If theory T' is a reaxiomatization of theory T , then the two must have the same set of theorems or truths, aside from the fact that a theorem in one theory might be called axiom in the other, and *vice versa*. *This is simply not true for “folk” time or Newtonian time as opposed to special relativistic time*. In particular, time is path-dependent in STR whereas it is not pre-relativistically. In STR the time difference between two timelike separated points as measured by an ideal clock whose worldline contains the two points depends upon the path taken between them, the time difference being longest for a path that represents the trajectory of an unaccelerated object. The path dependence of time is the conceptual basis for the so-called “twin paradox”¹⁰, and it has been measured experimentally.¹¹ It is a perennially startling feature of STR.

This difference between Time_2 and Time_1 in turn suggests that the guiding idea of Sellars, that primary temporal pictures must contain a “now”, a global hyperplane of simultaneous events (or, perhaps, episodes) is not correct. It may indeed be at the root of temporality in the framework of things, but time in STR is a local, path-dependent notion, not a global one. The ideal of fusing these two images seems to underestimate the gap between them. If one has to choose between them, then the relativistic notion must prevail, and jettisoning globality seems to be the right way to approach philosophy of time in a Sellarsian fashion, even though it runs counter to Sellars own approach. Time_1 must go the way of table_1 .

There is much else in TWO – many subtle reflections on the existence of the past and future and on the intricacy of tensed language. I have focussed here only on the themes in it most relevant to the metaphilosophy of PSIM, hoping that it may stimulate others to think more either about TWO or about philosophy of time in Sellarsian manner.

¹⁰ See Arthur (2010) for a thorough explanation.

¹¹ See, for example, Hafele and Keating (1972a, 1972b).

But I will add one final reflection. During the time that Sellars was writing TWO, Adolf Grünbaum was composing his long essay, *Geometry, Chronometry, and Empiricism* which appears immediately before it in Volume III of the Minnesota Studies series. Sellars adds in footnote 26 that Grünbaum provided «many helpful comments and criticisms» during the writing of TWO. Since Grünbaum would doubtless differ with Sellars about STR and relativity in general, any notes or records of their discussions could prove illuminating.

REFERENCES

- Arthur, R.T.W. (2010). Minkowski's proper time and the status of the clock hypothesis. In V. Petkov (Ed.), *Space, Time, and Spacetime. Physical and Philosophical Implications of Minkowski's Unification of Space and Time*. Berlin, Heidelberg, New York: Springer-Verlag, 159–179.
- Austin, J. (1962). *Sense and Sensibilia*. Oxford: Oxford University Press.
- Eddington, A.S. (1928). *The Nature of the Physical World*. Cambridge: Cambridge University Press.
- Einstein, A. (1949). Autobiographical Notes. In P. Schilpp (Ed.), *Albert Einstein: Philosopher-Scientist*. La Salle, Illinois: Open Court.
- Hafele, J., & Keating, R. (1972a). Around-the-World Atomic Clocks: Predicted Relativistic Time Gains. *Science (new series)*, 177(4044), 166–168.
- Hafele, J., & Keating, R. (1972a). Around-the-World Atomic Clocks: Observed Relativistic Time Gains. *Science (new series)*, 177(4044), 168–170.
- Hooker, C. (1977). Sellars' Argument for the Inevitability of the Secondary Qualities. *Philosophical Studies*, 32(4), 335–348.
- Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York: The Free Press.
- Savitt, S. (1994). The Replacement of Time. *Australasian Journal of Philosophy*, 72(4), 463–474.

- Savitt, S. (2002). On Absolute Becoming and the Myth of Passage. In C. Callender (Ed.), *Time, Reality and Experience*. Cambridge: Cambridge University Press, 153–167.
- Savitt, S. (2006). Presentism and Eternalism in Perspective. In D. Dieks (Ed.), *The Ontology of Spacetime*. Amsterdam: Elsevier, 111–127.
- Sellars, W. (1956). Empiricism and the Philosophy of Mind. In H. Feigl, & M. Scriven (Eds.), *Minnesota Studies in the Philosophy of Science*, vol. I. Minneapolis: University of Minnesota Press, 253–329. Reprinted with additional footnotes in W. Sellars (1963), *Science, Perception and Reality*. New York: The Humanities Press, 127–196.
- Sellars, W. (1962). Time and the World Order [TWO]. In H. Feigl, & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science*, vol. III. Minneapolis: University of Minnesota Press, 527–618.
- Sellars, W. (1963). Philosophy and the Scientific Image of Man [PSIM]. In W. Sellars (1963), *Science, Perception and Reality*. New York: The Humanities Press, 1–40. First pub. in R.G. Colodny (Ed.) (1962), *Frontiers of Science and Philosophy*. Pittsburgh, PA: University of Pittsburgh Press, 35–78.
- Stein, H. (1968). On Einstein-Minkowski Space-Time. *The Journal of Philosophy*, 65, 5–23.

The Unity of the Manifest and Scientific Image by Self-Representation *

Keith Lehrer[†]

lehrer@email.arizona.edu

ABSTRACT

Sellars (1963) distinguished in *Empiricism and Philosophy of Mind* between ordinary discourse, which expressed his “manifest image”, and scientific discourse, which articulated his “scientific image” of man-in-the-world in a way that is both central and problematic to the rest of his philosophy. Our contention is that the problematic feature of the distinction results from Sellars theory of inner episodes as theoretical entities. On the other hand, as Sellars attempted to account for our noninferential knowledge of such states, particularly in correspondence with Castañeda, discussed by Lehrer and Stern (2000), he is lead to account of representation of such states that incorporates the states into what Lehrer has called *exemplar representation* (2004, 2011a) and Ismael *reflexive self-description* (2007). What is common to the three accounts, with some differences, is that such states may be function reflexively in selfrepresentation. Our argument is that the elaboration of this account, suggested in Sellars, shows how the discourse of the manifest image can be transformed into the discourse of the scientific image as self-representations of scientific entities.

Sellars (1963) distinguished in *Empiricism and Philosophy of Mind* (EPM hereafter) between ordinary discourse, which expressed his “manifest image”, and scientific discourse, which articulated his “scientific image” of man-in-the-world in a way that is both central and problematic to the rest of his philosophy. Our contention is that the problematic feature of the distinction results from

* I have used the first plural form in this paper to express my indebtedness to David Stern for earlier discussion leading to our earlier joint paper and to Jenann Ismael for more recent discussion leading to this paper, though I alone remain responsible for the final content of it.

[†] University of Arizona.

Sellars theory of inner episodes as theoretical entities. On the other hand, as Sellars attempted to account for our non-inferential knowledge of such states, particularly in correspondence with Castañeda, discussed by Lehrer and Stern (2000), he is led to an account of representation of such states that incorporates the states into what Lehrer has called *exemplar representation* (2004, 2011a) and Ismael *reflexive self-description* (2007). What is common to the three accounts, with some differences, is that such states may function reflexively in self-representation. Our argument is that the elaboration of this account, suggested in Sellars, leads to an explanation, which is central to Sellars, of how this special form of self-representation can explain how the self-represented states can provide an anchor for theory in experience, in the manifest image and the scientific image as well, and an explanation of how the discourse of the manifest image can be transformed into the discourse of the scientific image as self-representations, which give us what Ismael (2007) has called a phenomenal profile, can be decoupled from the discourse of the manifest image and added the discourse of the scientific image. We shall focus on the discourse of color, including appearances and sensations of color, as the test case for our argument that self-representation effects the transformation of the discourse of the manifest image to that of the scientific image into a coherent, though dynamically changing, image of man-in-the-world.

Our argument may strike some as close to a position that Sellars (1963) attacked as the “Myth of the Given”. Self-represented states may be described, and are by Ismael, as *self-presenting*, which is a vocabulary Sellars would have eschewed. However, our notion of self-representation is contained in Sellars in the Castañeda correspondence, cited in Lehrer and Stern (2000) and suggested already in EPM in his discussion of the reporting role of discourse of thoughts and feelings used in first-person reports. So, how can Sellars consistently combine self-representation of inner episodes with his rejection of the Myth of the Given and account of inner episodes originating as theoretical entities? It is important, and useful for our exposition, to clarify the matter at the outset. First of all, the target of the Myth of the Given is a specific claim that the existence of inner episodes *entails* knowledge and conception of them. His argument against the view is that conception and knowledge require the learning of language supplying a conceptual framework of reasons and justification. Since it is possible, and he holds, in fact the case, that inner episodes, sensations, for example, may occur prior to the acquisition of language, the occurrence of such episodes does not logically entail conception

or knowledge that they occur. Our account of self-representation does not conflict with Sellars rejection of the Myth of the Given.

The account of self-representation we advocate and believe to be central to explaining how the manifest image can be transformed into the scientific image, does not have the consequence that the occurrence of the episodes that are self-represented entail conception and knowledge of them. The occurrence of the episodes, sensations of color, for example, does not entail that they are self-represented or that we have knowledge of such states. The self-representation of inner episodes by exemplarizing them or profiling them is a process that uses the states as vehicles of representation reflexively representing themselves. It is possible that a human being could have sensations but lack the capacity to use those sensations as vehicles of representation. Sellars consistently held that such episodes, even if originally introduced in the manifest image to explain behavior within a theory postulating them as theoretical entities, do not by their simple occurrence logically entail representation of them, not even self-representation.

In terms of what Sellars was denying in the “Myth of Agreement”, he is consistent in suggesting a form of self-representation as a method of using inner episodes as reflexive vehicles to represent themselves. We concede, however, a possible area of disagreement. Sellars was convinced that the capacity to represent the world, our own states included, was acquired by learning a conventional language. Since his time, the arguments in favor of innate representational systems, especially by Fodor (1983), makes Sellars insistence on the actual role of language acquisition on the capacity to represent the world highly controversial. However, in fairness to Sellars, the argument against the Myth of Given is an argument against the view that the occurrence of inner episodes logically entails the representation of them. For that argument to succeed all that is required is the logical possibility that language learning should be necessary for acquiring the ability to represent or self-represent inner episodes. We acknowledge that logical possibility, which suffices for us to agree with Sellars. The conditions necessary for realizing the capacity for self-representation may be left open for the purposes of the present discussion, though the insistence of Sellars on the necessity for meta-representation for genuine descriptive use of language may return us to the issue.

We shall need to present that argument again published in Lehrer and Stern (2000) for the thesis that Sellars held the view that, although the

description of inner episodes was originally introduced in the manifest image as inferred theoretical entities, such episodes become self-representational as description of inner episodes takes on a reporting role in which first person reports are reliable and knowledge of the inner episodes becomes non-inferential. Castañeda asked Sellars for an account of how this transition occurs. Lehrer and Stern (2000) argue that Sellars answer is that the inner episodes act like exemplars used to represent themselves, as well as other phenomenal states, in the process of exemplarization. Lehrer (2007) suggested that self-representation, exemplarization, served the purpose of anchoring discourse to experience in response to the Quine-Neurath claim that we must rebuild the discourse of science floating at sea. Ismael (2007), in her writing directly on Sellars argues that once phenomenal states take on this role of self-representation, showing us what properties the episode exemplify by a kind of indexical ostension, the phenomenal profiles of self-representation can be decoupled from one form of descriptive discourse in which it anchors a descriptive map in experience to attach it another descriptive map again securing the anchor.

So both Lehrer and Ismael are implying, Ismael in more direct connection to Sellars whom she discusses, that the transition from description in the manifest image to description in the scientific image connects with the world and our place in the world by self-representation. As we move from the discussion of Eddington's two tables to the discussion of the relationship between inner episodes such as thoughts and sensations to states of the brain as neurological states and their subatomic constituents, we can decouple, in Ismael's terminology, the self-representations of the former and transfer them to the latter. So the self-representation of the same state can be used to show us what a reddish appearance is like in the descriptive discourse of the manifest image as well as being used to show us what red things look like under standard conditions, in that use of discourse, while showing us at later time in scientific understanding what a specific kind of r-brain state is like that occurs. We shall depend, at a certain stage at correlating the reddish appearance with the r-brain state to effect the transition, but once effected, we shall be able to report when we are in the r-brain state from the self-representation of the exemplarized state in our phenomenal profile. It is as though, given the profile, we shall be in position to report, "There it is, the r-brain state", because we have decoupled the self-representation from one form of report to use it for another. Moreover, though the transition may at first be inferential, based on the observed

correlation, it may become noninferential much in the way that we must at first infer where we have to strike a key to type a letter and subsequently perform the action noninferentially. Sellars suggested that we could recondition ourselves to respond to experience with the description in the scientific image instead of the one in the manifest image. That would leave us with a problem, as he would concede, of how we know noninferentially what the r-brain state described in scientific discourse is like when we report the occurrence of it from our experience. The answer, explained by Sellars in the Castañeda correspondence, is that the self-representation of the state anchors the description in our scientific image of the world and ourselves in the world.

Let us now turn to a more detailed argument that Sellars appealed to self-representation to give an account of how reports of inner episodes can constitute noninferential knowledge. Sellars is famous for his thesis in EPM that the language of inner episodes is introduced as a theory to explain behavior. This account has the advantage that the justification of claims about inner episodes, thoughts, for example, have the same structure as the justification of claims about theoretical entities, electrons, for example. This initial idea contains the suggestion that the justification of claims in the manifest image and the conceptual framework of it have the same structure in principle as claims in the scientific image and the conceptual framework of it. Moreover, his appeal to a behavioral theory of conditioned responses giving rise to the functional role of discourse about thoughts and sensations ties the meaning of discourse of inner episodes to behavior in a way that blocks the problem of the justification of claims about the thoughts and sensations of others. Such claims may be defeated, of course, but the meaning of claims about thoughts and sensations of others, inner episodes of others, commits one to considering behavior as evidence for claims of such episodes because such claims obtain their meaning from their role in explaining behavior. Behavior is evidence for claims about inner episodes by virtue of the meaning of the words used to describe them. The problem of our knowledge of other minds becomes the problem of our knowledge of theoretical entities. The unity of our descriptions of inner episodes in the manifest image with our description of theoretical entities in the scientific image seems embedded in the way such discourse is introduced to achieve explanation.

Though this has some plausibility, Sellars confronted an obvious objection, provided an answer, and was pressed to elaborate it by Castañeda, as presented in Lehrer and Stern (2000). Sellars acknowledged the problem. The problem

simply put is that first person reports are reports of noninferential knowledge of inner episodes, thoughts and sensations, for example, that the person does not infer from his behavior. So, the theory that introduces talk about inner episodes, in what Sellars formulates as his myth of Jones, who introduces the discourse of inner episodes to explain behavior, must add something to the role of discourse to explain how someone can know noninferentially about the existence of his own inner episodes. Sellars solution, which he calls the denouement in the saga of Jones, introduces a contrast between theoretical and nontheoretical in the language of thought as follows:

For once our fictitious ancestor, Jones, has developed the theory that overt verbal behaviour is the expression of thoughts, and taught his compatriots to make use of the theory in interpreting each other's behavior, it is but a short step to the use of this language in self-description. Thus, when Tom, watching Dick, has behavioural evidence, which warrants the use of the sentence (in the language of the theory) "Dick is thinking 'p'" (or "Dick is thinking that p"), Dick, using the same behavioral evidence, can say, in the language of the theory, "I am thinking 'p'" (or "I am thinking that p"). And now it turns out – need it have? – that Dick can be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe his own behavior. [...] Our ancestors begin to speak of the privileged access each of us has to his own thoughts. *What began as a language with a purely theoretical use has gained a reporting role.* (EPM, 1963, p. 189)

Two observations are in order. The first is that the "need it have?" expression is intended to block the objection that the existence of the inner episode *entails* that Dick or the rest of us have the ability to give self-descriptions of inner states. The obvious suggestion, elaborated in terms of conditioning, which Tom effects, is that it is a contingent matter whether Dick can master the use of the reporting role and has the cognitive ability to obtain noninferential knowledge of the inner episode, of his thought in this case.

The second observation leads us to the objection Castañeda raised in the correspondence (Castañeda and Sellars, 1961–1962), namely, that Sellars has not given us any explanation of how Dick can master the reporting role. Castañeda asks,

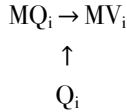
What exactly is what Jones reports in the new use of "I am thinking that p"? How is it that he can make correctly such a statement without observing his behavior? It is not easy to see how on your view these questions can be answered, if they are not answered, it is difficult to see exactly what your view is

accomplishing. (Castañeda and Sellars, 1961–1962, letter of March 6, 1961; cf. Lehrer and Stern, 2000, pp. 205–206)

He concludes by suggesting that to complete his account Sellars must fall back into the Myth of the Given and to admitting mental entities whose «*occurrence entails my knowledge of them*» (2000, p. 206). That is the challenge to Sellars. Either explain how the discourse of inner episodes, thoughts, acquires a reporting role as the result of conditioning resulting in noninferential knowledge or acknowledge that it is logically impossible to have the inner episodes, thoughts, without knowing that you have them.

Sellars reply will take us to the theory of self-representation, exemplarization and phenomenal profiles. Here is the core of Sellars reply.

The important thing to note is that the *core* of Dick’s learning to report what he is thinking is a matter of his acquiring a tendency (*ceteris paribus*) to respond to his thought that-p by saying “I am thinking that-p”. Everything hinges on the force of word “respond” in this connection. *It is being used as a technical term borrowed from learning theory.* The following diagram will help clarify matters:



where Q_i is a thought that-p, MQ_i is a meta-thought “I am thinking that-p” and MV_i is a meta-statement “I am thinking that-p”. (Castañeda and Sellars, 1961–1962, letter of April 3, 1961; cf. Lehrer and Stern, 2000, p. 207)

Sellars insists that the arrows are in the first instance the result of conditioning, and he concludes,

The above type of account explains the “privileged access” a person has to his own inner episodes. For (although worlds are possible in which this is not the case) only the person who has a thought that-p can respond to it [...] with the thought that he has the thought that-p. (Lehrer and Stern, 2000, pp. 207–208).

That is the account the Sellars offers in his reply to Castañeda’s objection that Sellars must offer an account of our self-description that avoids the Myth of the Given.

Brief reflection on the diagram above will reveal that Sellars answer rests on an assumption of the self-representation of inner episodes. He thereby avoids the Myth of the Given and sustains the role of self-representation in the

transition from the manifest to the scientific image. The short argument for self-representation, noted by Lehrer and Stern (2000), is that MQ_i contains Q_i in a generalization that with the use of dot quotes converts Q_i into a sortal, that is, a general representation that uses Q_i as an exhibit of the kind or class of objects it represents. We note the reflexive character of this form of representation, the use of Q_i as an exemplar to represent the class represented. Therefore, self-representation is involved in Sellars explanation of self-description. The vertical arrow in the first line of the diagram indicates the direction of the externalization of the internal representation, reflexive self-description or exemplarization, to the statement, which may be expressed as an utterance in the conventional language.

There are details in this account of self-representation, which are a modification of the original account offered by Lehrer (1996) and contained in Lehrer and Stern. The modifications were evoked by the work of Ismael (2007) and Fuerst (2010). Papineau (2002, 2007) and Lehrer (2000) had suggested that the representation of consciousness states using the conscious state to represent itself, as well as other states and objects, was best explained as something like quoting a word to obtain a representation of the word and then disquoting it to use it in self-representation. This is somewhat misleading in the case of reflexive representation as Ismael and Fuerst argued leading Lehrer to alter his view. The issue is that quotation and disquotation involve two tokens, the one quoted and the one disquoted as is apparent in a simple example from Sellars,

“red” means red

in which quotation and disquotation involve two different tokens, one quoted, the other not, in the sentence. There is no individual that refers to itself in the sentence. In exemplarization, by contrast, an exemplar is being used as an exhibit of a class of things represented and refers reflexively to itself. To take a different example, suggested by Goodman (1968), if I use a piece of cloth as a sample to represent a kind of cloth, Harris Tweed, for example, the sample refers to instances of cloth, and it is itself an instance of that cloth. The one sample is used as an exemplar to represent instances of cloth, and refers to the token of itself in the use of the token. The token loops back onto itself in exemplarization and reflexive self-description. Moreover, there is some security in the token loop of reference that is lost when one token refers to another. We are not arguing that it is logically impossible for the token to fail to represent itself, however. The token becomes representational and

represents what it does because of the way that it is used representationally. The token reflexive loop of self-representation depends on the exercise of a human capacity which, like all such capacities, lacks the logical infallibility of representation assumed by the Myth of the Given. The security of reference obtained from reflexive representation or exemplarization is that success depends only on the process or activity of reflexive representing and exemplarizing.

The correspondence with Castañeda we have discussed is focused on the question of our representation of thoughts and our noninferential knowledge of them. Sellars gave the same account of sense impressions or sense contents. Jones initially introduces them as explanatory entities, indeed, as inner states. In the case of the impression of a red triangle, it is, Sellars writes:

That state of the perceiver – over and above the idea that there is a red and triangular physical object over there – which is common to those situations in which

- (a) he sees that the object there is red and triangular
- (b) the object over there looks to him to be red and triangular
- (c) there looks to him to be a red triangular physical object over there. (EPM, 1963, p. 190)

He continues to say about the theoretical entities introduced by Jones,

[T]he hero of my myth postulates a class of inner – theoretical – episodes which he calls, say, *impressions*, and which are the end results of the impingement of physical objects and process on the body, [...] the eye. (EPM, 1963, p. 191)

Finally, Sellars adds the reporting role as in the case of thoughts,

Jones teaches his theory of perception to his followers. As before in the case of *thoughts*, they begin by using the language of impressions to draw theoretical conclusions from appropriate premises. [...] Finally, he succeeds in training them to make a *reporting* use of this language.... (EPM, 1963, p. 194)

Once again the myth helps us to understand that concepts pertaining certain inner episodes – in this case *impressions* – can be primarily and essentially *intersubjective*, without being resolvable into overt behavior symptoms, and that the reporting role of these concepts, their role in introspection, the fact that each of us has a privileged access to his impressions, constitutes a dimension of these concepts which is *built on* and *presupposes* their role in intersubjective discourse (EPM, 1963, p. 195).

Thoughts and sense impression are, according to Sellars, introduced in his myth of Jones as theoretical inner episodes. But these entities are real, not fictional, and the impressions like the thoughts, we propose from the above analysis of the correspondence concerning thoughts, become self-representational as they are exemplarized and become reflexive self-descriptions. Sellars insists that it takes a theory to produce, a conceptual framework, to effect the self-representation. He denies that we begin with self-representation of inner episodes as data in search of a conceptual framework. We locate the inner episodes within a conceptual framework. Jones may introduce a conceptual framework of thoughts and impressions, but the question that confronts Dick is how to find his thoughts and impressions in this framework? He needs to find fixed points of reference to locate his thoughts and impressions in the framework. Those fixed points enable Dick to report his location in the conceptual framework of his world, in himself in his world and in his world in himself.

How does self-representation solve the problem of how the framework represents the world? The exemplarized entities can exhibit at one and the same time what they are like, what a red impression is like, for example, and what a red object is like. The impression, Sellars insists, is not an object of perception. It is a state that can be used in representation to represent itself, giving us noninferential knowledge of what it is like when the representation finds a place in a conceptual framework. At the same time, it can be used to represent what a red object is like in the external world, referring at the same time to itself and to something beyond itself. So the inner episode, exemplarized in reflexive representation, shows us both what the external object is like, exhibiting what it is like for us, and how we represent the external object, exhibiting how we represent it. In that way it exhibits to us what we are like as we represent our world.

It is important for understanding Sellars to appreciate that fixed points of self-representation of thoughts and impressions are not the chronological starting points of representation and knowledge. They are found within discourse as we seek fixed points of reference and representation for ourselves. However, having found them when seeking fixed points for discourse and our conceptual framework, they may take us beyond the discourse and framework. For, we may use those self-representations as fixed points, as exemplarized reflexive representations, knowing what the states represented are like and how we use them to represent the world, in the conceptual framework of the

manifest image and beyond that in the scientific image. Self-representations, once found within a conceptual framework to show us what our experience of our world is like may be decoupled, as Ismael (2007) has put the matter, from the initial framework and transferred to another as Lehrer (2011b) formulates the transition.

The coupling and transfer is familiar enough from everyday experiences of art as well as science. Look at a painting of a historical figure, Madame Pompadour by Boucher, for example, notice how she is portrayed in the painting. You have a conception of her in the framework of the painting from your impression and thoughts of the painting. Now the question arises of whether to transfer your impressions and thoughts from the framework of the painting to the historical figure. You know what she is like in the framework of painting even as you turn away from the painting itself. For you know what your thoughts and impressions are like as you carry away your exemplarized reflexive representations of them. Now you may decouple those self-represented states and transfer them to your conception of her in the court of Louis XV. Whether you transfer them would influence how you acted in the court if you were a member of it, and your conception of the historical events in present day.

This familiar reflection about decoupling and transfer of the content of representation shows us how to bridge the gap between the conceptual framework of the manifest image and the conceptual framework of the scientific image. Having located thoughts and impressions in the framework and descriptions of the manifest image, you exemplarize those states into reflexive self-representations and in phenomenal profiles. But having located them in self-representation, you may decouple them from the words “thoughts” and “impressions”. Those states represent themselves whatever words we attach to them. As we transfer those self-representations to the discourse of science, to the discourse of “neural activation”, for example, they connect that discourse with the same fixed points of self-representation in our experience of our world, of ourselves, and of ourselves in our world. They provide an arch of representation connection between the conceptual framework of the manifest image and the scientific image. The unity of two frameworks does not consist of the reduction of one to the other or even the inclusion of one in the other. It consists instead of finding an arch of self-representation that connects the one with the other. Self-representation, exemplarization and reflexive self-description, can be decoupled from

discourse to connect the manifest image with scientific image with a representational arch.

As Sellars closed EPM, he asked concerning the myth of Jones,

But is my myth really a myth? Or does the reader not recognize Jones as Man himself in the middle of his journey from the grunts and groans of the cave to the subtle and polydimensional discourse of the drawing room, the laboratory, and the study, the language of Henry and William James, of Einstein and of the philosophers who, in their efforts to break out of discourse to an *arché* beyond discourse, have provided the most curious dimension of all. (EPM, 1963, p. 196)

The most curious dimension of them all, articulated by Sellars in his correspondence with Castañeda, may be the dimension of self-representation of inner episodes given birth to within the society of intersubjective discourse which, once mature, is free to move in a private or social manner from one form of discourse to another.

Finally, Sellars argues that the efforts to break out of discourse, which may succeed in taking us to self-representation, will not by itself take us to *knowledge* of the self-represented states. Knowledge, even noninferential knowledge of inner episodes, is tied to the framework of others concepts and claims within the framework «placing it in the logical space of reasons, of justifying and being able to justify what one says» (EPM, 1963, p. 169). We conclude with our concurrence revealed in our own writings on knowledge and acknowledge gratefully the precedence of *Empiricism and Philosophy of Mind*.

REFERENCES

- Castañeda, H., & Sellars, W. (1961–1962). *Correspondence between Hector Castañeda and Wilfrid Sellars on Philosophy of Mind*. Ed. by A. Chrucky. Available on-line at: <http://www.ditext.com/sellars/corr.html> Accessed on : April 15th 2012
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fürst, M. (2010). *A Dualist Version of Phenomenal Concepts*. In Draft.
- Goodman, N. (1968). *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Bobbs-Merrill.

- Ismael, J. (2007) *The Situated Self*. Oxford: Oxford University Press.
- Lehrer, K. (1996). Skepticism, Lucid Content, and the Metamental Loop. In A. Clark, J. Ezquerro & J. M. Larrazabal (Eds.), *Philosophy and Cognitive Science*, Dordrecht: Kluwer, 73–93.
- Lehrer, K. (2000). Meaning, Exemplarization and Metarepresentation. In D. Sperber (Ed.), *Metarepresentation*, a volume of Vancouver Studies in Cognitive Science. New York: Oxford University Press, 299–310.
- Lehrer, K. (2004). Representation in Painting and Consciousness. *Philosophical Studies*, 117(1–2), 1–14.
- Lehrer, K. (2007). Loop Theory: Knowledge, Art and Autonomy. John Dewey Lecture. *Proceedings and Addresses of the American Philosophical Association*, 81(2), 121–136.
- Lehrer, K. (2011a). What Intentionality is Like. *Acta Analytica*, 26(1), 3–14.
- Lehrer, K. (2011b) Art. Self and Knowledge. Oxford: Oxford University Press.
- Lehrer, K., & Stern, D.G. (2000). The “denouement” of “Empiricism and the Philosophy of Mind”. *History of Philosophy Quarterly*, 17(2), 201–216.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- Papineau, D. (2007). Phenomenal and Perceptual Concepts. In T. Alter & S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press, 111–144.
- Sellars, W. (1963). Empiricism and the Philosophy of Mind [EPM]. In W. Sellars, *Science, Perception and Reality*. New York: Humanities Press, 127–196.

Normativity and the Realist Stance in Semantics

Giacomo Turbanti *
turbanti.giacomo@gmail.com

ABSTRACT

Recent attempts to define and support realism in semantics seem to acknowledge, as the only defence from skeptical attacks to the notion of meaning, a flat acceptance of the existence of representational relations between language and things in the world. In this paper I reconsider part of the mistrust about the normative character of meaning, in order to show that some of the worries urging the realists to cling on representationalism actually rest on misconceptions. To the contrary, I suggest that normativity is the main strength of a stable realist stance in semantics. Support to this suggestion comes from the reanalysis of some oft-ignored sellarsian themes.

1. Introduction

1.1. Is meaning normative?

Is meaning normative? This question has been haunting philosophy of language at least since Wittgenstein's remarks on rule-following, but, after Kripke (1982), the debate on the normativity of meaning has been hinged, to a large extent, on the problem of how to abstract linguistic rules from a naturalistic account of speakers' behavior. The frustration of this enterprise usually leads to a stark choice. The first option is to accept the normativity of meaning, but to give up on semantics: meanings are normative but we don't grasp them good enough to represent them. The second option is to keep the semantic analysis of meanings, but to try defusing underdetermination issues by separating meaning from norms. I reject both these pigeon holes, and, in what follows, I'll try to unhinge the whole framework with a sellarsian lever. Just as any lever has three parts, so too has mine: the analysis of the "Myth of

* Scuola Normale Superiore, Pisa, Italy.

the Given” will provide a fulcrum, the functional analysis of meanings will carry the load of the framework, and the analysis of normative vocabulary will allow to apply force. Once the framework will be lifted up enough, we’ll be able to see that it is nourished by roots which dig deeper into epistemology, and that its semantic consequences are just some of its more visible fruits. In fact, I claim that realism in semantics is compatible with a non reductive account of the normativity of meaning.

1.2. The skeptical stance and the realist stance

Let me try to sketch, briefly, the rough outline of the relevant part of the debate on the normativity of meaning.

The first of the aforementioned options is usually set up as a *skeptical* stance, introduced along the well-known line of the arguments in Goodman (1954), Quine (1968) and Kripke (1982). Although these arguments differ in many respects, their barebone structure is basically the same, and it goes like this: consider some linguistic content and put forward a definition for it which is *adequate* to its public usage, then either you have some independent peg to hang an analysis of such an adequacy or, inevitably, another content is deliverable which satisfies your definition while being, in fact, incompatible with the first one. Several pegs have been proposed (for instance, logical ones like “projectibility”, naturalistic ones like psychology, pragmatic ones like linguistic communities), none of which has proven to be firm enough: hence the skepticism about meaning.

These arguments seem to be compelling, so the reaction against this unwelcome result tends to be drastic. The alternative option is usually set up as a militant *realist* stance. By this side of the debate, the skeptical argument is construed, at bottom, as a semantic staging of the “naturalistic fallacy” in ethics: if meaning is an evaluative notion, then no descriptive analysis may account for it. In this sense, however, the whole charge can be dismissed simply by noticing that normativity of meaning is an unsupported assumption. The burden of the proof falls back on the skeptics who have to show where the norms come from. Indeed, the realists claim, if there are norms in the nearby of meaning, they are *constitutive* of linguistic practices, in the sense that they *describe* what it is for an expression ϕ to mean F in a given practice, e.g.,

“ ϕ ” means $F \rightarrow \forall a (\phi \text{ adequately applies to } a \leftrightarrow a \text{ is } F)$ ¹

There’s no “*ought*” hidden in such a description: here adequacy has to be measured empirically and represented extensionally.

2. Mythbusting

My purpose in this section is to provide an Archimedean point outside the conflict of these two stances. I want to make room for two claims: (a) normativity of meaning doesn’t trigger skepticism (Section 2.1); (b) normativity of meaning can be represented from a realistic perspective (Section 2.2). Together, (a) and (b) draw the outline of an argument against an insidious misconception which hides inside the formal apparatus of first order quantified logic. In my opinion, this misconception is among the main causes of contemporary relapses of the form of the Myth of the Given clearly described in §30 of Sellars (1956):

[U]nless we are careful, we can easily take for granted that the process of teaching a child to use a language is that of teaching it to discriminate elements within a logical space of particulars, universals, facts, etc., of which it is already indiscriminately aware, and to associate these discriminated elements with verbal symbols.

2.1. The regress of interpretations

Doubts about Kripke’s account being a proper analysis of Wittgenstein’s views have often and correctly been raised, and yet it established a standard way to carve a skeptical argument in the *Philosophical Investigations*. I would push for a slight but insightful and consequence-laden adjustment of this carving, which I borrow from McDowell (1984).

McDowell pictures the bundle of problems Wittgenstein was dealing with as a complex dilemma: on the one horn, the familiar correspondentistic representation of truth as congruence between meanings and facts, that Wittgenstein rejects, and, on the other horn, the whole famous paradox of §201: «no course of action can be determined by a rule, because every course of action can be made out to accord with the rule». In this adjusted picture Wittgenstein’s argument does not rest in the formulation of a paradoxical

¹ This is slightly adapted from Hattiangadi (2007, p. 56).

conclusion. Rather, it draws a path moving from the rejection of the augustinian correspondentistic interpretation of meanings (the first horn of the dilemma), through the analysis of the difficulties of the paradox of rule-following (the second horn), to their eventual solution. Remarkably, the solution comes just in §201 of *Philosophical Investigations*, when Wittgenstein points out the misunderstanding which supports the whole dilemma: the way the paradox is generated – i.e., by providing explicit definitions for the meaning of a linguistic expression as interpretations of its public usage – preempts the realization that «there is a way to grasp a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases». Once this misunderstanding is dropped, one can avoid the paradox without committing to any mythical peg where to hang a permanently adequate description of meaning. And this means that skepticism can be defused right from the start.

This is enough for my purposes, so I'll drop McDowell's account here, since I'm not willing to follow him in his analysis of the noninterpretative way in which we grasp meanings. But I need to point out one more crucial thing. As it is well known, Wittgenstein's own conception of the noninterpretative grasp on meanings is a “bedrock” of linguistic behavioral practices. Such a bedrock is “fundamental” to any interpretation of linguistic content, in the sense that our inquiries into the adequacy of our definitions can't dig under it. Let's now try to add this idea to the picture. So, on the one side we reject the idea of self-authenticating non-verbal *data* which could grant the adequacy of a definition of meaning, on the other side we drop the requirement of *justification* for the meaningfulness (the following of a rule) of all verbal episodes: hence we are left with a bedrock of verbal episodes whose meaningfulness doesn't require an interpretation. But here's a worry: granted that our linguistic practices are meaningful and granted that we entertain contents, what can we say about them? Since the bedrock can't be interpreted, how can we represent contents? It seems that we can't do semantics. I turn now to defuse this worry.

2.2. The givenness of the logical space

Very roughly, in Sellars's account, the Myth of the Given is the idea of a prelinguistic sort of epistemic awareness of conceptual contents. This is often construed as the awareness of certain *qualia* – e.g., the preconceptual awareness of *green*, as opposed to *red*. Here, I'm pointing at a more

fundamental level of the Myth: the idea of a prelinguistic awareness of the *determination* – whether it be qualitative or not – of conceptual content. My purpose will be to ensure that, however fundamental the criticism against the Myth may be, it is compatible with a realist stance. In this section, I'll try to show how Quine, the great foe of empiricist dogmas, fell prey of the Myth of the Given. Recall that I need this piece of argument in order to disentangle the thesis that meaning is “inscrutable”, so let me focus on Quine (1968). The outcome of Quine's argument for the inscrutability of reference could be hastily sketched as follows: since “meaning” is too a vague notion for the naturalist, she tries getting a firmer grip on conceptual content by recurring to *extensionalism*, but she has to realize that reference is indeterminate as well. This story is well known, but let me rehearse it very briefly. Consider an extensional semantics in which a domain of objects D is *given* and concepts ϕ , ψ are interpreted as functions whose ranges are collections $K = \{a_1, \dots, a_n\}$ of objects in D , i.e.,

$$x \in K =_{\text{def}} x = a_1 \vee x = a_2 \vee \dots \vee x = a_n$$

and predication is defined as

$$\phi x =_{\text{def}} x \in K.$$

Now, in this framework the analysis of meanings reduces to the inspection of collections K s, whose set theoretical relations generally can be made explicit in terms of quantified sentences. So we *say*, for instance, that rabbits are mammals in terms of

$$\forall x(\phi x \rightarrow \psi x).$$

Going backwards, this is how we inspect the ontology of a theory through the quantificational apparatus. Such is the power of the extensional representation. But, according to Quine, this approach to the analysis of meaning is flawed, since it entirely depends on the givenness of the domain D . For suppose another domain D' was given: collections K' s could be rearranged on D' so to correspond to our concepts and satisfy each and every quantified sentence we used to express their relations. Again, the point is that, although we are provided with all the information expressed in the quantified sentences of our theory, still we can't tell whether ϕx is defined in terms of K or K' . Hence, Quine claims, ontology is *relative* in the sense that it can be specified only with

relation to a given background. The problem is obviously that no background domain is given to us in semantic analysis.²

Up to here I just described a familiar piece of the toolbox of the philosopher of language, now I want to show that it is prone to malfunctioning if it is not properly used. Quine construes the idea that the domain of a theory can be specified only against the background of another given theory in the following way:

We may picture the vocabulary of a theory as comprising logical signs such as quantifiers and the signs for the truth functions and identity, and in addition descriptive or nonlogical signs, which, typically, are singular terms, or names, and general terms, or predicates. Suppose next that in the statements which comprise the theory, that is, are true according to the theory, we abstract from the meanings of the nonlogical vocabulary and from the range of the variables. We are left with the logical form of the theory, or, I shall say, the *theory form*. Now we may interpret this theory form anew by picking a new universe for its variables of quantification to range over, and assigning objects from this universe to the names, and choosing subsets of this universe as extensions of the one-place predicates, and so on. (Quine, 1968, p. 204)

Clearly his concern is that the bare *theory form* provides no information to discriminate suitable models, in the absence of a given background domain where objects could be picked. Now, from a sellarsian perspective, the idea of a «theory form» sounds very suspicious: it amounts to the view that a logical backbone – a formal vocabulary of quantifiers, predicative letters and variables – might be sterilely transplanted from one theoretical body to another. Brandom described this sort of suspect in some detail, by providing reasons against Quine’s famous “gavagai” example.³ Assuming that there is an incompatibility between the sortals the theorist may use to translate “gavagai”, e.g., “rabbit” and “undetached rabbit part”, he must have at his disposal the linguistic resources to make such incompatibility explicit by contrasting two

² It may be (and it has been) objected that no background domain is given to us with the exception of the only one that really guarantees an objective extensional representation of contents: the domain of *all possible* particulars. From the perspective adopted in the present paper, this modal way out might be resisted by noticing that possible world semantics simply does not provide a definition of meanings which is adequate to their usage in linguistic practices, because actual speakers just have no suitable access to such a metaphysical domain. I’ll stick to this rejoinder here for two reasons. First, for the sake of my argument, I don’t want to just cut the gordian knot of adequacy. Second, in an important sense, as I am going to argue, there are *not enough* possible particulars to do the job.

³ See Brandom (1994, pp. 409–412).

general ways to reindividuate items of the different sorts. That is to say, if “gavagai” is a genuine sortal then it must apply either to individual rabbits or to individual undetached rabbit parts, and it must be so applicable in language. So, for instance, it must be possible to say that «“gavagai” *a* is the same as “gavagai” *b*» and check whether *a* can be substituted with *b* in contexts like «... is a mammal» or «... is a broken foot». Surely the theorist has the linguistic resources to do that. The question is: do the natives have similar resources? If they do, then obviously the translation wouldn’t be indeterminate. But if they don’t, then how could a sortal which doesn’t match the natives’ resources for reindividuation be a proper translation?

Notice these reasons do not amount to a refutation yet. Brandom looks at Quine’s example from his already refined inferentialist perspective, and his purpose is to defend *that* perspective from indeterminacy. Quine, from his own point of view, could have easily dismissed the question about the use of sortals by recurring to extensionalism. In fact, what guarantees the feasibility, *in principle*, of the procedure described by Quine in the above quotation is just the extensional definition of concepts: in order to see whether ϕ is incompatible with ψ it is enough to check whether K and K' are disjoint in DUD' .

But then, is there something wrong with extensionalism? This point is a tricky one. Sellars explicitly tackled it at the very end of Sellars (1957, §§ 102–108). Now, that is a mouthful of sellarsian philosophy, but here I’ll try to chew just the very bit I really need:

[T]he logic of variables and quantification involves not only the *momentary* crystallized content of the language at a cross section of its history, but also its character as admitting – indeed demanding – modification, revision, in short, development, in accordance with rational procedures. (Sellars, 1957, § 105)

Sellars makes two crucial claims here: (a) the extensional description of contents pictures an idealized phase of linguistic practices in which conceptual resources are completely and definitively made explicit; (b) such a description implies, indeed requires, the possibility of content to be continuously improved. I won’t be able to say anything about (b) until the end of this paper, so, for the time being, I’ll focus on (a).

This remark of Sellars’s highlights what goes wrong with Quine’s strategy: one just can’t recur to the quantificational apparatus as a sterile scalpel to carve ontologies out of theoretical bodies and compare them, because the

quantificational apparatus itself is part of those bodies. What does that mean out of metaphor? Consider a standard tarskian semantics and try to ask, for instance: what is a model for a sentence p ? It seems that we can provide two answers: (i) a model for p is an *interpretation* that gives to p a designated semantic value; (ii) a model for p is a *possible world* that contains the state of affairs represented by p . Indeed, the fact that (i) and (ii) can often be treated as equivalent is the key of the success of model-theoretic semantics. But that doesn't mean we can simply equate them: on the one side we have the interpretation of linguistic expressions, on the other side we have the representation of states of affairs. My purpose here is not to highlight a gap between (i) and (ii). To the contrary, I want to claim that since there is no gap and we make explicit the interpretation of language in terms of an extensional representation, we can't account for the variability in the determination of contents inside this representation: we'll never get different interpretations of the same concept, we'll always get just different concepts. In this sense, Quine fell prey of the Myth of the Given to the extent that he took the quantificational apparatus of a language as conceptually prior to the determination of its contents. However, and this is what Sellars's remark (b) is about, it is just because we provide an explicit interpretation of contents that we can modify them. So we can't just throw away semantic theory.

3. A theory of meaning

In the previous section I tried to open up some space for the possibility to endorse both a realist stance in semantics and a non-reductive account of the normativity of meaning. Now I want to substantiate this possibility. Let's start from scratch once again by asking: what do we do, when we do formal semantics?

3.1. Picturing and meaning

First of all it must be acknowledged that the formal semantics we've all learned, model-theoretic semantics, is couched in the representational tradition. Intuitively, the paradigm of representation is a *picture*. But consider a prototypical statement like "the particle a has spin- s ", and ask: what does it *represent*? The answer, clearly, will be that it represents the particle a as having a certain spin. Now, in a sense this means that the statements *says of a that it*

has spin-*s*. In another sense it means that the statement *pictures* a *complex object*, e.g., the *fact that a* has spin-*s*, in the same way as a plan represents a building (where a building is a complex object and a single brick is not). This latter sense is quite misleadingly supported by the formalization of the statement in first order logic, say “*S(a)*”: $*S(a)*$, as a sign design, is a complex object as well, so that $*S(*)*$ and $*a*$, as sign designs, are *part of* it. In this latter sense our statement is construed as a complex name for the complex object it pictures.

Sellars repeatedly denounced the risk that the confusion of these two senses may engender the confusion between asserting and picturing. In particular it leads to think that the representational purport of linguistic expressions can be explained in terms of a picturing relation holding between them and things they *name* in the world: indeed, while linguistic expressions do, in a sense, picture things in the world, it is this very fact that begs the question about how represent*ings* can point beyond themselves to represent*eds*.

3.2. Linguistic roles

So, we must also acknowledge that when we do formal semantics we do not, in the first instance, *describe* a relation between a language and the world. But then, what are meaning statements about? Consider again:

“ ϕ ” (in L) *means* F

Let me lay down some platitudes. First, “ ϕ ” on the left, a quotation of the sign design $*\phi*$, is the name of an expression in language L. Second, F on the right can’t be the name of a linguistic expression in L as well, on pain of regress. Third, however, if F were not a name then either it would not occur in a truthfunctional sense or *meaning* would not be a relation (but then what would it be?). So F must be the name of something, and it must be something we have some knowledge of in order for meaning statements to be explanatory. Indeed, F is usually construed as an expression of *our* meta-language. But what does it mean to have some knowledge of an expression of our language? It means that we know how to use it: we know how to deploy it with relation to the context, we know its *linguistic role*. Linguistic roles provide a functional classification

of linguistic sign designs. If we adopt Sellars's dot-quotation⁴ to single out these roles, we can provide an alternative interpretation of meaning statements:

“ ϕ ” (in L) means $F \rightarrow$ “ ϕ ” (in L) is an $\bullet F \bullet$.

Distributive terms like $\bullet F \bullet$ designate (both intra- and inter-linguistically) expressions which are used in the same way, that is to say, those expressions which occupy the same place in the net of relations – paradigmatically, inferential relations – established on linguistic sign designs by their use in linguistic practices. Notice that it's easy to accommodate in this framework the semantic realist's interpretation of linguistic rules, since part of the linguistic role of $\bullet F \bullet$ s might be, for instance, to apply whenever $\bullet f \bullet$ s apply as well. Notice also that all this is compatible with representational semantics – indeed, this is purported to be just its correct interpretation.

So what we do when we do formal semantics is to provide a model for linguistic roles in order to *explain* the use of linguistic sign designs. Where the world comes into the picture is in the evaluation of the model. As it was clearly stated just in Tarski (1944), formal semantics is quite independent from the ideas we may entertain about the *nature* of semantic contents, but, obviously, that doesn't mean that formal semantics doesn't explain anything about language and mind. In fact it's easy to see how models work here. In Section 3.1 we've noticed that we can talk of sentences as complex sign designs that *picture what it is*, without implying that this fact should explain how they *say of what it is that it is*. We've acknowledged that the explanation runs in the opposite direction: it is because we use certain sign designs according to certain rules in order to say something, that those sign designs picture it (rather than something else). So, semantic models do not apply to a picturing relation between expressions and things in the world, because that would not

⁴ Just to briefly sum up, dot-quotation applies to expressions in a given *familiar* language to build *distributive singular terms* referring to any expression in any language that play the same linguistic role of the quoted expression. So, as the *distributive singular term*

the pawn

refers to any *piece* (however materially realized) that is subject to certain rules in a chess game, in the same way the *distributive singular term*

\bullet triangular \bullet

refers to any sign design (however linguistically realized) that is subject to certain rules in a language game. For further details see Sellars (1963a).

provide an explanation of linguistic sign designs. That is, to repeat, meaning statements do not establish relations between linguistic items and non-linguistic items. Instead, semantic theory provides models which represent *linguistic roles* in order to explain the use of linguistic sign designs.

Thus, to take the realist stance as to shortcircuit the explanation of the contents by interconnecting the linguistic domain of sign designs with the non-linguistic domain of the model is just a mistake.

3.3. The realist scruple

The notion of linguistic role might seriously worry the naturalistic biases of the semantic realist, and she may be willing to protest that I'm just weighing her down with ontological burdens while refusing to answer the only relevant semantic question: how are linguistic expressions related to things in the world?

What I would be missing can be best highlighted by appealing to common sense. So, suppose one morning I wake up with a terrible rash. I go to the doctor who, after checking me up, declares: "you have chicken pox!" So, I adopt all the necessary cures, which involve, according to Italian lore, to devote myself to Saint Anthony. My behavior ensues from the fact that "chicken pox", in the doctor's claim, means *chicken pox*, rather than, for instance, *encephalitis*, and "you" means *me*. Otherwise I would have had to devote myself to Saint Paul, or simply do nothing at all.⁵ In other words, the complaint is that I'm wavering on the notions of *Reference* and *Truth*. But, as a matter of fact, it is just at this point of the analysis that such notions can be properly introduced. And the reason is that we can now see clearly what they are not. On the other hand, a proper account, as it's easy to realize, would greatly exceed the space of this paper, so I'll have to be very schematic.

Let's begin with *Truth*: it is not a relation between sentences and states of affairs, nor a relation among sets of propositions. The isomorphism between the structure of linguistic sign designs and the structure of states of affairs is a necessary yet not sufficient requirement for the picturing relation between language and world to hold. If the meaning of a propositional sign design is its propositional linguistic role, then to say that a sentence is true is to say that it is *correctly* used according to its role – paradigmatically, when it is correctly

⁵ The example is ironically adapted from the introduction of Hattiangadi (2007).

asserted according to the inferential rules of the language. Notice that these inferential rules involve the *context*, and, in particular, the consideration of both possibly perceptual premises and possibly practical consequences of the assertion: it is crucial, in order to understand how the uses of expressions get in touch with the world, to take into account in this sense full-fledged linguistic practices. Notice also that these rules establish the validity of not only *formal*, but also *material* inferences. The sellarsian distinction between formal and material inferences is defended in Sellars (1953). Here a concise but harshly bolzanian definition could be put as follows: formal inferences are valid because they are substitutional occurrences of patterns in which “logical” vocabulary is fixed, while material inferences are valid because they are substitutional occurrences of patterns in which “non-logical” vocabulary is fixed.

Now, all this holds for subsentential expressions as well, so, in a sense, this is also account of *Reference*. But such an account, as it stands, would be incomplete in two main aspects. First, subsentential roles are not all of the same sort. So, just to get the idea, consider again the sign design $*S(a)*$: part of “*a*” being an $\bullet a \bullet$, as contrasted to being an $\bullet S(\) \bullet$, is that “*a*” is a singular term. This difference must be accounted for in some way, and it must be explained how the linguistic role for sentential sign designs like $*S(a)*$ is composed by the linguistic roles of $\bullet S(\) \bullet$ s and $\bullet a \bullet$ s.⁶ Second, our use of sortals elicits regularities which pertain to our concepts of thing-kinds (or “essences”). The way in which the linguistic role of sortals establish those regularities must be accounted for as well. In Sellars (1957) that is done in terms of the modally robust, and yet defeasible, inferential rules which define the linguistic role of sortals. So, for instance, the inferential rules which define the linguistic role of $\bullet \text{match} \bullet$ may involve conditionals like

⁶ This point has been often construed as revealing the main technical flaw of the whole picture: the semantic structure and the syntactic structure have to be isomorphic, such is the requirement of *compositionality*, but linguistic roles do not compose. However, this is wrong. To begin with, Chapter 6 of Brandom (1994) shows how to evaluate the inferential role of subsentential expressions by exploiting just compositionality. Still, some doubts are raised by the exploitation of compositionality in an holistic framework (Fodor and Lepore, 2001, 2007): the proof of inferential conservativity for the logical vocabulary of *Incompatibility Semantics* in Brandom (2008) shows how they can be dismissed. In this respect, the general point to be realized is that, from a formal point of view, compositionality is but a trivial problem, as it is clearly explained in Westerstahl (1998). More interesting issues rise when the semantic and syntactic structures are already defined according to other theories and independent assumptions.

“This match would light if scratched”,

i.e., they entitle one to move, defeasibly, from the application of a **•match•** and a **•scratched•** to an application of a **•lighted•**.

To sum up, semantic notions like *Reference* and *Truth* are *normative* notions, in the sense that they are employed to specify correct use.

Here, a common misunderstanding should be dismissed straight away. As a general objection against this analysis of meaning in terms of rules, one may notice that it is just a mistake to try to define semantic notions in terms of *prescriptions* on the use of expressions, since these latter have to do with the pragmatics of language.⁷ This worry is legitimate, but, *qua* objection, it loses all its bite as soon as it is noticed that it is grounded on the confusion that consists in interpreting rules *directly* as patterns of behavior. Behavior involves *actions*, and rules for linguistic actions are obviously of the pragmatic sort, like: *tell the truth!* Now, while linguistic episodes can (and usually do) manifest themselves in terms of overt verbal performances, the rules that define their contents in the sense here intended are not, to use a Sellarsian turn of phrase, rules of performance, but rules of *criticism*. That is, they do not specify what one *ought to do*, but what one *ought not*. So are, for instance, the rules of the Highway Code: they do not say where one ought to go, but they say that one ought not to cross a red light. Since linguistic rules are usually construed as inferential rules, I'm afraid that sometimes this confusion may be backed by a certain hasty suggestion about the notion of *following logically*, a suggestion already mocked by Lewis Carroll but rather die hard, according to which a logical inference, e.g., from *p* to *q*, amounts to a *prescription* of asserting *q* once *p* is asserted (or, at least, to believe *q* once *p* is believed): this, as the tortoise tried to explain to Achilles, is obviously nonsense.

4. Normative vocabulary

Almost every piece of my argument is in its place, and still, it seems, my results are quite poor. Even if I was successful in the criticism of the standard picture of the debate about normativity of meaning, I haven't yet indicated any other practicable way to get off the ground in semantics. In order to provide an alternative I need to introduce, as the last bit of my argument, the analysis of the role of normative vocabulary.

⁷ See for instance Wikforss (2001), Hattiangadi (2007), Glüer and Wikforss (2009).

The fundamental role of normative vocabulary eventually comes to the fore in the attempt to cash out the notion of *correctness* established by the rules which define linguistic roles. In fact, if we were to ask how to evaluate the correctness of a particular application of the rules, immediately we would have to face the same old regress arising again and again. However, the key to defuse the regress once and for all lies just in a proper analysis of normativity.

Before we go ahead, let me briefly recapitulate. Formal semantics provides models to explain linguistic behavior, but this theoretical enterprise seems to be bound to failure because of the irreducibility of the normative character of human behavior to descriptions of matters of fact – such is the wittgensteinian image of the bedrock. In this paper we have established (i) that neither the idea of such a bedrock nor the possibility of its theory is incoherent, and (ii) that, once a functional characterization of contents is accepted, the realist stance is compatible with a theory of linguistic roles. Still, it must be clarified how the normative analysis of linguistic roles may fit into the explanation of linguistic behavior provided by formal semantics. In other words, through the functional characterization of conceptual contents in semantics we may obtain scientific theories whose models explain our linguistic behavior, but still fall short of explaining it as rational behavior.

But now I have all I need to start the last bit of my sellarsian analysis. As it's easy to notice, the problem we face here is but a particular occurrence of one of the major themes of Sellars's: he described it in Sellars (1962) as the problem of fusing in a "stereoscopic vision" the "manifest" and the "scientific image of man in the world".⁸ I won't try to approach a proper analysis of the general theme, but I hope that my account of this particular instance was clear enough to let the reader understand how Sellars's solution applies here. Thus, the key to obtain such stereoscopy is the following:

[T]o complete the scientific image we need to enrich it *not* with more ways of saying what is the case, but with the language of community and individual intentions. (Sellars, 1962, p. 78)

Here the notion of community intention is crucial, so it's worth taking a break and making it clear. First, the notion of intention is, as expected, a normative notion. In this context, intentions are not dispositions to act. Rather, they are the sort of things which can move rational beings, those who are sensitive to the force of reasons. Thus, for instance, it is *because* you have an intention

⁸ See Sellars (1962, pp. 40–41).

whose content is expressed by “I shall raise my arm” that you raise your arm. As such, intentions are evaluated in terms of *reasonableness* with relation to practical reasoning. For instance, we express the fact that my intention (not) to raise my arm is reasonable in circumstances of kind *C*, by saying “I ought (not) to raise my arm, in circumstances of kind *C*”. Second, the evaluation of the reasonableness of intentions poses the problem of *objectivity*. Notice that this might be a harmful trigger for the regress of interpretations. However, and this is the crucial point, at the level of intentions, a direct solution to the regress can be provided.

Let’s begin by asking what it means to generalize subjective intentions. The idea is to move from principles of the form

I ought to do actions of kind *A*, in circumstances of kind *C*

to principles of the form

Anyone ought to do actions of kind *A*, in circumstances of kind *C*.

Such a generality can’t be achieved by intentions of the form

Everyone shall do actions of kind *A*, in circumstances of kind *C*,

for we know that moral principles, *qua* normative principles, are not reducible to what everyone does. So our question turns out to deal with the analysis of

Anyone shall do actions of kind *A*, in circumstances of kind *C*.

Here, Sellars suggest, “anyone” refers to any of those who *share* the intention. To share an intention in this sense is not to have the *same* subjective intention, but to have a community intention of the form

We shall do actions of kind *A*, in circumstances of kind *C*.

It is the sharing of intentions that constitutes a community.⁹ The constitution of a community of rational agents who share intentions supports the application of normative values to every domain pertaining to their agency, in particular to the epistemic and the practical domains. It is in this sense that Sellars’s solution to the problem of fusing the images consists in *extending* the descriptive vocabulary of theoretical representations of the world by introducing the normative vocabulary of communities of rational agents.

Now let’s go back to our argument. In order to accept Sellars’s suggestion here we have to block two preliminary objections.

⁹ See Chapter VII of Sellars (1968).

The first objection is a rehearsal of a worry we already considered in Section 3.3: does this approach commit us to intentions *qua* mental sort of things? The answer is no, because we only need to accept into our models theoretical objects which can be defined as functional analogues of sign designs. This, notice, entitles us to dismiss the hardest part of Sellars's own problem, i.e., the reception of *persons* inside the scientific image: since we are interested in formal semantics, we have just to deal with the functional analysis.

The second objection may be formulated as a request for clarification: Kripke already invoked the perspective of the community but failed to establish it as anything more than an intersubjective point of view, so how is Sellars's proposal different? The answer is that the two proposals are indeed pretty different: according to Kripke, the community establishes the horizon of the generality of norms, while, according to Sellars, the universality of norms establishes the horizon of the community. In this sense, the very task of making the contents of the norms explicit is the task of constituting the community of the agents who follow them. This is probably the hardest point to acknowledge, but it's also probably the most important: surely, without it the whole approach would be idle.

Then, the application of Sellars's solution to the semantic domain is quite straightforward. In fact, *Truth* turns out to be the sort of reasonableness of community intentions which pertain to the semantic domain, and its evaluation has to be made explicit in terms of normative principles pertaining to linguistic performances. Typically these will be principles of the form

One ought not to refuse to apply " ψ ", if one accepts to apply " ϕ ".

These principles establish relation among the sign designs – e.g., $\ast\phi\ast$, $\ast\psi\ast$ – which are employed in the practice they regulate. In this sense each sign design is identified by the *functional role* – e.g., $\bullet\phi\bullet$, $\bullet\psi\bullet$ – it acquires in these relations. The goal of formal semantics is to represent these linguistic roles in terms of relations among the elements of the model of a semantic theory.

5. Concluding remarks

I'm afraid that I haven't put forward any really new thesis in this paper. Or at least, it seem to me that I've just collected pieces of reasoning which had been already in good sight there on the table, although they are often ignored in the debate about the normativity of meaning. My attempt here was to tidy things up a bit: sometimes that is enough to put them back into good use. Thus, I hope

that the alternative realist stance was stated neatly enough. In this last section I want to suggest some reasons to adopt it. I already pointed out its assets with relation to some traditional problems in philosophy of language, so what I still have to show is that it is suitable to satisfy the prior needs of the realist. In other words, the question I still want to ask is: is it really a *realist* stance? Let me pick a paradigmatic list of the *desiderata* for a realist stance just from a scholar who aims to defend semantic realism against skeptical attacks. In the introduction of Hattiangadi (2007) the position of the semantic realist is basically summarized in two tenets: (a) to understand the meaning of an expression is to grasp its correctness conditions; (b) ascriptions of meaning are subject to correctness conditions as well.

Indeed, the approach presented in this paper satisfies both these *desiderata*. With respect to (a), the present approach takes meanings to be defined by the inferential rules of a language and to be represented in terms of linguistic roles: according to the present approach to understand the meaning of an expression is to grasp its linguistic role, which represents the conditions for its correct application. Therefore the present approach satisfies tenet (a). With respect to (b), the present approach takes the objectivity of the rules of language to be defined by the horizon of the community of speakers who share the same normative space: according to the present approach the correctness of meaning ascriptions is evaluated in terms of the linguistic norms of the community of speakers. Therefore the present approach satisfies tenet (b).

In spite of this, the semantic realist might still feel unsatisfied. The reason is that there is a third idea implicit underneath tenets (a) and (b), the idea that correctness conditions have to be checked against some sort of “fact of the matter”, as Hattiangadi suggests. Depending on how the notion of “fact of the matter” is construed, this idea may turn out to be a third tenet (c) characterizing the semantic realist’s position. There is a sense in which the present approach satisfies tenet (c) as well. The present approach takes formal semantics to picture states of affairs by representing semantic rules pertaining to them, so that according to the present approach there’s a precise sense, formally specifiable in terms of semantic models, in which correctness conditions are checked against facts of the matter. In this sense, however, tenet (c) is already contained in (a) and (b). There is obviously another way to maintain (c). It consists in construing the notion of “fact of the matter” as a given source of conceptually determinate information available out there (or in here) to be grasped. It’s just one major credit of Sellars’s that of having

delivered to contemporary philosophers the consciousness that such a given source is a myth. In this latter sense (c) is an autonomous tenet, which, however, I think, doesn't (and shouldn't) belong to a characterization of semantic realism. Actually, in this latter sense (c) has often been construed as marking the boundary between realism and antirealism. In fact, it's always been the antirealist strategy, at least since Gorgias, that of denying (a) and (b) by casting doubts on the existence of objective facts of the matter. But, as we have seen, it is just a mistake to believe that objectivity is given independently of normative practices, and the semantic realist has no need and no gain in following her opponent in this mistake.

I wish to conclude by adding that the sort of semantic enterprise envisaged at the end of the previous section is not is not just wishful thinking. On the one side, the inferential analysis of meaning is an ongoing logical enterprise being presently developed, among others, by Jaroslav Peregrin and by Dag Prawitz and his followers.¹⁰ On the other side, the account of normativity providing the interpretation for inferential relations has reached some practicable results, mainly due to Robert Brandom's elaboration of sellarsian themes.¹¹

REFERENCES

- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Brandom, R. (2008). *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.
- Fodor, J., & Lepore, E. (2001). Brandom's Burdens: Compositionality and Inferentialism. *Philosophy and Phenomenological Research*, 63(2), 465–481.
- Fodor, J., & Lepore, E. (2007). Brandom Beleaguered. *Philosophy and Phenomenological Research*, 74(3), 677–691. Reprinted in Weiss & Wanderer (2010).
- Glüer, K., & Wikforss, Å. (2009). Against Content Normativity. *Mind*, 118(469), 31–70.

¹⁰ See Prawitz (2006) and Peregrin (2009, 2010).

¹¹ See, again, mainly Brandom (1994, 2008).

- Goodman, N. (1954). *Fact, Fiction & Forecast*. London: University of London.
- Hattiangadi, A. (2007). *Oughts and Thoughts: Rule-Following and the Normativity of Content*. Oxford: Oxford University Press.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
- McDowell, J. (1984). Wittgenstein on Following a Rule. *Synthese*, 58, 325–364. Reprinted in McDowell (1998).
- McDowell, J. (1998). *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
- Peregrin, J. (2009). Inferentialism and the Compositionality of Meaning. *International Review of Pragmatics*, 1(1), 54–181.
- Peregrin, J. (2010). Inferentializing Semantics. *Journal of Philosophical Logic*, 39(3), 255–274.
- Prawitz, D. (2006). Meaning Approached Via Proofs. *Synthese*, 148(3), 507–524.
- Quine, W.V.O. (1968). Ontological Relativity. *Journal of Philosophy*, 65(7), 185–212. Reprinted in Quine (1969).
- Quine, W.V.O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Sellars, W. (1953). Inference and Meaning. *Mind*, 62(247), 313–338. Reprinted in Sellars (1980) and Sellars (2007).
- Sellars, W. (1956). Empiricism and the Philosophy of Mind. In H. Feigl & M. Scriven (Eds.), *Minnesota Studies in the Philosophy of Science*, vol. I. Minneapolis: University of Minnesota Press, 253–329. Reprinted with additional footnotes in W. Sellars (1963), *Science, Perception and Reality*. New York: The Humanities Press, 127–196. Republished as a volume (1997) with an introduction by R. Rorty and a study guide by R.B. Brandom, Cambridge, MA.: Harvard University Press.
- Sellars, W. (1957). Counterfactuals, Dispositions, and the Causal Modalities. *Minnesota Studies in The Philosophy of Science*, II, 225–230. Reprinted in part in E. Sosa (Ed.) (1975), *Causation and Conditionals*.

Oxford University Press.

- Sellars, W. (1962). Philosophy and the Scientific Image of Man. In R.G. Colodny (Ed.), *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press, 35–78. Reprinted in Sellars (1963b) and in Sellars (2007).
- Sellars, W. (1963a). Abstract Entities. *Review of Metaphysics*, 16, 627–671, Reprinted in Sellars (1967); reprinted in Sellars (2007).
- Sellars, W. (1963b). *Science, Perception and Reality*. New York: The Humanities Press. Reprinted by Ridgeview Publishing Company, Atascadero, California, 1991.
- Sellars, W. (1967). *Philosophical Perspectives*. Springfield, ILL: Charles Thomas. Reprinted as : W. Sellars (1977). *Philosophical Perspectives*, 2 voll. (vol.1: *History of Philosophy*; vol. 2: *Metaphysics and Epistemology*). Atascadero, CA: Ridgeview Publishing Co.
- Sellars, W. (1968). *Science and Metaphysics: Variations on Kantian Themes*. New York: Humanities Press.
- Sellars, W. (1980). On Reasoning About Values. *American Philosophical Quarterly*, 17(2), 81–101.
- Sellars, W. (2007). *In the Space of Reasons: Selected Essays of Wilfrid Sellars*. Cambridge, MA: Harvard University Press.
- Tarski, A. (1944). The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, 4(3), 341–376.
- Weiss, B., & Wanderer, J. (Eds.) (2010). *Reading Brandom: Making it Explicit*. London: Routledge.
- Westerståhl, D. (1998). On Mathematical Proofs of the Vacuity of Compositionality. *Linguistics and Philosophy*, 21(6), 635–643.
- Wikforss, Å. (2001). Semantic Normativity. *Philosophical Studies*, 102(2), 203–226.
- Wittgenstein, L. (2009). *Philosophical Investigations*, 4th edition. Trans. and ed. by P.M.S. Hacker and J. Schulte. Oxford: Blackwell.

Sellarsian Synopsis: Integrating the Images ^{*}

Jay L. Garfield[†]
jgarfield@smith.edu

ABSTRACT

Most discussion of Sellars' deployment of the distinct images of "man-in-the-world" in *Philosophy and the Scientific Image of Man* focus entirely on the manifest and the scientific images. But the original image is important as well. In this essay I explore the importance of the original image to the Sellarsian project of naturalizing epistemology, connecting Sellars' insights regarding this image to recent work in cognitive development.

«To say that man is a rational animal is to say that man is a creature not of habits, but of rules. When God created Adam, he whispered in his ear, "In all contexts of action you will recognize rules, if only the rule to grope for rules to recognize. When you cease to recognize rules, you will walk on four feet»

(Sellars, *Language, Rules and Behavior*, §15)

Introduction

In *Philosophy and the Scientific Image of Man* (1963b), perhaps the best introduction to Sellarsian philosophy available, Sellars distinguishes between three images of man-in-the-world. Most subsequent philosophers, myself included, who have written about the relationship between the images have focused on the two to which Sellars himself gives the greatest emphasis – the

^{*} The ideas developed here emerged from long dialogue with Dr Marco Fenici. I thank him for that interchange, from which I learned a great deal.

[†] Smith College, Northampton, MA; The University of Melbourne, Australia; Central University of Tibetan Studies, Sarnath, India.

manifest and the *scientific* – in the quest either to give one precedence over the other (Churchland, 1986) or to vindicate a binocular vision (Garfield, 1988; deVries, 2005; Rosenberg, 2007). But there is a third image in play as well, one to which Sellars himself gives relatively less attention, and which his successors have ignored almost entirely – the *original* image. This is unfortunate, for the project of naturalizing epistemology on Sellarsian lines, and making sense of the locus of normativity in the natural world, requires attention to all three images. In the present essay I will take Sellars' vision beyond stereoscopy, bringing all three images into play in the quest for an epistemology fully naturalized on Sellarsian lines, consistent with evidence from contemporary developmental and cognitive psychology.

The *manifest* image is the view of the world and of our place in it delivered by sophisticated common sense. It is the view that philosophical speculation attempts to refine. Most importantly, it is the image in which we are present as *persons*, beings who institute and respect norms in our thought, action and social arrangements, in which meaning emerges as a property of language and thought, and in which this normativity and meaning emerges in the context of a natural world that is in general governed by purely descriptive natural laws, and which is in general devoid of semantic content.

The *scientific* image is the world as our best science represents it. It is the world of microphysics, of cosmology, of chemistry and biology. It is a world of natural phenomena governed by natural law. Most importantly, the scientific image is an image devoid of *persons*, devoid of normativity, and devoid of meaning. In this image we have been explained away through the categories of the life and physical sciences in which the categories of normativity and meaning are not to be found. These images are, although from the perspective of the manifest image, *toto genere* different and apparently irreconcilable, from the standpoint of the scientific image, quite continuous. Sellars notes: «[T]his difference in level [of description in the two images] appears as an irreducible discontinuity in the *manifest* image, but as, in a sense requiring careful analysis, a reducible difference in the *scientific* image» (1963a, p. 6). This continuity, and the possibility of joining the images, is essential to the project of making real sense of the possibility of knowledge, sense that takes both its normative and biological dimensions seriously.

While the scientific image differs from the manifest in being *devoid* of persons and the conceptual categories they implicate, the *original* image is the image in which *everything*, or at any rate, everything *salient* is a person. In this

image storms are intentional agents, trees and non-human animals are partners in dialogue, and the universe is animated with meaning and intention. Sellars suggests that this is a primitive image, one principally of historical interest from which humankind (or at least that part involved in modern and post-modern culture) long ago emerged, and of interest only by way of contrast.

[T]he refinement of the ‘original’ image into the manifest image, is the gradual ‘depersonalization’ of objects other than persons. ...

A primitive man did not *believe* that the tree in front of him was a person, in the sense that he thought of it both as a tree *and* as a person, as I might think of this brick in front of me as a doorstep. If this were so, then when he abandoned the idea that trees were persons, his concept of a tree could remain unchanged, although his beliefs about trees would be changed. The truth is, rather, that *originally*, to be a tree was *a way of being a person*, as to use a close analogy, to be a woman is a way of being a person, or to be a triangle is a way of being a plane figure. (Sellars, 1963a, p. 10)

But as we will see, the original image may be *original* in an *ontogenetic* as well as a sociogenetic sense; the degree to which even in development, even in modernity, we transcend it may be more limited than we suppose; and its importance as a basis for both the manifest and scientific images, and for the life of persons as persons may be more synchronic and more pervasive than Sellars himself realized.

In what follows, I will begin by considering the mutual presupposition of the manifest and scientific images, and the importance of this interdependence for understanding the place of normativity in the natural world and the demands that normativity makes on our understanding of that natural world. I will then consider the sense in which the original image forms not just a mythic-historical backdrop for the two more familiar images, as Sellars’ own presentation suggests, but also a psychological, ontogenetic and evolutionary understanding of the empirical conditions of the possibility of distinctively human life. My discussion is not meant to be a reconstruction of what the man Wilfrid Sellars actually argued, but rather a hermeneutic argument for the best way to take up the set of distinctions he introduces in the service of the account of meaning, knowledge and human cognitive life he articulates in his corpus as a whole, an approach to reading this text which Sellars the man would have introduced wholeheartedly. (He was wont to say, commenting on the

advantages of historical perspective in hermeneutic practice, «we understand Plato far better than Plato could have understood himself»).

1. The Interdependence of the Manifest and Scientific Images

Churchland (1986) argues that the scientific image is the final arbiter of reality, in a Spinozistic sense the locus of the ultimate explanation of the approximate truth of the claims made by the manifest image, and in the end, destined to replace the manifest image as a way of understanding the world and the place of humanity in it. On this eliminativist view, the manifest image is just what we do until the scientist comes along, destined for the scrapheap of cultural history, just as the original image – on his view – has been scrapped. Persons, Churchland and other eliminativists argue, are no more real than storm gods; intentional or normative predication of members of *Homo sapiens* or their behavior is no more apposite than similar predication of tides and their behavior.

The motivations for this view are clear, and indeed have *some* basis within Sellars' own thinking. If science is the ultimate measure of the real, and if the description of the world science delivers is at odds with that delivered by sophisticated common sense, we should discard common sense in favor of science. If, for instance, common sense tells us that whales are fish, and science that they are mammals, we discard common sense and go with science. And, after all, science indeed tells us that we are nothing but collections of atoms in the void, insignificant moments in an insignificant, law-governed universe, best understood in the terms of physics, chemistry and biology, none of which has time to talk about persons, let alone moral or semantic value.

I have responded to this view at length (Garfield, 1988, 2000) and will not rehearse those arguments in detail here (see also Rosenberg, 2007). But the main points of the reply are easy to outline, and are all suggested explicitly by Sellars, either in *Philosophy and the Scientific Image of Man* or elsewhere. The scientific image cannot dispose of the manifest image because it *presupposes it*. Science aims at knowledge, and knowledge is *justified true belief* (plus or minus a bit of Gettier). Justification is a *norm-governed activity*, belief is a *meaningful, personal* state. Science itself is an intensely norm-governed activity, and its deliverances are theories, which, if they are to explain, must be both *justified* and *meaningful*. It is hence a transcendental condition of the possibility of the activity of science, and hence of the vindication of the

scientific image itself that the manifest image be in place as the context for scientific endeavor. Independent of the categories of the manifest image, science simply wouldn't be *science*, and here would be no *reason* to adopt the scientific image.

This transcendental argument is obviously Kantian in structure. It is a version of the transcendental deduction of the categories – the demonstration that without a rule-deploying subjectivity no knowledge is possible, recapitulated in Sellars' *Language, Rules and Behavior* (1949) and *Some Reflections on Language Games* (1954). And that is no accident, of course: Sellars was nothing if not Kantian. His own conception of the ineliminability of the normative was drawn directly from Kant (see the epigraph to this essay). We might also note that there is a second reason that the scientific image cannot dispose of the manifest. The scientific image contains the resources to *explain* the possibility of the manifest image, of our norm-governed behaviour, and hence of science itself. Explanation in science is a form of ontological *vindication*, not a reason for elimination.

But this is not to give pride of place to the manifest image. For just as the scientific image presupposes the manifest, the manifest image presupposes the scientific image as its extension and completion, and this regulative role that science plays in our very self-conception is one of Sellars' most profound extensions of Kantian ideas – in this case ideas drawn not from the *Transcendental Analytic*, but rather from the *Ideal of Pure Reason*. The manifest image is the locus of our awareness of ourselves as bound by norms, including centrally *epistemic* norms. These epistemic norms come to us sometimes in the form of what Kant would have recognized as *categorical imperatives*, in this case, imperatives to come to know, to understand. But of course in Sellars' hands, the analysis of their imperative force, and hence of their normativity, has a social dimension:

[T]he essentially social character of conceptual thinking comes clearly to mind when we recognize that there is no thinking apart from common standards of correctness and relevance, which related what *I do* think to what *anyone ought* to think. The contrast between '*I*' and '*anyone*' is essential to rational thought. [...] A group isn't a group in the relevant sense unless it consists of a number of individuals each of which thinks of himself as '*I*' in contrast to '*others*'. Thus a group exists in the way in which members of the group represent themselves. Conceptual thinking is not by accident that which is *communicated* to others, any

more than the decision to move a chess piece is buy accident that which finds an expression in a move on a board between two people. (Sellars, 1963a, pp. 16–17)

This social dimension turns out to be essential to the naturalization of epistemology. Sellars himself saw that it had to be; as we will see, contemporary cognitive science vindicates that verdict, though in surprising ways that make the links to Kant explicit.

We encounter ourselves in the manifest images as persons. Persons cannot, as Schopenhauer (1813/2003) pointed out, experience anything without asking “*why?*”. We will return below to the psychological dimension of this demand. But nothing we say or could say under that more empirical head can undermine the fact that the demand for explanation, for deeper understanding, is *experienced* in the first instance *as a demand*. To refuse to inquire, to refrain from demanding understanding, is to recuse oneself from the epistemic community or persons. Not only, as Kant put it, does “all our knowledge begin with experience”, but all experience is but the beginning of knowledge.

The manifest image hence contains – in virtue of our representation of ourselves as persons, in virtue of the necessity of persons to constitute and to conform to norms, and in virtue of the fact that those norms include not only moral and linguistic, but epistemic norms – the seeds of systematic inquiry whose flowering is the institution and practice of science, the fruit of which is the scientific image.

Thus the conceptual framework of persons is the framework in which we think of one another as sharing the community intentions which provide the ambience of principles and standards (above all, those which make meaningful discourse and rationality itself possible) within which we live our individual lives Thus the conceptual framework of persons is not something that needs to be *reconciled with* the scientific image, but something to be *joined* to it. (Sellars, 1963b, p. 40)

Without the *telos* of scientific understanding, the manifest image is incomplete. These two images are hence not only *mutually consistent* – as many have argued – and not only *complementary* in developing a binocular, and hence more complete vision of the world and of humanity within that world – as many have also argued – but are also mutually *entailing*, and each

presupposes the other as its transcendental condition. Sellars hints at this point:

But if in Spinoza's account, the scientific image, as he interprets it, dominates the stereoscopic view (the manifest image appearing as a tracery of explainable error), the very fact that I use the analogy of stereoscopic vision implies that as I see it the manifest image is not overwhelmed in the synthesis. (Sellars 1963b, pp. 8–9)

2. The Original Image

But what of the third image – the original image? The original image, as we have seen, is introduced by Sellars as a kind of historical myth of origin of human civilization. Seen this way, the original image is something we have collectively outgrown in a trajectory leading from shamanism to science. It is easy then, to pass over the original image in reading *Philosophy and the Scientific Image of Man*, and easy to refrain from asking why Sellars introduces it in the first place, given the focus of the essay on the tension between the manifest and scientific images. But the original image repays close attention and careful reflection, and indicates a tension in Sellars' own thought that we may be in a position to resolve productively in the service of his greater intellectual vision.

Sellars characterizes the human intellectual progress represented by the transition from original to the manifest to the scientific image as the progressive “depersonalization” of nature: first everything is a person; then we alone are persons; in the end nothing is a person.

... [T]he manifest image is the modification of an image in which *all* the objects are capable of *the full range* of personal activity, the modification consisting of a gradual pruning of the implications of saying with respect to what *we* would call an inanimate object, that it *did* something. Thus, in the original image to say of the wind that it blew down one's house would imply that the wind *either* decided to so with an end in view, and might, perhaps, have been persuaded not to do it, *or* that it acted thoughtlessly (either from habit or impulse

In the early stages of the development of the manifest image, the wind was no longer conceived as acting deliberately [...] Nature became the locus of ‘truncated persons’ [...] Inanimate things no longer ‘did’ things

in the sense in which persons do them. (Sellars, 1963a, pp. 12–13)

That narrative makes a certain kind of sense. But it invites a prior question. And the most obvious answer to that question raises a further question regarding the cogency of the Sellarsian framework itself. Why, in the first place, do those who take up with the world through the original image personalize the entire world? What is the motivation? And whence do the categories of intentionality and normativity that must be in play in treating *anything*, let alone virtually *everything* as a person, come?

We can reframe this question if we take an ontogenetic view of the transition between the images. We might think of the original image as that of the very young child who sees not only persons, but also inanimate objects such as dolls or other toys, as well as animals, as persons, attributing to them intentionality, mental states and processes, and even moral properties. Maturation into a reflective person leads us to restrict these categories to our conspecifics, and indeed only our mature, reasonably healthy conspecifics; overgeneralization to the severely disabled, the impaired, the infantile or the senescent, let alone to non-human animals or machines, is seen as a kind of immaturity. When we become reflective adults, we turn to science as the measure of reality, allowing it a kind of ontological and epistemological primacy in certain domains; though, as I argue above, we never allow it to displace the manifest image in the way that we do expect the manifest image to displace the original developmentally.

But now we can raise the question posed a moment ago in a new register. If the ontogenetically original image is one of excessive personalization, and the restriction of personalization to other reasonably intact *Homo sapiens* is a later development, whence come these over-applied normative and intentional categories? And here we come upon a dilemma, one with an obvious analogue in the sociogenetic register employed in *Philosophy and the Scientific Image of Man*. They are either innate or of social origin. If the former, we seem to give up on the Sellarsian picture of the categories of intentionality as emerging from collectively constituted norms, and of the theoretical model of the introduction of the concepts of inner episodes made famous in the “Myth of Jones”. If the latter, it seems impossible to understand how the original image antedates in development the manifest. With an eye on this conundrum, we will turn to recent results in developmental and cognitive psychology to better understand the role of the original image.

3. Empirical Psychology and the Philosophy of Mind

Sellars himself notes the *prima facie* difficulty faced by any account of the emergence of conceptual thought, and hence of the emergence of the capacity to attribute intentionality:

The conclusion is difficult to avoid that the transition from pre-conceptual patterns of behaviour to conceptual thinking was a holistic one, a jump to a level of awareness which is irreducibly new, a jump which was the coming into being of man. (Sellars, 1963a, p. 6)

Here Sellars is focusing on the constitution of the manifest image, where, he says, «man first encounter himself as man». But we should note that on his own view, this “transition” must occur much earlier if the original image is indeed to be an image in which the categories of personhood are at work. This only sharpens the problem. Things get more problematic, though, when we focus on the crucial difference between the manifest and scientific images themselves, in the context of the account of the theoretical introduction of the concepts of inner episodes such as thoughts and impressions presented in *Empiricism and the Philosophy of Mind*. In that discussion – surely the most influential and oft-cited fragment of the vast Sellarsian corpus – Sellars urges that we think of these concepts as introduced as theoretical entities – as unobserved *explanans* of intelligent behaviour. But the categories of thoughts and impressions are surely part and parcel of our image of ourselves as persons, and so of the manifest image. However, even if they attain a “reporting role”, if the myth is to have any force, the semantic properties of overt states are conceptually prior to those of inner episodes, and this appears to be inconsistent with the preclusion of theoretical entities from that image:

[T]he conceptual framework which I am here calling the manifest image is, in an appropriate sense, itself a scientific image. It is not only disciplined and critical; it also makes use of those aspects of scientific method which might be lumped together under the heading ‘correlational induction’. There is, however, one type of scientific reasoning which it, by stipulation, does *not* include, namely that which involves the postulation of imperceptible entities, and principles pertaining to them, to explain the behaviour of perceptible things. (Sellars, 1963a, p. 7)

So, if we take the Myth of Jones seriously – and surely that is central to the Sellarsian framework – how are we to make sense of the manifest or of the original images as images whose primary conceptual framework is that of *persons*, but as devoid of theoretical entities? This set of problems must be solved if we are to make any real progress in the project of naturalizing epistemology in the Sellarsian framework. For that involves telling a story within the scientific image of how normativity and the categories of intentionality that define the epistemic domain emerge within the original and manifest images. To solve these problems, it is appropriate to turn to the science of psychology, and in particular to the literature on the acquisition of “Theory of Mind” (the capacity to attribute cognitive states to others and to predict and explain their behaviour on the basis of these attributions) and on the propensity to attribute intentional states to objects in the environment.

The literature on the development of Theory of Mind is vast, and it is well beyond the scope of this paper to survey it.¹ For present purposes, we can distinguish three principal phases in this literature. Initially (the 1980’s and 1990’s) a consensus developed grounded in an impressive array of studies (the classics are Wimmer and Perner, 1983 and Perner, Leekam and Wimmer, 1987, but there are hundreds of kindred results) that prior to the fourth year, young children were unable to make use of belief-attribution in predicting and explaining behavior, in virtue of their regular failure in such false belief tasks as the unseen displacement and misleading container task and their ilk.

Many psychologists early on attributed this to the maturation of a Theory of Mind module responsible for attributing inner states and reasoning about them. The regular developmental track for this range of abilities, its stability across cultures, and the fact that there seems to be in the autism spectrum a selective impairment of this capacity lent credibility to the hypothesis that an innately determined cognitive module is at work (Baron-Cohen, 1995; Baron-Cohen, Leslie, and Frith, 1985; Carruthers, 2006; Leslie, 1987; Scholl and Leslie, 1999). Arguments from evolutionary psychology led added support to the modularist hypothesis. After all, if we consider the obvious selective advantage to individuals able to tell what others are thinking, it is clear that there would be selection pressure for a module that would subservise such a function. And indeed we see in our closest biological kin – the other great apes

¹ For an excellent overview discussion and assessment of the history and significance of that literature, see Fenici (2011).

– precursor abilities, such as the ability to monitor attention and an understanding of the relation between perception, motivation and action in simple situations.

In the next decade or so, a second phase of theoretical thought developed, spurred by the work of de Villiers and de Villiers (2000) on the connection between language development and the development of Theory of Mind.² A substantial body of literature involving both developmental studies and important comparisons of the performance of language-impaired and non-impaired populations on Theory of Mind tasks established a powerful case for the claim that the ability to attribute and to reason about mental states is strongly dependent upon linguistic development, and specifically upon the mastery of the syntax and semantics of tensed sentential complement clauses (Hale and Tager-Flusberg, 2003; Peterson and Siegal, 2000; Tager-Flusberg and Joseph, 2005; Pyers and Senghas, 2009). While more recent work has cast doubt on the tight link between complement mastery and Theory of Mind, and some researchers have suggested a more general connection between capacities such as *irrealis* linguistic representation (Astington and Baird, 2005; Clark, 1998; Fenici, 2011; Garfield, Peterson, and Perry, 2001; Perner, Sprung, Zauner, and Haider, 2003) or narrative competence (Bruner, 1991; Fenici, 2011; Hutto, 2007, 2008, 2009; Nelson, 2009), suggesting more Vygotskian models of acquisition (Garfield et al., 2001; Fenici, 2011; Fernyhough, 2008; Harris, 2005) the role of language in passing classic Theory of Mind tasks appeared unshakeable.

Over the past few years, however, a third wave of theory has washed over the Theory of Mind literature, inspired by the work of Baillargeon (Onishi and Baillargeon, 2005) and her colleagues as well as by Carpenter, Nagell, and Tomasello (1998). A series of very impressive studies have shown fairly conclusively that pre-linguistic children as young as 12 months old, at least implicitly attribute both true and false beliefs to others, understand the relationship between perception and belief formation, act on those attributions (Caron, 2009; Clements and Perner, 1994; Kuhlmeier and Bloom, 2003; Mitchell and LaCohée, 1991; Warrenken and Tomasello, 2006, 2007) and evince surprise when the behavior of others fails to conform to reasonable belief attributions. (Southgate, Senju, and Csibra, 2007; Surian, Caldi, and Sperber, 2007) This literature substantially undermines the thesis that there is

² See also de Villiers (2009) and de Villiers and Pyers (2002).

a deep connection between language and Theory of Mind. These new results also suggest that there is a significant innate component to human attributions of intentionality.³

Nonetheless, these results do not refute the claim that fully mature Theory of Mind is strongly language dependent, and this language dependence is important. Children require mastery of *irrealis* linguistic constructions, of which sentential complements in English are one class (though infinitival constructions seem equally potent (Cheung et al., 2004, Perner et al., 2003), in order to explain behavior induced by false belief, to predict behaviour in the context of over-riding reality expectations or emotional valence, and to execute non-spontaneous, deliberative false-belief reasoning, as opposed to the spontaneous reactions evinced by infants in these paradigms.

Before we ask just how all of this is relevant to the task of naturalizing epistemology, reconciling the images and explaining the particular role of the original image in human life, let us consider one more surprising discovery from the cognitive science laboratory. We noted above that on the Sellarsian picture, emergence from the original image involves the depersonalization of the non-human world. Not only infants (Csibra et al., 1999), but even educated adults, when viewing a video display of geometric shapes moving about a screen spontaneously describe their movements and relations to one another in intentional terms (“the triangle is chasing the circles”) and attribute emotional states to them (“the circles are afraid”) (Heider and Simmel, 1944; Michotte, 1946). Moreover, fMRI scans of subjects viewing these displays demonstrate that the areas in the brain associated with spontaneous intentional attributions to persons are active when viewing these displays. (Castelli, Happé, and Frith, 2000). How far have we emerged from the original image?

4. Naturalizing the Normative; Norming Nature

Results such as those of Castelli et al. (2000) suggest that we think about the original image ontogenetically, and not merely as a stage we go through, but about a primordial mode of taking up with the world. As we mature into a life lived in the manifest image, we learn to override that basic disposition to attribute intentionality to whatever moves autonomously or looks roughly animate, in conscious thought, but it never really leaves us. The original image

³ See also Brooks and Meltzoff (2002) and Baron-Cohen (1995) regarding infant gaze monitoring.

is original, contemporary cognitive science suggests, because the viewpoint it encodes is biologically determined. We have simply evolved to attribute intentionality. Those who lack this ability are failures in the competition for resources and mates in human societies. The hard task for the normally developing *Homo sapiens* is not to learn *to* interpret, but to learn *not to*; to make the transition from the original to the manifest image.

To attribute intentionality or belief to others spontaneously is necessary, but not sufficient for full social life. Without that propensity built in to us (Dapretto et al., 2005; Meltzoff, 2005, 2007; Meltzoff and Moore, 1977), it is hard to see how we could ever acquire it. But that propensity only scaffolds and enables, and does not *constitute* full personhood, the personhood that emerges in the manifest image and which includes the ability to attribute full personhood to our fellows. To put that point more precisely, without the innate fundamental propensities to attribute intentionality, and to engage spontaneously with one another on the basis of those attributions – propensities that, as we have seen, never leave us – we could not coalesce into societies, into communities of language users and norm-enforcers, as we would never engage with one another as collaborators in this project. The discovery of this spontaneous capacity and propensity is hence part of the explanation of the ontogenesis of communities.

Communities of attributers permit the constitution of norms and rules governing the use of terms, governing behavior, and governing assertion and justification. And language permits the development of narrative, collaboration in joint ventures, theoretical endeavour, and explicit discourse about belief true and false, desire requited and unrequited, action successful and unsuccessful. It permits explanation, understanding, reflection, and knowledge in the full sense. This in turn makes the collective practice of personhood possible. In these communities, constrained by norms and thereby limited from mere habits, the resultant discourse, investigations and articulation enables us to act as persons in virtue of recognizing ourselves and others as persons – not only as subjects and as objects for one another – but as rational interlocutors both responsible to and responsible for the norms that constitute our collective human life (Fogel, 1993; Lewis et al., 2009). Thus arises the manifest image from the original, perhaps not historically, but ontologically. The original, from the standpoint of the scientific, hence explains the manifest.

And as we noted at the outset, once epistemic norms are in place, science inevitably follows. The ground of its possibility, however, is the constitution of norm-governed epistemic activity by beings biologically like us, in the sense that they are wired for intentional attribution. In this sense as well, even as scientists, we never leave our biological origins behind. We are in the end complex animals, but complex animals capable of reflecting in a way that only emerges from our social matrix, a matrix we are wired to construct, and which permits the transcendence of nature realized in self-understanding that nonetheless can be explained as a natural phenomenon.

To naturalize epistemology is nothing more than to come to understand ourselves well enough as natural objects to be able to explain how organisms like us can come together to constitute social collectives, and then to supplement that understanding with an understanding of how those social collectives can constitute norm-governed practices that enable knowledge. That is the work of psychology and social theory, and we have seen that while that work may not be complete, it is well underway. Naturalizing epistemology in this way allows us to see just why all three images are necessary in order to understand our being-in-the-world.

But this Sellarsian naturalization of epistemology also amounts to a norming of the natural world. For in doing so, we come to see ourselves not only as persons, but also as *animals*, animals that have evolved to occupy a particular ecological niche with a particular innate endowment that suits us to live in a particular – and particularly complex – way with one another. To live in that particular, natural way, is to live a norm-governed life; such a norm-governed life is hence not even, as Aristotle or Hume would have it, *second* nature; for us, it is *first* nature. To fall short of that life would be to fall short of what Marx felicitously called our “species-being”.

REFERENCES

- Astington, J.W., & Baird, J.A. (Eds.) (2005). *Why Language Matters for Theory of Mind*. New York: Oxford University Press.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: The MIT Press.
- Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the Autistic Child Have a “Theory of Mind”? *Cognition*, 21(1), 37–46.

- Brooks, R., & Meltzoff, A.N. (2002). The Importance of Eyes: How Infants Interpret Adult Looking Behavior. *Developmental psychology*, *38*(6), 958–966.
- Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry*, *18*(1), 1–21.
- Caron, A.J. (2009). Comprehension of the Representational Mind in Infancy. *Developmental Review*, *29*(2), 69–95.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social Cognition, Joint Attention, and Communicative Competence From 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i–vi, 1–143.
- Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. New York: Oxford University Press.
- Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *Neuroimage*, *12*(3), 314–325.
- Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Wang, S.P., & Mo, L. (2004). Relative Roles of General and Complementation Language in Theory-of-Mind Development: Evidence From Cantonese and English. *Child Development*, *75*(4), 1155–1170.
- Churchland, P. (1986). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
- Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary Themes*. Cambridge: Cambridge University Press, 162–183.
- Clements, W.A., & Perner, J. (1994). Implicit Understanding of Belief. *Cognitive Development*, *9*(4), 377–395.
- Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal Attribution Without Agency Cues: the Perception of ‘Pure Reason’ in infancy. *Cognition*, *72*(3), 237–267.

- Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., & Iacoboni, M. (2005). Understanding Emotions in Others: Mirror Neuron Dysfunction in Children with Autism Spectrum Disorders. *Nature Neuroscience*, 9(1), 28–30.
- de Villiers, J.G. (2009). Complements Enable Representation of the Contents of False Beliefs: the Evolution of a Theory of Theory of Mind. In *Language Acquisition*. Basingstoke: Palgrave Macmillan.
- de Villiers, J.G., & de Villiers, P.A. (2000). Linguistic Determinism and the Understanding of False Beliefs. In P. Mitchell & K.J. Riggs, *Children's Reasoning and the Mind*. Hove, UK: Psychology Press, 191–228.
- de Villiers, J.G., & Pyers, J.E. (2002). Complements to Cognition: a Longitudinal Study of the Relationship between Complex Syntax and False-Belief-Understanding. *Cognitive Development*, 17(1), 1037–1060.
- deVries, W. (2005). *Wilfrid Sellars*. Montreal: McGill-Queen's University Press.
- Fenici, M. (2011). *The False Belief Test for Dummies*. Università degli Studi di Siena, Italy, PhD dissertation in Cognitive Science.
- Fernyhough, C. (2008). Getting Vygotskian about the Theory of Mind: Mediation, Dialogue and the Development of Social Understanding. *Developmental Review* 28(2), 225–262.
- Fogel, A. (1993). *Developing through Relationships: Origins of Communication, Self, and Culture*. Chicago: University of Chicago Press.
- Garfield, J.L. (1988). *Belief in Psychology: A Study in the Ontology of Mind*. Cambridge, MA: The MIT Press.
- Garfield, J.L. (2000). The Meanings of 'Meaning' and 'Meaning': Dimensions of the Sciences of Mind. *Philosophical Psychology*, 13(4), 421–440.
- Garfield, J.L., Peterson, C.C., & Perry, T. (2001). Social Cognition, Language Acquisition and the Development of the Theory of Mind.

- Mind and Language*, 16(5), 494–541.
- Hale, C.M., & Tager-Flusberg, H. (2003). The Influence of Language on Theory of Mind: a Training Study. *Developmental Science*, 6(3), 346–359.
- Harris, P.L. (2005). Conversation, Pretence and Theory of Mind. In J.W. Astington & J.A. Baird (Eds.), *Why Language Matters for Theory of Mind*. New York: Oxford University Press, 70–83.
- Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *American Journal of Psychology* 57(2), 243–259.
- Hutto, D.D. (2007). The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology. *Royal Institute of Philosophy Supplement*, 60, 43–68.
- Hutto, D.D. (2008). *Folk Psychological Narratives*. Cambridge, MA: The MIT Press.
- Hutto, D.D. (Ed.). (2009). Narrative and Folk Psychology. *Journal of Consciousness Studies*, 16(6–8), 9–39.
- Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of Dispositional States By 12-Month-Olds. *Psychological Science*, 14(5), 402–408.
- Leslie, A.M. (1987). Pretense and Representation: The Origins of “Theory of Mind”. *Psychological Review*, 94(4), 412–426.
- Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B., & Huang, Z. (2009). Culture, Executive Function, and Social Understanding. *New Directions for Child and Adolescent Development*, 2009(123), 69–85.
- Meltzoff, A.N. (2005). Imitation and Other Minds: The “Like Me” Hypothesis. In S. Hurley & N. Chater, *Perspectives on Imitation: From Neuroscience to Social Science*. Cambridge, MA: The MIT Press, 55–77.
- Meltzoff, A.N. (2007). “Like me”: a Foundation for Social Cognition. *Developmental Science*, 10(1), 126–134.
- Meltzoff, A.N., & Moore, M.K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198, 75–78.

- Michotte, A. (1946). *La Perception de la Causalité*. Oxford: Etudes Psychol. Vol. VI.
- Mitchell, P., & Lacochee, H. (1991). Children's Early Understanding of False Belief. *Cognition*, 39(2), 107–127.
- Nelson, K. (2009). Narrative and Folk Psychology. *Journal of Consciousness Studies*, 16(6–8), 69–93.
- Onishi, K.H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719), 255–258.
- Perner, J., Leekam, S.R., & Wimmer, H. (1987). Three-Year-Olds' Difficulty With False Belief: the Case for a Conceptual Deficit. *British Journal of Developmental Psychology*, 5(2), 125–137.
- Perner, J., Sprung, M., Zauner, P., & Haider, H. (2003). Want That Is Understood Well before Say That, Think That, and False Belief: A Test of de Villiers's Linguistic Determinism on German-Speaking Children. *Child Development*, 74(1), 179–188.
- Peterson, C.C., & Siegal, M. (2000). Insights into Theory of Mind from Deafness and Autism. *Mind and Language*, 15(1), 123–145.
- Pyers, J.E., & Senghas, A. (2009). Language Promotes False-Belief Understanding: Evidence from Learners of a New Sign Language. *Psychological Science*, 20(7), 805–812.
- Rosenberg, J. (2007). *Wilfrid Sellars: Fusing the Images*. Oxford: Oxford University Press.
- Scholl, B.J., & Leslie, A.M. (1999). Modularity, Development and "Theory of Mind". *Mind & Language*, 14(1), 131–153.
- Schopenhauer, A. (1813/2003). *On The Fourfold Root of the Principle of Sufficient Reason*. Tr. by EFJ Payne. Chicago: Open Court.
- Sellars, W. (1949). Language, Rules and Behavior. In S. Hook (Ed.), *John Dewey: Philosopher of Science and Freedom*. New York: Dial Press, 289–315.
- Sellars, W. (1954). Some Reflections on Language Games. *Philosophy of Science*, 21(3), 204–228.

- Sellars, W. (1963a). Empiricism and the Philosophy of Mind. In W. Sellars, *Science, Perception and Reality*. London: Routledge and Kegan Paul, 127–196.
- Sellars, W. (1963b). Philosophy and the Scientific Image of Man. In W. Sellars, *Science, Perception and Reality*. London: Routledge and Kegan Paul, 1–40.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation Through Attribution of False Belief by 2-Year-Olds. *Psychological Science*, *18*, 587–592.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of Beliefs by 13-Month-Old Infants. *Psychological Science*, *18*(7), 580–586.
- Tager-Flusberg, H., & Joseph, R.M. (2005). How Language Facilitates the Acquisition of False Belief in Children with Autism. In J.W. Astington & J.A. Baird (Eds.), *Why Language Matters for Theory of Mind*. New York: Oxford University Press, 298–318.
- Warneken, F., & Tomasello, M. (2006). Altruistic Helping in Human Infants and Young Chimpanzees. *Science*, *311*(5765), 1301–1303.
- Warneken, F., & Tomasello, M. (2007). Helping and Cooperation at 14 Months of Age. *Infancy*, *11*(3), 271.
- Wimmer, H., & Perner, J. (1983). Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. *Cognition*, *13*(1), 103–128.

Science and the Modest Image of Epistemology

Owen Flanagan^{*}
ojf@duke.edu

Stephen Martin^{*}
sjm30@duke.edu

ABSTRACT

In *Philosophy and the Scientific Image of Man* (1963) Wilfrid Sellars raises a problem for the very possibility of normative epistemology. How can the “scientific image”, which celebrates the causal relation among often imperceptible physical states, make room for justificatory relations among introspectible propositional attitudes? We sketch a naturalistic model of reason and of epistemic decisions that parallels a compatibilist solution to the problem of freedom of action. Not only doesn’t science lead to rejection of our account of normative reasoning, science depends on, sophisticates, and explains how normative reasoning is possible.

The final problem of philosophy is to connect the fact and content of knowledge with its conditions. How does knowing occur in the kind of world that is actually known? Knowing is a fact and must be connected up with the world which the sciences study. Thus a system of philosophy answering this question is the capstone of science.

(Roy Wood Sellars, *Evolutionary Naturalism*, 1922, p. 6)

^{*} Department of Philosophy, Duke University, Durham NC, USA.

1. The Space for Reasons in a World of Causes

Does empiricism, the epistemology of the sciences, remain recognizable as epistemology as the sciences advance? Epistemology provides the standards of *justified true belief*, the norms of *warranted assertability*, the criteria for determining the *quality* of evidence and inference. What is the place of such a discipline in the world described and explained in terms of causal interactions among unfamiliar events and processes, some of them imperceptible? (Sellars, 1963).

The problem can be expressed in a variety of ways, all of which are disquieting from the perspective of one who thinks of philosophy, generally, and epistemology, in particular, as foundational, autonomous, antecedent to, or independent of inquiry; or, to put it another way, who thinks of epistemology in a way that makes its core concepts and categories immune from doubts resulting from the very inquiry it promotes and endorses.¹ How does *truth*, specifically the *truth relation*, where ‘p’ is true iff p, fit with the dominant relation that the “scientific image” celebrates, the *causal relation*? How, if they do, do beliefs, which along with *desires* are a fundamental element of folk psychology, survive into the scientific image? How can the sciences preserve the *knowledge relation*, as the perennial philosophy conceives it, if the processes of acquiring beliefs, evaluating beliefs, and consulting epistemic standards are analyzed reductively in terms of the flow of information, where the flow of information is understood in causal-neural-computational terms?

In *Empiricism and the Philosophy of Mind* (1997) Sellars marks critical thinking, the operation of questioning, evaluating, and deciding to endorse or

¹ Famously Descartes tries to find epistemic *terra firma* for inquiry independent of inquiry. But he doesn’t succeed. Thus “the Cartesian circle”, the problem of the foundationalist not being able to doubt everything and also engage in inquiry. In Descartes’ case, he isn’t able to doubt language, logic, the principle of clarity and distinctness and also to engage in the thought experiment that yields the *cogito*, the security of logic, and the principle of clarity and distinctness. The American pragmatists – specifically Peirce – recommends that we develop standards of “warranted assertability”, is a response to both Descartes’ problem, and to the additional or separate fact that science just never seems to yield certainty. One might say the shift from the *Theaetetus* account of knowledge in terms of justified true belief to a warranted assertability account – the distance between Plato and Peirce – was caused by science, by observing its history. Even the standards of epistemology itself are open to critique, revision, and expansion. Consider the evolution of the straight rule of induction as a tool of science. We start with the rule: if I observe that A & B co-occur to m/n, then infer that the next A will come with B to m/n. But we learn to insert rules in the antecedent such as so infer “if the sample is representative and large enough”. Eventually the straight rule yields, *a la Reichenbach* (1949), the canons of inductive logic, statistics and probability theory.

not endorse our own epistemic states, as the act of «placing [these states] within the logical space of reasons, of justifying and being able to justify what one says» (p. 76). What is this “logical space of reasons”? How, and where, is it, “reason”, situated in the four dimensional manifold of space-time, and how does it operate to produce claims that are of high quality and not merely of a certain quantity? Where in a world of causes is there space for reason?

The “task”, according to Sellars, is one of

[S]howing that categories pertaining to man as a *person* who finds himself confronted by standards (ethical, logical, etc.) which often conflict with his desires and impulses, and to which he may or may not conform, can be reconciled with the idea that man is what science says he is. (1963, p. 38)

In one guise, this problem is to reconcile the freedom in our lives suggested by the ability to decide what to believe with laws of nature that might seem to leave no room for decision or free action. In another guise, it is to account for the appearance of epistemic normativity within the causal, descriptive enterprise of the natural sciences. Failure to resolve this problem suggests that knowledge, as understood by epistemology, is mere appearance, and that our self-image as rational animals is an illusion.

The problem is made even more acute by the fact that the sciences themselves make knowledge claims. It is in this respect that the scientific image seems to be founded on, and thus incapable of replacing, what Sellars calls the “manifest image”, an ideology organized around a suite of concepts required for the operation of reason. The idea is that (a) the sciences are made intelligible as sciences by their ability to describe phenomena and evaluate claims about them, and (b) this ability has a parent in philosophical thinking’s broader freedom to call anything whatsoever into question. For the sciences to question and reject the ability to make up and change our minds about what to believe according to epistemic standards would be for them to undercut both their own intelligibility as a conglomeration of rational research programs, and, as a consequence, any epistemic authority those research programs might appear to have.²

² Hilary Putnam seems to have worries along these lines when he laments Quine’s call for naturalizing epistemology as “mental suicide” (1982, p. 20).

2. Ontological Commitment to the Kinds of Intentional Psychology

The problem of epistemic normativity is to make sense of the position from which we decide what to believe, and to identify the scope and limits of what empirical knowledge can tell us about the quality of those decisions. Sellars describes the problem as one of integrating two related logical frameworks that govern inquiry into the world and our place in it. These frameworks are his *manifest* and *scientific* images.

The manifest image is the system in which persons come to recognize themselves as persons, that is, as conscious, rational, free beings. It is organized around concepts used to both describe and evaluate various aspects of human life. The descriptive concepts are concepts like *perception, sensation, pleasure, pain, belief, desire, intention, goal, fear, hope, love, hate, choice, decision*. They all fall within the classificatory scheme of what is called intentional psychology as it is deployed within the manifest image to describe aspects of human life. The normative concepts are concepts like *good, evil, success, failure, justification, evidence, responsibility, duty, beauty, and wellbeing*. These are applied to assess and evaluate various qualities of the phenomena described by intentional psychology. These are second-order concepts in presupposing the legitimacy of those descriptions.

According to the manifest image, there are many kinds of cognitive entities that we are capable of considering, evaluating, and accepting or rejecting. At one end of the spectrum, there are phenomena like the contents of sensory experience, phenomena that Kant, for example, describes as what we receive, passively, in sensible intuition. At the other end of the spectrum, there are comprehensive ideologies: deeply held convictions, considered opinions, and whole systems of belief. To question these ideologies in deliberation is to consider the possibility that a whole way of understanding the world and our place in it is mistaken. Sellars reminds us in *Philosophy and the Scientific Image of Man* that the manifest image is an ideology like any other. Is there any basis for thinking that it is immune to doubt or rejection as inquiry advances?

If the inquiry is conducted by the sciences, there are indeed grounds for doubt about the survival of manifest image. Let us provisionally assume, with Sellars, that the sciences have final say about what is true. The scientific image departs from the manifest image at just the point where theoretical claims make reference to entities that have no basis in the concepts of familiar experience, among them the concepts of intentional psychology. If the sciences of human

behavior *can* proceed without deploying those concepts, the manifest image will have turned out to be an ultimately dispensable ideology. If the sciences *must* proceed without deploying those concepts, matters will be worse. The manifest image, as the home of both intentional psychology and epistemology, will have turned out to be mistaken, misleading, foolish, or fictional, not worth using if we wish to advance our conception of reality as it is. In hopeful moments Sellars conceives of the situation as one in which the ontology of intentional psychology (and of common sense, generally) is preserved as science provides “a needlepoint of detail” to that image; in other moods, he broaches the eliminativist possibility.³

3. Two Arguments Against the Possibility of Traditional Epistemology

The sciences motivate two lines of argument against the possibility of epistemology. The first of these threatens the legitimacy of intentional psychology and, in particular the concept of *belief*. Paul Feyerabend and Richard Rorty were the first to develop the idea, but the most straightforward version is an argument of Paul Churchland’s (Cf. Churchland, 1981).⁴ Churchland takes seriously Sellars’s idea from *Empiricism and the Philosophy of Man* that intentional psychology is a theory in which thought is conceived by analogy to overt speech acts.⁵ Churchland thinks it is a bad theory, on par with long-discredited Aristotelian physics and creationist biology, and thinks that the concepts of thought it deploys should and will be replaced by concepts from the neurosciences. He writes:

A look at the history of [intentional psychology] does little to allay such fears, once raised. The story is one of retreat, infertility, and decadence. The presumed domain of [intentional psychology] used to be much larger than it is. In primitive cultures, the behavior of most of the elements of nature were understood in intentional terms. The wind could know anger, the moon

³ Sellars, in the sort of hopeful moment we have in mind, says this of philosophy in the Platonic tradition: «Let me elaborate on this theme by introducing another construct which I shall call – borrowing a term with a not unrelated meaning – the perennial philosophy of man-in-the-world. This construct, which is the “ideal type” around which philosophies in what might be called, in a suitably broad sense, the Platonic tradition cluster, is simply the manifest image endorsed as real, and its outline taken to be the large scale map of reality to which science brings a needle-point of detail and an elaborate technique of map-reading» (1963, p. 8).

⁴ See Paul Feyerabend (1963) and Richard Rorty (1965,1970) for the earliest contemporary expressions of eliminative materialism.

⁵ Sellars rehearses this view in *Philosophy and the Scientific Image*.

jealousy, the river generosity, the sea fury, and so forth. These were not metaphors. [...] To use Imre Lakatos' terms, [intentional psychology] is a stagnant or degenerating research program, and has been for millennia. (Churchland, 1981, p. 75)

The argument is that there may be no such things as beliefs, and, moreover, that there is *evidence* that there are no beliefs, as they are conceived by perennial philosophy. And if there are no beliefs, then there are no justified true beliefs, that is nothing to which concepts of justification and truth may be applied. There are no good witches and no wicked witches because there are no witches. One hope is that epistemology in the scientific image might recover something belief-like, something representational, that could play the role that beliefs conceived as discrete mental states play within the manifest image.

The second line of argument targets the justification relation. It is also very straightforward. It goes like so: What makes a science a science are its laws. Laws are counterfactual-supporting generalizations about causal relations. Therefore the only relations there are within the sciences are causal relations. On the assumption that the sciences have final say about what is true, causal relations are therefore the only relations there are; *sui generis* justificatory relations are an illusion.

As straightforward as this argument is expressed it is not as powerful as its proponent thinks. First, quantum physics questions the fundamentality of the causal relation in physics. Even if, as many say, causality is fundamental above the quantum level and even if the sciences above that level understand causal relations to be ontologically, or at least explanatorily, fundamental, there are nevertheless many non-causal relations that the sciences at least recognize, and, in some cases, find indispensable. There are spatial relations, expressed by claims like "Plato is to Aristotle's right", and "Michelangelo is painting upside-down again". There are temporal relations, expressed by claims like "Confucius was born before" Socrates. There are quantitative relations, expressed by claims like "Russell has more whiskey in his glass than Whitehead". There are mereological relations, expressed by claims like "Part of the painting is smudged". There are statistical relations, expressed by claims like "The majority of Leonardo's artwork is unfinished". This is not to mention all of mathematics: systems of relations that are (a) non-causal, and (b) theoretically indispensable to the natural and social sciences.

Despite all the non-causal relations the sciences either recognize or, with mathematics, presuppose, the thought remains that the sciences, as part of a descriptive enterprise, could never assess and prescribe epistemic standards. Jaegwon Kim (1988) gives a well-known version of this idea in objecting to W.V. Quine's suggestion in *Epistemology Naturalized* that «epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science» (Quine, 1969). Kim objects to Quine's apparent recommendation to replace the study of a normative relation, justification, with the study of a causal relation, the one that holds between sensory input and verbal and written output. «[Quine] is asking us to set aside the entire framework of justification-centered epistemology. That is what is new in Quine's proposals. Quine is asking us to put in its place a purely descriptive, causal-nomological science of human cognition» (1988, p. 388). Kim's worry is that if the sciences have final say about what is true, and if all the sciences are authorized to speak about is the way things are, and why things are as they are, then they cannot (also) say anything about what we *ought* to believe, about the *norms* of belief, since *oughts* and *norms* are not the sorts of thing about which science, according to its job description, has, or is entitled to have, opinions about. Thus there can be no epistemology, traditionally conceived, within the scientific image.⁶

4. Eddington's Tables

The surprising twist in the plot of scientific advancement is that the more we come to know about the world, the less we understand it. The less we understand it because the less our commonsense language seems to grasp things as they “really” are. This is why Sellars invokes Arthur Eddington's “two tables” when discussing “the clash of images” (1963, pp. 35–36) in *Philosophy and the Scientific Image of Man*. In the passage Sellars refers to, Eddington worries about the potential conflict between the world of everyday experience and the conclusions of scientific inquiry:

⁶ One can dispute Kim's reading of Quine (Flanagan, 2006). The point of Quine's remark is that the (descriptive) project of rational reconstruction ought to be replaced by the (descriptive) project of neuropsychology, *not* that the *prescriptive* project of traditional epistemology ought to be replaced by the *descriptive* project of correlating sensory input with verbal and textual output. Nevertheless, it remains to be seen what the norms of naturalized epistemology come *to* if they are to come *from* the empirical sciences, given that the empirical sciences seem to be limited to descriptive claims.

I have settled down to the task of writing these lectures and have drawn up my chairs to my two tables. Two tables! Yes; there are duplicates of every object about me – two tables, two chairs, two pens. [...] One of them has been familiar to me from earliest years. It is a commonplace object of that environment which I call the world. How shall I describe it? It has extension; it is comparatively permanent; it is coloured; above all it is substantial. [...] My scientific table is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed; but their combined bulk amounts to less than a billionth of the bulk of the table itself. (1928, p. ix–x)⁷

We start in inquiry, eventually in the sciences, with what is most direct and familiar, only to find our initial understanding undercut by continual theoretical ramification and revision. This is bewildering enough when the subject is physics, but what happens when the sciences turn their attention to understanding and knowledge in their own right? It is one thing for the world at large to become more alien with the development of theory, it is another to form an alien self-conception, a picture in which we no longer understand ourselves to be who we thought we were. As the maturation of physical theory undercuts our familiar conception of the world, it stands to reason and is indeed happening (witness the revolutions in evolution, genetics, and now in neuroscience) that the scientific study of persons will not leave our own familiar self-conception intact, including the very concept of a person itself.

We can sum up so far: Preserving normative epistemology as more than a shadow of its former self demands two things: (1) finding room for the ontological categories of rational thought, or something close enough, within the scientific image; and (2) understanding how “believing” and its suite are more than just causally produced and causally productive informational states, but are actually knowledge-yielding. For our familiar self-conception to survive scientific advancement, the manifest and scientific images must share a common model of reason, what we have been calling critical thinking.

⁷ For a similarly provocative comment on the “knowing” vs. “understanding” divide from a famous physicist, we have Richard Feynman saying in *The Character of Physical Law*:

«On the other hand, I think I can safely say that nobody understands quantum mechanics. So do not take the lecture too seriously, feeling that you really have to understand in terms of some model what I am going to describe, but just relax and enjoy it. I am going to tell you what nature behaves like. If you will simply admit that maybe she does behave like this, you will find her a delightful, entrancing thing. Do not keep saying to yourself, if you can possibly avoid it, “But how can it be like that?” because you will get “down the drain”, into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that» (1967, p. 129).

Following Descartes we can also describe the capacity for reason as the faculty of judgment. Under any name, this capacity essentially involves the application of normative concepts to the psychological states postulated by intentional psychology. It is this that places them, as Sellars says, within the space of reasons. What does this capacity come to? How is it possible for material beings in a material world to execute this capacity (Flanagan, 2007)? Answering such questions involves at least two steps: first, we need to accurately describe the phenomenology of the space of reasons; second, we need to provide a naturalistic account that explains how the phenomenology can be realized by mammals like us.

5. The Phenomenology

John McDowell nicely describes the phenomenology of standing in the space of reasons in this passage:

For that kind of locution to fit, in the sense in which I intend it, the subject would need to be able to step back from the fact that it is inclined in a certain direction by the circumstance. It would need to be able to raise the question whether it *should* be so inclined, and conclude that it should. Acting on the inclination – supposing the verdict of the inquiry that is opened by this stepping back is positive – now takes on a dimension of freedom. (Unpublished, p. 7, Section 4)⁸

McDowell gives us the explanandum. What we need is a model of reason, immanent to the manifest image and portable to the sciences, and then refined perhaps by the sciences, but not overturned or eliminated by them, which explains how this phenomenology is possible. Methodologically, we proceed as Kant did, by starting with an adequate description of the phenomenology from which we can infer hidden cognitive processes. The methodological principle is that a system must be constructed such that it can perform the operations it actually performs. This is why it is so important to specify operational performance accurately. For the Kantian cognitive scientist, the good phenomenology functions to provide precise design specifications.⁹

⁸ See Christine Korsgaard's *The Sources of Normativity* (2004) for a similar account of the phenomenology of the operation of reason.

⁹ Flanagan identifies Kant as laying «both the substantive and methodological foundations for modern cognitive science» (1984, 1991). The transcendental deduction is a method of inferring underlying causes from overt phenomena. (Kant knew Newton's *Principia* very well, and it is likely that he was

6. How Epistemic Freedom is Possible

The next step is to provide a naturalistic account – one endorsed by the scientific image – that realistically models cognitive processes that might support the phenomenology of freedom to decide what to believe. Happily, we can help ourselves here to a solution that has been worked-out on the closely related problem of freedom in the domain of action.

Consider Daniel Dennett’s control-based approach to freedom of the will (1984). Dennett wants to explain a familiar aspect of human life – the apparent freedom we have to choose what to do – in light of information that constrains what that freedom could come to – laws of physics, for example, together with the now plausible, post-Darwin assumption that we are entirely material beings.

Dennett flags several major elements of the phenomenology of free choice as needing explanation: (1) Freedom of choice is experienced as coming in degrees: there are easy and difficult choices, no-brainers and effortful tasks, things I can do and things I can’t do (some of which I wish I could do); there are near goals and distant goals, and so on; (2) We can feel stuck or trapped when we have only bad options (or, as in the theory of cognitive dissonance, when we have multiple equally good, but incompatible, options; (3) We can be overwhelmed by having too many options to consider; (4) Normally, only some options are live options; and (5) Modal language about alternative possibilities, about whether and how things could be, and/or could have been otherwise, seems apt when speaking about choices-to-be-made or that-have-been-made.¹⁰

The first thing to note is the strong similarity between Dennett’s description of the phenomenology of free choice and McDowell’s account of how operating in the space of reasons can seem. The next thing to note is how

inspired by Newton’s own stated method of inferring hidden forces of nature from propositions about phenomena of motion. For more on Newtonian methodology, see Stein, 1990; Smith, 2007.

¹⁰ Dennett analyses systems that possess a certain kind of natural freedom, i.e., freedom that is naturalistically possible, this way: “A system has a degree of freedom when there is an ensemble of possibilities of one kind or another, and which of these possibilities is actual at any time depends on whatever function or switch controls this degree of freedom. Switches (either on/off or multiple-choice) can be linked to each other in series, in parallel, and in arrays that combine both sorts of links. As arrays proliferate, forming larger switching networks, the degrees of freedom multiply dizzyingly, and the issues of control grow complex and non-linear. Any lineage equipped with such an array confronts a problem: What information *ought* to modulate passage through this array of forking paths in multi-dimensional space of possibilities? (2004, p. 162)”

we can learn from Dennett's move to answer the question: how is such freedom possible, if it is, in the natural world, the world described by science?

Dennett goes beyond the phenomenology in seeking to provide a model of freedom in terms of the operation of control over what Dretske (1983) calls the *flow of information*. What is control? Dennett writes:

The root idea of control, which has been elevated into a technically precise concept in cybernetics and automata theory, is (in ordinary terms) that *A controls B* if and only if the relation between *A* and *B* is such that *A* can *drive B* into whichever of *B*'s normal range of states *A wants B* to be in. (1984, p. 52)

The promise of Dennett's model of freedom and practical reason lies not only in its scientific viability, but in its being a model around which we can construct an account of theoretical reason, in which the flow of information is, or at least can be, controlled by consulting epistemic standards, or better, by processing through epistemic norms. Let us sketch out that account, starting with three aspects of McDowell's phenomenological story to focus on as design specifications to be met.

But first we must remind ourselves that critical reasoning, second-guessing, wondering, and changing beliefs may all be relatively rare cognitive processes. Many beliefs, perceptual ones, most familiarly, often just arrive, we go with them, and things work out. Usually when a perceptual experience gives us pause, makes us wonder, not all possibilities are entertained – we wonder whether that was a hawk or an owl, not whether that was some kind of bird, or some kind of kite, or some kind of extra-terrestrial, or some kind of hallucination. But sometimes the world or a text or a fellow inquirer or conversant gives us pause and makes us engage in wandering about the space of reasons. Then the familiar phenomenology ensues.

Here are some features of the phenomenology: First, the act of belief examination or, what is different, justification is experienced as some kind of wondering about or assessment of what one is, or was initially, inclined to believe, an evaluation of what is on the table as a candidate for epistemic endorsement. The cause of the wondering, the second-guessing and double-checking, is that something in the world has given one pause, has put the stops on automatic assent or dissent. Normally, when we are betwixt and between beliefs only some possible beliefs are live options. Considering a set of live options, plausible contenders for endorsement, is in fact, all it normally means to open an inquiry, or, alternatively, to have an open mind. The act of belief

examination or justification is really experienced as a process. The act is really an activity, experienced as an admixture of choate and inchoate, disciplined and undisciplined thoughts, resulting in a decision about what to believe for now.

7. Epistemology Naturalized

The free and critical operation of our rational capacities lie at the center of the manifest image endorsed by traditional philosophical thought (Cf. Flanagan, 2002). The topic of their formulation and legitimacy singularly define perennial philosophy as a series of footnotes to Plato, who was the first to treat them systematically. Consider the psychology of the *Republic*. In response to Adeimantus's request for a constructive defense of justice, Socrates suggests understanding the hidden activity of the soul by analogy to overt city life. He proceeds to identify reason with the rulers of a just *polis*, whose job is to weigh the preferences of the military and working classes against the best interest of the polis as a whole. The soul and the *polis* each operate rationally when alternative courses of action are considered for selection by the ruling class.

Plato does two very interesting things, here. He characterizes the soul in terms of parts that have functional configurations, and he identifies justice as the virtue realized by one configuration in particular. This configuration is the one in which the calculating part of the soul is free to assess the recommendations of the emotional and appetitive parts against its own concern for the whole. Plato's hypothesis is that *practical* reason functions to improving decisions about what to *do*. Our suggestion is that we think of *theoretical* reason, the ability to think critically about what to *believe*, by analogy to Plato's functional conception of practical reason, and if we can, is this a model of reason the sciences can accept?

In both metaethics and meta-epistemology, the issue is whether an ability, a sort of freedom to decide how to act and what to believe, respectively, remains available to us in light of information we have about how the world – including most relevantly, the mind – works. Some say knowledge about how the world works – discovery of the laws of physics and psychology – seems to crowd out deliberate action and thought. One can simply deny that we have any such freedom – what you find with hard determinism in ethics and the pure descriptivism Kim is worried about in epistemology – or make room for it by foisting a compatibilist view.

Our proposal is similar to John Dewey's – a compatibilist in ethics and epistemology – in analyzing the kind of freedom we care about in both ethics and epistemology as the ability to produce differential responses (beliefs and actions among them) to future circumstances based on feedback about past successes and failures.¹¹ To possess this sort of freedom is to be *response-able*, or *responsible* with an 'a'.¹² This is a modest conception of freedom, to be sure, but *if* it licenses an operational model of theoretical reason in the mold of Plato's account of practical reason, and if the sciences can accept a model of practical reason like Dennett's, this conception of freedom further licenses a model of reason that has potential staying-power inside the scientific image.¹³

The fact that the model of reason fundamental to the manifest image *may* be acceptable to the sciences, that it is not *inconsistent* with science, does not entail its usefulness (or indispensability) to them. Does the suite of concepts required for the operation of reason play a theoretical role within the scientific image? There are really two questions here. (1) Do the sciences apply an operational model of reason according to which we reflect upon, evaluate, and decide what to believe, and (2) do the sciences describe this operation as actually resulting in the acquisition of justified true beliefs?

The answer to (1) is yes, though by focusing on the use of intentional psychology in everyday experience – naive intentional psychology – critics have overlooked the extent to which the concepts of belief and desire (or some naturalistic descendants of them) are deployed within the scientific image. Recall Paul Churchland's comment that intentional psychology is a «stagnant or degenerating research program, and has been for millennia» (1981, p. 75). Churchland may be right that intentional psychology, or parts of it, did

¹¹ See Dewey (2002): «For morals has to do with acts still within our control, acts still to be performed. (p. 18) [...] The moral issue concerns the future. It is prospective. [...] The moral problem is that of modifying the factors which now influence future results» (p. 19). Flanagan discusses the sense in which reason is forward-looking in *The Science of the Mind* (1991, p. 51).

¹² See Flanagan's *The Really Hard Problem* (2007): «Dewey says the moral problem concerns the future. I treat you as an intelligent being, capable of self-control, if I call you on inappropriate or non-virtuous actions. If you are receptive and paying attention, my response gives you reason to behave better in the future. I call this *responsibility* to indicate that it incorporates the credible assumption that our characters, our hearts and minds, are plastic to some degree. Social communities are dynamic systems in which complex feedback mechanisms help us adjust our beliefs, desires, feelings, emotions, and behavior» (p. 35).

¹³ Compare with Paul Churchland's (2012) view that reasoning is pattern and meta-pattern recognition. For a dissenting view, Alex Rosenberg (2011) continues to hold the line of eliminative materialism.

stagnate for millennia, but this stagnation abruptly ended with Pascal's inaugural treatment of probabilistic judgment in the 17th century, which continues to lead to advances in formal epistemology, broadly construed. Another research program that preserves – by bringing “needlepoint of detail” to – the concepts, categories, and processes that the manifest image endorses, is cognitive science and neuropsychology, a field that Churchland originally thought would eliminate the manifest image, but, in fact, has only sophisticated it.

Consider the field of formal epistemology. Pascal wanted to know how a person ought to choose from an array of uncertain options. He thought that two variables needed to be known to answer this question: the magnitude of an option's reward, and the probability of its being acquired. Pascal then identified the best option, the option that ought to be selected, as the product of those two variables. Importantly, he conceived of the objects of choice as objective quantities, in that the values of alternatives were not taken to be relative to a person's beliefs or desires. In fact, beliefs for Pascal only entered into the picture as potential *objects* of choice, which themselves could be evaluated as the product of their probability of being true and the reward a person would receive for holding them. Famously, Pascal claimed that choosing to believe in God is a better bet than not.

The major problem with Pascal's idea was that it did not adequately describe the decisions that people actually make. That is, even if events have objective probabilities of occurring, and even if those events have objective reward magnitudes, people systematically fail to choose options with the highest objective expected reward. In response to this *descriptive* failure, Daniel Bernoulli (1954) subjectivized Pascal's choice values by relativizing them to a person's expected (believed) utility (desirability). Bernoulli thus imported the notions of belief and desire from naive intentional psychology, quantified them as Pascal had quantified objective probabilities and reward values, and identified their product as the object of human decision making, what has become known as *subjectively expected utility*.

After Bernoulli, Pareto (1927/1971), von Neumann and Morgenstern (1944), and Samuelson (1937, 1938, 1983) successively developed and improved the concept of subjectively expected utility. Despite continual refinement, it remains a concept both (a) born and with roots in, naive belief-desire psychology, and (b) the central notion in economic theory, in social choice theory, and many parts of psychology.

More recently, Richard Jeffrey (1990) has proved that the decision matrices of decision theory, which were traditionally populated by goods and actions, can be interpreted to be populated by propositions or states of affairs, opening the door to a conception of decisions about what to believe as decisions between what we have called possible beliefs.¹⁴ And in the last several years neuroscientists like Michael Shadlen and William Newsome (Cf. Shadlen and Newsome, 1996), and Paul Glimcher (Cf. Glimcher, 2010; Platt and Glimcher, 1999) have identified likely neural correlates of both subjectively expected utility values and the processes of choosing among options on the basis of those values.¹⁵

The point of this rush through the history of decision theory is that by developing a fully quantitative, scientifically respectable notion of rationality in terms of preference consistency, empirical theories of decision-making have been providing “needlepoint of detail” to the roughed-out design of intentional psychology handed off to the sciences by perennial philosophy. And so the proper reply to eliminative materialism is that intentional psychology survived into the scientific image as microeconomics, statistics and probability theory, game theory, and, lately, neuroeconomics, and neuropsychology, and that these disciplines remain secure within the scientific image under the umbrella of the decision sciences

What about (2), the question about whether the sciences describe the operation of reason as one that actually results in justified true beliefs, warranted beliefs, or something in their vicinity.? The answer to this question is complicated. The first complication is that there may be circumstances in which having true beliefs or acquiring information is a bad thing for you, if, that is, your primary interest is personal happiness. Under these circumstances, you are better off being under a *positive illusion* (Cf. Taylor and Brown, 1988; Flanagan, 2007; McKay and Dennett, 2009; Flanagan, 2009). To be better off under a positive illusions is to be better off maintaining an uncritical stance toward your own epistemic states, the risk in questioning them being the disruption of a way of life that (a) makes you happy, and (b) is unrecoverable once disrupted.¹⁶

¹⁴ This is what attracts Donald Davidson to Jeffrey’s view in his own attempts to naturalize reason. See Davidson (1995) and (2004).

¹⁵ Glimcher’s (2010) account of the history of economic thought and its recent integration with the neurosciences is especially illuminating.

¹⁶ Flanagan (1991, 2007, 2009) argues that some positive illusions are not beliefs and thus are not illusions. These are better analyzed as hopes or wishes, possibly unrealistic hopes or wishes.

The second complication is that there may be circumstances in which more information is a good thing, but deliberating about what to believe is unlikely to deliver that information. Since the 1970s, we have become more familiar with the role in decision making of unconscious cognitive heuristics. Indeed many choices do seem to be better made without withdrawing from our circumstances to make them. Gerd Gigerenzer and his group, for instance have advanced a version of Herbert Simon's (1955) idea that the computational complexity of real-time decision making means that we are *in general* better off making choices about what to believe using "fast and frugal heuristics" (Cf. Gigerenzer and Todd, 1999).

There are nevertheless pitfalls to a systematic refusal to think about what to believe. Daniel Kahneman gives a favorite example from the work of his colleague Shane Frederick. Students at Princeton and the University of Michigan were given the following problem:

"A bat and a ball cost \$1.10 in total. The bat costs \$1 more than the ball. How much does the ball cost?" (2003, p. 1450)

The answer subjects are inclined to give, "10 cents", is wrong; and of course upon reflection we see quickly that the correct answer is "5 cents". But why the slip? Kahneman rehearses this study in his *Thinking, Fast and Slow* as an illustration of the function of our capacity for deliberation, what he calls System 2 cognition: «One of the main functions of System 2 is to monitor and control thoughts and actions "suggested" by System 1, allowing some to be expressed directly in behavior and suppressing or modifying others» (2011, p. 43) Kahneman argues that in absence of a prompt to double-check your own epistemic states, something signaling the need to scrutinize them, there is nothing to cause our brains to engage the metabolically costly routine of deliberating about what to believe.

The point can be made metaphorically as well. Why is it so hard to throw a Wiffle ball with any great velocity? The obvious answer is that a Wiffle ball is not heavy enough to throw very fast, and the obvious answer is correct. But it is not correct for obvious reasons. One may suppose that the lightness of the ball makes it more susceptible to drag forces, as a feather's lightness prevents it from falling to the earth with the acceleration of a bowling ball, and perhaps there is some truth to this idea. The real problem, however, does not have to do with drag forces. The problem is that Wiffle balls are not heavy enough to stimulate the full recruitment of the muscle fibers needed to generate the

power required to throw them with great velocity.¹⁷ And just as some measure of resistance is required to recruit our most powerful skeletal muscles, so too some measure of resistance is required to recruit our most powerful mental muscles. Resistance, in the latter case, takes the form of feedback that leaves us not knowing what to believe.

Scientific advancement gives us feedback that leaves us not knowing what to believe about our own capacity for critical thought. This is far more significant than feedback that our first impressions are mistaken, or that more muscle fibers must be innervated to accelerate an object through space. But by characterizing this capacity operationally, in a way that runs back to Plato's understanding of critical thinking, we are coming to understand how a model of reason endorsed by perennial philosophy survives into the scientific image *as* a model of reason. If reason is the capacity to evaluate alternatives in light of the need for further information, and if reasoning about what to believe results in the acquisition of information, then critical thinking remains the sort of faculty that is good to have in situations where revising one's belief or acquiring more information in order to check one's beliefs is to one's advantage.

Our claim is that the sciences are equipped to make sense of reason, of our freedom to think critically and make up our minds about what to believe, if we extend Dennett's control-based account to cover decisions about what to believe. To do this we must (a) define the space of reasons as a matrix of possible beliefs;¹⁸ (b) define critical thinking as the consultation of (potentially implicit) epistemic standards by which we evaluate the evidence for contender

¹⁷ The example comes from Mark Rippetoe's *Starting Strength*. «This is due to several factors involving the physiology of skeletal muscle contraction, among them the fact that a very high velocity movement does not allow enough time for the nerves to recruit many of the components that contribute to muscle contraction. Like trying to throw a wiffleball [sic], a very light weight moving very fast does not provide enough resistance to push against effectively. A baseball is pretty good to throw, because it's just about the right weight to throw hard and fast. Power is at a maximum when throwing a 16 lb. shot, due to the combination of weight and velocity. But a great big rock would be too heavy to allow for the production of much power, because of the very slow velocity even a very strong man could produce. So the load must be optimum for power production» (2007, p. 176).

¹⁸ Because we are concerned here with the psychological side of reasoning, we characterize "reasons" first internally, as psychologically real particulars, along Davidsonian lines. If I am wondering about the best way to get from Durham to Chapel Hill, the routes, whatever they are, that I am now considering are my reasons, the only ones I have or know about. But Google (formerly God in such thought experiments) has information about the very best route. It is an objective fact that I have reason to consult Google Maps. But if I don't know about Google Maps, this is an external reason, one that exists, but that I don't have (internally) as a reason.

beliefs; and (c) explain epistemology as the descriptive-genealogical explication and normative critique of epistemic standards we actually consult. Epistemic standards are the norms, rules, and principles governing the process of deciding what to believe, or, more directly in the province of epistemology itself, what epistemic norms and decision procedures to advocate and utilize. The best explanation of epistemic norms and decision procedures is that they are cognitive strategies developed over the course of history, through personal and social experiences, to guide and improve decisions about what to believe.

8. Epistemic Normativity

We've claimed that overcoming worries about the *bona fides* of epistemology in a world seen from the perspective of the scientific image requires two kinds of work. First, we need to provide a naturalistic model that plausibly describes and explains how the familiar phenomenology of critical reasoning is possible, indeed, why it is as it is. We've shown that a compatibilist model in epistemology, where freedom of thought and belief, freedom of reasoning, is modeled on a compatibilist solution that responds to parallel eliminativist worries about free action, can satisfy this demand. This is the descriptive-genealogical part of the project. It paves the way to address the second part of the problem, the problem of normativity. This is the task of showing epistemology to be capable of discovering, expanding, critiquing, and endorsing the decision procedures that the sciences both utilize and describe and explain.

The epistemic normativity problem seems especially difficult. Why? Because neither the ability to *describe* aspects of human life *as* thought, nor the ability to describe thought as *deliberate* or *critical*, seems to license the ability to *evaluate* the standards employed in deliberation and critical thinking. This is Kim's worry about what the program of naturalized epistemology comes to, an apparent recapitulation of the ethical *is-ought* problem, which Sellars notices and mentions *as* a recapitulation of the ethical *is-ought* problem at the end of *Philosophy and the Scientific Image of Man* (1963, p. 39). The question, then, is how to get claims about the epistemic standards a person ought to consult from the sciences (or from philosophy conceived naturalistically, as continuous with science or as beholding to the scientific image) if all the sciences can do is set out descriptive-genealogical-explanatory claims. This worry does not depend on, and thus does not require therapy to overcome, the false belief that

the causal relation is the only legitimate relation. It depends, at this point in the dialectic, on the beliefs that science only traffics in description, genealogy, and explanation and that these do not yield oughts. Both beliefs are false.¹⁹

To see this, consider Quine's own response to the objection that in *Epistemology Naturalized* he had set out to abandon epistemic normativity:

Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me, normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in more cautiously epistemological term, prediction. Like any technology, it makes free use of whatever scientific findings may suit its purpose. It draws upon mathematics in computing standard deviation and probable error and in scouting the gambler's fallacy. It draws upon experimental psychology in exposing perceptual illusions, and upon cognitive psychology in scouting wishful thinking. It draws upon neurology and physics, in a general way, in discounting testimony from occult or parapsychological sources. There is no question here of ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction. The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed. (1986, pp. 664–665)

The key idea is this: If we stipulate that “truth” is the end, “the terminal parameter”, either given by our platonic, with a small ‘p’, nature, or by the set of social practices we call “science”, then epistemology is the practice of discovering, developing, critiquing, and endorsing the norms that produce it, “the terminal parameter”, truth.

Normative epistemology survives into the scientific image as a body of informed norms that prescribe epistemic decision procedures that reliably produce truth (or, at least, have done so thus far).²⁰ Our decisions about what to believe are good decisions if they meet criteria for how we ought to make up

¹⁹ The key is to deny the antecedent of the conditional in the claim that: “if all the sciences can do is set out descriptive-genealogical-explanatory claims”, then no normativity can emerge. Science can extract norms, not demonstratively or deductively, but abductively, for its own practices that yield knowledge. When the scientific community does this, it goes 2nd order and does what we call “meta-science”. When statisticians, decision scientists, logicians, and philosophers engage in the same activity it is broadly the science of reasoning or, for simplicity, epistemology. The norms are both used to make first order claims and extracted from sets of such claims. Vicious circularity is avoided.

²⁰ See Flanagan (1982, 1988, 2006), where naturalized epistemology is set out as «the enterprise of sorting reliable techniques of knowledge acquisition from unreliable ones in theoretical domains» (p. 541). See also Goldman (1986), Nozick (1994), Lauden (1990), Kitcher (1992), Rosenberg (1990), and Kornblith (1993).

and change our minds, and the criteria for how we ought to make up and change our minds depend on the canons of epistemology conceived along the lines just articulated. Reichenbach's (1949) pragmatic justification of induction (see footnote 1), like the research program initiated by Pascal, is an example of epistemology in just this spirit. The norms of knowing have been refined in epistemology proper, as well as in the epistemic sciences that include statistics and probability theory, game theory, neuroeconomics, and everything else under the umbrella of the decision sciences. One ought not think that naturalized epistemology is in its early stages. It is not. Indeed, the best way to allay fears about its *prospects* is to come to appreciate its *history* as involving a large and precise body of norms that yield knowledge in the sciences, and in ordinary life.²¹

9. Conclusion: Language and the Space of Reasons

So, the scientific image endorses and sophisticates, rather than undermines or eliminates, the manifest image of persons as creatures who operate inside the space of reasons. What does the scientific image say about the relationship between epistemic agency and language? Are humans the only epistemic agents, the only creatures who traffic in reasons? What about nonhuman animals, infrahumans (e.g., infants), and extrahumans – the World Wide Web, the iCloud? From an empirical point of view, what degree of epistemic agency is language necessary for, exactly?

²¹ Here's a thought for unification of ethics and epistemology that might be attractive to the naturalist – defense would require another paper. Insofar as epistemology and ethics involve imperatives, they are all hypothetical imperatives. The "insofar" is important. The ends of epistemology and ethics, might be categorical ends or goods, "truth" for epistemology, "good" for ethics (Flanagan, 2007). The idea that epistemology survives into the scientific image on the back of an instrumental conception of epistemic normativity might be thought to echo Philippa Foot's thesis that morality is a system of hypothetical imperatives (1972). Early Foot thought that moral norms are like other norms in being means to ends. She came later to explain the apparent necessity of moral norms as rooted in the *de facto* ubiquity and necessity (i.e., mandatory nature) of some ends (see *Natural Goodness* 2001). If we put both strands together, something Foot did not think could be done smoothly, ethics is naturalized by conceiving of it as an inquiry into the sorts of things we ought to do if we want to live a rewarding life as a *eudaimon*, where the exact causes and constituents of *eudaimonia* are open to empirical inquiry. Likewise, epistemology naturalized can be understood as the set of instrumental norms that serve the unconditional end of knowing. All the norms are hypothetical, where the antecedent sets the task as truth-seeking. The end of truth-seeking, the command, if you will, to seek truth is itself not conditional, hypothetical, or instrumental. The truth is good in itself.

There is a line of thought in Sellars's work, one taken up by Brandom (cf. 1995), and McDowell (cf. 1994; Unpublished), according to which language is necessary for possession of the suite of concepts required for the operation of reason and critical thought. Sellars, Brandom, and McDowell all acknowledge a continuity between non-linguistic animals and human beings but deny that non-linguistic animals can exhibit legitimately critical thinking. They cannot exhibit legitimately critical thinking because, *ex hypothesi*, they cannot, without language, bring into view for evaluation the very standards they use to evaluate their epistemic states and decide what to believe. The claim, then, would be that language, while not necessary for knowledge of the world in some attenuated, "as-if" sense of knowledge, is necessary for the evaluation and adjustment of the norms according to which such knowledge, or proto-knowledge, is acquired. But this claim is more than the claim that language is necessary for being a participant in the space of reasons; it is a claim about engagement in the special critical reasoning practice of epistemology.

Let us distinguish three phenomena: critical thinking or belief acquisition and revision; epistemology, or critical thinking about, or belief formation and revision with respect to, the standards by which beliefs are acquired and revised; and meta-epistemology, or critical thinking about, or belief formation and revision with respect to, the scope and limits of epistemology. Meta-epistemology involves scrutiny of deeply held convictions, considered opinions, and whole systems of belief: what we have previously identified as comprehensive ideologies.

Critical thinking, as we have described it, is a widespread phenomenon. Furthermore, only critical thinking is necessary for being an epistemic agent. Non-human primates and infants, when provided with environmental feedback in the form of cues and prompts for greater scrutiny, what we earlier called "resistance", are capable of checking, second-guessing, even double-checking their environments to acquire more information relevant to a surmise and in order to reach secure conviction. Dogs sometimes come to make sure a toy has been thrown before darting off for it. Any animal that lives in an environment with variable opportunities for food and sex must learn from experience about these opportunities in order to increase its likelihood of surviving, and thus of passing its genes to successive generations. Learning by checking surmises, by bringing epistemic states into view for endorsement, is sufficient to place such surmises, expectations, and other epistemic states within the space of reasons.

Non-linguistic animals cannot do epistemology as it is practiced in philosophy departments, but can they do anything beyond merely double-checking and second-guessing when confronted with reasons for greater scrutiny? We offer that any animal that has the ability to improve its learning strategies through experience, for example by becoming increasingly wary of small-sample sizes while foraging and thus behaving as though has adopted a new and improved version of Reichenbach's straight rule that has adjusted for sample size is doing something more complex than ordinary critical thinking. Darwinian gradualism encourages us to call such normative refinement among smart non-human animals, "proto-epistemology".

What language is undoubtedly necessary for is meta-epistemology, the central problem of which Roy Wood Sellars identifies as one of connecting knowledge up with the world, which the sciences study. This problem becomes urgent in *Philosophy and the Scientific Image*, where Wilfrid Sellars brings into relief the possibility that the very categories upon which the practice of epistemology depends may not themselves have a place within the scientific image. We have argued that even as the categories of intentional psychology and normative discourse are quantified and formally refined, they are not eliminated. The categories of true belief, justification, and warranted assertability, categories required to make sense of the sciences as sciences, have in fact survived into a multifaceted program of naturalized epistemology that is well underway.

REFERENCES

- Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, 22(1), 23–36.
- Brandom, R. (1995). Knowledge and the social articulation of the space of reasons. *Philosophy and Phenomenological Research*, 55(4), 895–908.
- Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2), 67–90.
- Churchland, P.M. (2012). *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. Cambridge, MA: MIT Press.

- Davidson, D. (1995). Could there be a science of rationality?. *International Journal of Philosophical Studies*, 3(1), 1–16.
- Davidson, D. (2004). A unified theory of thought, meaning, and action. In D. Davidson, *Problems of Rationality*. Oxford: Oxford University Press.
- Dennett, D.C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.
- Dennett, D.C. (1996). *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon and Schuster.
- Dennett, D.C. (1997). *Kinds of minds: Toward an understanding of consciousness*. New York: Basic Books.
- Dennett, D.C. (2004). *Freedom evolves*. ePenguin.
- Dewey, J. (1921/2002). *Human nature and conduct*. New York: Dover Publications.
- Dretske, F. (1983). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Eddington, A.S. (1928). *The nature of the physical world*. New York: MacMillan Company.
- Feyerabend, P. (1963). Materialism and the mind-body problem. *The Review of Metaphysics*, 17(1), 49–66.
- Feynman, R.P. (1967). *The character of physical law*, Cambridge, MA: MIT press.
- Flanagan, O. (1988). Pragmatism, ethics, and correspondence truth: response to gibson and quine. *Ethics*, 98(3), 541–549.
- Flanagan, O. (2006). Varieties of naturalism. In P. Clayton (Ed.), *Oxford handbook of religion and science*. New York: Oxford University Press, 430–452.
- Flanagan, O.J. (1982). Quinean ethics. *Ethics*, 93(1), 56–74.
- Flanagan, O.J. (1984–1991). *The science of the mind*. Cambridge, MA: MIT Press.

- Flanagan, O.J. (2002). *The problem of the soul: Two visions of mind and how to reconcile them*. New York: Basic Books.
- Flanagan, O. J. (2007). *The really hard problem: Meaning in a material world*. Cambridge, MA: MIT Press.
- Flanagan, O.J. (2009). “Can do” attitudes: Some positive illusions are not misbeliefs. *Behavioral and Brain Sciences*, 32(06), 519–520.
- Foot, P. (1972). Morality as a system of hypothetical imperatives. *The Philosophical Review*, 81(3), 305–316.
- Foot, P. (2001). *Natural Goodness*. Oxford: Oxford University Press.
- Gigerenzer, G., & Todd, P.M. (1999). Fast and frugal heuristics: the adaptive toolbox. In G. Gigerenzer, P.M. Todd & the ABC Research Group, *Simple heuristics that make us smart*. New York: Oxford University Press, 3–34.
- Glimcher, P.W. (2010). *Foundations of neuroeconomic analysis*. Oxford: Oxford Univ Press.
- Goldman, A.I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard Univ Press.
- Jeffrey, R.C. (1990). *The logic of decision*. Chicago: University of Chicago Press.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449–1475.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar Straus & Giroux.
- Kim, J. (1988). What is “naturalized epistemology”? *Philosophical perspectives*, 2, 381–405.
- Kitcher, P. (1992). The naturalists return. *The Philosophical Review*, 101(1), 53–114.
- Kornblith, H. (1993). Epistemic normativity. *Synthese*, 94(3), 357–376.
- Korsgaard, C.M. (2004). *The sources of normativity*. Cambridge: Cambridge University Press.

- Laudan, L. (1990). Normative naturalism. *Philosophy of Science*, 57, 44–59.
- McDowell, J. (1994) *Mind and World*. Harvard University Press
- McDowell, J. (Unpublished). *Sellars and the space of reasons*.. Retrieved from <http://www.pitt.edu/~brandom/mc-core/downloads/McD%20Cape%20Town%20talk--Sellars%20EPM.doc>
- McKay, R.T., & Dennett, D.C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32(6), 493–561.
- Nozick, R. (1994). *The nature of rationality*. Princeton: Princeton University Press.
- Pareto, V. (1927/1971). *Manual of political economy*. Tr. by A.S. Schwier. New York: Augustus M. Kelly.
- Platt, M. L., & Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741), 233–238.
- Putnam, H. (1982). Why reason can't be naturalized. *Synthese*, 52(1), 3–23.
- Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Quine, W.V.O. (1969). Epistemology Naturalized. In W.V.O. Quine, *Ontological Relativity and Other Essays*. New York: Columbia University Press, 69–90.
- Quine, W.V.O. (1986). Reply to Morton White. In L.E. Hahn & P.A. Schilpp (Eds.), *The Philosophy of WV Quine*. The Library of Living Philosophers, vol. XVIII. La Salle, IL: Open Court Publishing Company, 663–665.
- Reichenbach, H. (1949). *The Theory of Probability*. Berkeley and Los Angeles: University of California Press.
- Rorty, R. (1965). Mind-body identity, privacy, and categories. *The Review of Metaphysics*, 19, 24–54.
- Rorty, R. (1970). In defense of eliminative materialism. *The review of metaphysics*, 24, 112–121.
- Rosenberg, A. (1990). Normative naturalism and the role of philosophy. *Philosophy of Science*, 57, 34–43.

- Rosenberg, A. (2011). *The Atheist's Guide to Reality: Enjoying Life Without Illusions*. New York and London: WW Norton & Company.
- Samuelson, P.A. (1937). A note on measurement of utility. *The Review of Economic Studies*, 4(2), 155–161.
- Samuelson, P.A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 5(17), 61–71.
- Samuelson, P.A. (1983). *Foundations of economic analysis*. Cambridge, MA: Harvard University.
- Sellars, R.W. (1922). *Evolutionary naturalism*. Chicago and London: Open Court Publ. Co.
- Sellars, W. (1963). Philosophy and the scientific image of man. In W. Sellars, *Science, perception, and reality*. London: Routledge & Kegan Paul, 1–40.
- Sellars, W., Rorty, R., & Brandom, R. (1997). *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.
- Shadlen, M.N., & Newsome, W.T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, 93(2), 628–633.
- Simon, H.A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118.
- Smith, G. (2007). Newton's philosophiae naturalis principia mathematica. An entry in the on-line *Stanford Encyclopedia of Philosophy* <<http://plato.stanford.edu>>
- Stein, H. (1990). "From the phenomena of motions to the forces of nature": Hypothesis or deduction? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 209–222. Available from <<http://www.jstor.org/stable/193069>>
- Taylor, S., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Prospects for a Stereoscopic Vision of our Thinking Nature: On Sellars, Brandom, and Millikan^{*}

James R. O'Shea[†]
jim.oshea@ucd.ie

ABSTRACT

In this article I consider how the very different but equally Sellars-inspired views of Robert Brandom and Ruth Millikan serve to highlight both the deep difficulties and the prospects for a solution to what is arguably the most central problem raised by Sellars's attempted "stereoscopic fusion" of the "manifest" and "scientific images": namely, the question of the nature and place of *norm-governed conceptual thinking* within the natural world. I distinguish two "stereoscopic tasks": (1) the possibility of integrating a naturalistic theory of animal representation *within* an irreducibly normative inferentialist account of conceptual content; and (2) the possibility of providing a naturalistic explanation *of* the normative "space of reasons" and conceptual thinking as such. Millikan embraces and Brandom resists the naturalistic representationalist hypotheses involved in (1); while Brandom embraces and Millikan resists the conception of pragmatically irreducible normativity involved in (2). The grounds of resistance in each case are arguably suspect.

Introduction

Sellars's 1962 article, *Philosophy and the Scientific Image of Man* (PSIM) is widely recognized as a classic presentation of the profound problems that confront any attempt to account for the nature of the human being – as a

^{*} My thanks to the participants at the *Normativity of Meaning: Sellarsian Perspectives* conference in May, 2011 in Prague organized by Jarda Peregrin, where an earlier version of this paper was delivered; and to University College Dublin for the research time afforded by the award of a President's Fellowship during 2010–11.

[†] University College Dublin.

consciously experiencing, conceptually thinking, and rationally active being – within the bounds of ontological sense that have seemed to Sellars and many other philosophical naturalists to follow from modern scientific conceptions of nature. In that article Sellars succeeded in developing the problems in more detail than he did his own envisaged solutions to those problems, but taking his works as a whole he did attempt to sketch solutions to each of the problems he raises. (See O’Shea 2007, 2009, and 2011 for my own take on Sellars’s overall synoptic vision of persons as sensing, thinking, and acting beings within a scientific naturalist ontology.) In what follows I propose to consider how the views of two well-known systematic philosophers whose views are strongly influenced by those of Sellars – Robert Brandom and Ruth Millikan – can be seen as highlighting both the deep difficulties and the prospects for a solution to what is arguably the most central synoptic problem raised by Sellars: the question of the nature and place of *norm-governed conceptual thinking* within the natural world.

It is of course a matter of vigorous contemporary dispute whether meaning and intentionality are constitutively normative phenomena. Here, however, I propose to examine certain synoptic issues that arise on the assumption of the correctness of the *normativity thesis*, as we might call it. These issues cluster around the familiar but important topic of the consequences of the normativity thesis for naturalism. Sellars, as is well known, defended strikingly comprehensive versions of both the normativity thesis on the one hand, and a thoroughgoing scientific naturalism on the other. Perhaps most controversial by current lights are the particular ways in which Sellars argued for what he conceived of as a *stereoscopic fusion* of (in effect) the normativity thesis and scientific naturalism, by analogy with how, as he put it, «two differing perspectives on a landscape are fused into one coherent experience» (PSIM, p. 4). What exactly is such a stereoscopic vision of our specifically *conceptual* capacities supposed to look like, on Sellars’s view? And what are its prospects in light of more recent developments?

I should note from the outset that “naturalism” on the approach I shall take here, though fully comprehensive, will for present purposes not be taken (*contra* Sellars) to entail any ostensible conflict with the manifest image ontology of ordinary persisting and coloured *physical objects*, such as trees and tables, but only with the manifest ontology of *persons and norms*, which for Sellars presents an importantly different set of problems. The aspects of Sellars’s naturalism that I shall discuss here are very widely shared in

contemporary philosophy and do not involve any particularly controversial conclusions concerning the ultimate *falsity* of the manifest image ontology of physical objects (cf. EPM, p. 173) – but this is naturalism enough to generate the familiar controversies pertaining to the normativity of meaning and conceptual content that I shall be discussing here.

Ruth Millikan has recently provided a particularly helpful entryway to the issues I want to focus on, in her essay, *The Son and the Daughter: On Sellars, Brandom, and Millikan* (2005). In this essay Millikan recounts how, in her own work, she has «pursued the picturing themes from the *Tractatus* that were carried through in Sellars’s discussions of that causal-order relation between language and the world that he called “representing”»; by contrast, she continues, «Brandom has followed Sellars’s interest in the language-games metaphor from *Philosophical Investigations*, expressed in Sellars as a form of inferential role semantics and in the thesis that one learns to think only as one learns to abide by the rules of a language» (Millikan, 2005, p. 77; cf. Brandom 1994). In this essay, however, Millikan ultimately contends that there was what she calls “a crack” in Sellars’s system that accounts for how it is that both she and Brandom remained faithful to central aspects of Sellars’s views while nonetheless radically diverging in their own respective views.

Perhaps surprisingly – although from my perspective, plausibly – Millikan contends that there was no “crack” or inconsistency, *per se*, in the way that Sellars attempted to combine seemingly incompatible central themes from both the early and the later Wittgenstein. There need not be any blatant crack here provided that certain systematic distinctions of level and of aims are recognized. Here is how Millikan briefly describes what she sees as Sellars’s in principle coherent attempt to (as I shall put it) stereoscopically combine certain broadly *Tractarian* and certain later-Wittgensteinian themes within one unified, multileveled account of human cognition:

Indeed, Sellars went to great pains to explain exactly how inferential role semantics was consistent with ‘*Tractarian*’ picturing. The idea was, roughly, that in an individual’s or a community’s following the rules of a language, the language being largely internalized as thought, a very abstract map of the world was in the process of construction. [Here Millikan quotes Sellars’s 1962 article, “Truth and ‘Correspondence’” (in Sellars, 1963, p. 215) on the «fantastically complex system of rules of projection» that are involved in naturalistic picturing or representation.] These fantastic complexities are introduced mainly by the inference rules [...] that govern ‘statement-

statement’ (hence judgment–judgment) transitions. [...] The map of the world produced by a language is not found sentence by sentence but only in the whole of the living *language cum thought* running isomorphically to the whole world in sketch. If there is a crack in the Sellarsian foundation, this is not where it lies, or anyway not precisely. (Millikan, 2005, p. 78)

The topic of this passage provides the first stereoscopic task that I will consider here: roughly, the attempt by Sellars to embed a substantive naturalistic account of mental representation – a conception applauded and extended by Millikan – *within* the sort of normative-inferentialist “space of reasons” account of conceptual thinking defended by both Sellars and Brandom, despite the arguably unnecessary resistance to such naturalistic representationalist accounts by Brandom and other neo-Sellarsian philosophers such as John McDowell. In section II, I shall then briefly consider a second, more problematic stereoscopic task, one likewise championed by Millikan: that of attempting to give a naturalistic account *of* the normatively rule-governed space of reasons and conceptual thinking itself. It is in relation to this task that Millikan will contend that where «there may be a crack» is «in Sellars’s treatment of the nature of linguistic rules and the relation of these to conceptual roles and thus to intentionality» (Millikan, 2005, p. 78).

I

The first stereoscopic task, then, is that of showing how (as highlighted in Millikan’s passage above) the normative dimensions of human conceptual cognition are not only consistent with, but in fact stand in intelligible systematic interrelations with an underlying naturalistic dimension of “language/world” or “mind/world” representational relations.

Both Brandom and McDowell in their differing but highly productive ways have defended the Sellarsian view that properly conceptual representation is possible only within a wider *logical space of reasons* (cf. EPM p. 169). Only within such a normative «ambience of rules of criticism», to use another phrase from Sellars (1968, p. 175), are conceptual thinking and rationally responsible judgment possible. From these deeply Sellarsian perspectives it can seem philosophically disastrous to traffic, as Sellars himself did, in any notions of mental and linguistic representation, at the level of properly human cognition, that cannot be reconstructed in terms of the complex interplay of normative standings within a conceptually structured space of reasons. But I

am not convinced that we ought to reject, in principle, Sellars's interesting attempt to stereoscopically combine his broadly inferentialist account of meaning and conceptual thinking – at least in certain basic empirical domains – with an underlying naturalistic theory of mapping-and-tracking “picturing” or *representational systems*, as he eventually called them in his late (1981) article, *Mental Events*. Let's look a little more deeply into this issue.

In *Mental Events*, after giving a crash course in the first three parts on both his «functionalist theory of meaning and intentionality» (1981, §37) and his Tractarian nominalist theory of predication and picturing-representation, Sellars in the final three parts proceeds to offer an explanation of «what linguistic and non-linguistic representational systems have in common» (1981, §50). In the following passages Sellars lays out some central aspects of his basic conception:

§56. Indeed, I propose to argue that to be a representational state, a state of an organism must be the manifestation of a system of dispositions and propensities by virtue of which the organism constructs *maps* of itself in its environment, and locates itself and its behavior on the map.

§57. Such representational systems (RS) or cognitive map-makers, can be brought about by natural selection and transmitted genetically, as in the case of bees. Undoubtedly a primitive RS is also an innate endowment of human beings. The concept of innate abilities to be aware *of something as something*, and hence of pre-linguistic awarenesses is perfectly intelligible.

§72. [...] (h) Putting it crudely, the fundamental thesis I have been advancing is that while prelinguistic RSs do not have ‘subjects’ and ‘predicates’ they do share with subject-predicate RS the duality of the functions of *referring* and *characterizing*. The fact that in a subject-predicate language these functions involve separate *subject* symbols and *predicate* symbols is, from this standpoint superficial.

(i) All of the above is compatible with the idea that the presence in a RS of subjects and predicates makes possible degrees of sophistication [for example, negation and quantification –J.O’S.] which would otherwise be impossible. But to develop this point would require a discussion of logic-using representational systems.

(And Sellars then does go on to discuss the explicitly *logical* representational resources that he argues distinguish such «logic-using representational systems» from other animal representational systems.)

One question that immediately arises for anyone who is at all familiar with Sellars's views is how the view expressed in the final sentence of §57 above is supposed to be consistent with Sellars's claim in *Empiricism and the Philosophy of Mind* (EPM) that «all awareness of *sorts* [...] is a linguistic affair» (EPM, p. 160), which was a key component of his famous argument against the “myth of the given”. This ostensible inconsistency, however, is part and parcel of the particular stereoscopic task that we are now to explore. My aim in this section is to examine whether such a naturalistic theory of representation as Sellars attempted to sketch in *Mental Events* and elsewhere – with whatever explanatory payoffs it might be thought to have – is coherently available and recommendable to Sellarsian inferentialists of Brandom's stripe, for instance.

One of the key elements in Sellars's account in *Mental Events* and elsewhere (e.g., Sellars, 1963, chs. 6 and 11), it seems, is the role of natural selection in generating the required systematic pattern of normal functioning – a natural biological *selection space*, as I shall call it, as opposed to a *logical space* of reasons – within which particular events or behaviors can be coherently understood as instances of correct or incorrect functioning. Teleosemantic theories such as Millikan's have subsequently attempted to account for the norms of proper functioning in terms of which an animal representational system can be coherently conceived to *misrepresent* various aspects of its environment, thus providing some actual cash for Sellars's schematic gestures toward natural selection in these contexts. Could one in principle embrace the heart of Millikan's teleosemantic conception of animal representational systems while departing from Millikan in maintaining, as Sellars does, a sharp, pragmatically irreducible distinction between logical space normativity and natural selection space normativity?

Consider Sellars's own account of human perceptual cognition; for example, as expressed in the sensorily passive, object-elicited “language entry transition” or non-inferential judgment that there is a red cube on the table. On the one hand Sellars holds that such perceivings require not only differentially receptive sensory capacities, but also the possession and spontaneous exercise of acquired conceptual capacities (involving the capacity to apply the concepts *red* and *cube*, for instance). On the other hand, in parts IV and V of *Empiricism and the Philosophy of Mind*, Sellars had already made clear that we can and should push further in our explanatory ambitions – even while remaining within the “manifest image”, and even without being drawn

into the “homogeneity” problem or “grain argument” that Sellars puts forward concerning sensory qualia (although that problem was admittedly crucial for Sellars himself: cf. PSIM, sections V–VI). In the course of EPM part IV, entitled *Explaining Looks* (note the “explaining”, rather than “analyzing” the concept of “looks”, which occurred in part III, *The Logic of “Looks”*), Sellars remarks as follows concerning what will ultimately amount to his own explanatory posit of “sense impressions” as inner adverbial states of sensing:

Now there are those who would say that the question ‘Is the fact that an object looks red and triangular to S to be explained [...] in terms of the idea that S has an impression of a red triangle?’ simply doesn’t arise, on the ground that there are perfectly sound explanations of qualitative and existential lookings which make no reference to ‘immediate experiences’ or other dubious entities. Thus, it is pointed out, it is perfectly proper to answer the question ‘Why does this object look red?’ by saying ‘Because it is an orange object looked at in such and such circumstances’. The explanation is, in principle, a good one, and is typical of the answers we make to such questions in everyday life. But because these explanations are good, it by no means follows that explanations of other kinds might not be equally good, and, perhaps, more searching. (EPM, p. 150)

As he remarked in a similar spirit in *Science and Metaphysics* a decade later: «Philosophy may perhaps be the chaste muse of clarity, but it is also the mother of hypotheses» (1968, p. 12). Sellars there states as follows the most general form of the more searching explanatory aim that is embodied in his theory of what he explicitly calls «non-conceptual representations» (1968, pp. 16–17, & *passim*):

If what might be called the ‘sense impression inference’ is an inference to an explanation, what specifically is it designed to explain? [...] If we construe physical objects, for the moment, in Strawsonian [that is, in manifest image] terms we can say that the aim is to explain the correlation of the conceptual representations in question with those features of the objects of perception which, on occasion, both make them true and are responsible for bringing them about. (Sellars, 1968, p. 17)

Sellars’s overall explanation, I think, goes roughly like this. The perception that there is a red cube on the table involves, inter alia, both the conceptual and the nonconceptual representation of a red cube. Most philosophers at this stage are broadly familiar with either Sellars’s or Brandom’s normative inferentialist account of what gives the concepts “red” and “cube” their *conceptual* content. Sellars’s further hypothesis, I suggest, is that this

perceptual cognition *also* involves underlying mapping and tracking representational relations of two different kinds: one is of a linguistic kind which, qua representation, is parasitic on its having been generated within a wider logical space of reasons; and the other is of a biological kind which, qua representation, is parasitic on its having been generated within a wider space of natural selection. Sellars's picture, to be explored in what follows, thus has the following components:

Three representational dimensions in Sellars's account of human perceptual cognition:

(1) *Conceptual content*, reference, etc., as accounted for in terms of an inferential space of reasons. Plus 2 kinds of underlying naturalistic representation:

(2a) "*Space of reasons parasitic*" *linguistic* representations (*qua* "natural linguistic objects"); and

(2b) "*Selection space parasitic*" *nonconceptual* representations (e.g., the "sensation of a red cube").

Consider first the "space of reasons parasitic" form of underlying representational relation (i.e., (2a)). To put it very crudely, I have been trained within a normative space of reasons to be disposed to token "red" in response to red objects and "cube" in response to cubical objects in such a way that my current inner or outer tokening of a •this red cube• kind or manner of representing *ought-to-be* causally "correlated" (as we saw Sellars put it oversimply above) with red cubical objects in my nearby environment – other things being equal, of course, and subject to all the very serious objections and qualifications that such causal-covariation accounts of perceptual representation must address. But at least *some* of the standard objections to such theories would be mollified by the fact that Sellars's is a very unusual causal representationalist account, precisely because the relevant causal patterns are in this case established and maintained not directly by nature but indirectly via the normative pragmatic ought-to-be rules (as Sellars calls them, cf. 1968, *passim*) of a logical space of reasons. Those thinkers who are subject to these ought-to-be norms will normally have no such underlying causal representational level or goals directly in view at the normative-pragmatic level (of the "life-world", as it were). But neither should we theorists be so chaste as to suffer from a philosophically imposed ban on whatever explanatory grounds

there may be for the theoretical investigation of such “language game parasitic” cognitive mapping relations. This integrationist or stereoscopic view of our logico-conceptual and natural-nonconceptual representational capacities is a “mother of hypotheses” that is worth further exploration. In particular, an embrace of the Sellarsian space of reasons and the normativity thesis ought not to lead us (as many neo-Sellarsian philosophers have suggested that it should) to avert our eyes from Sellars’s own delicately placed naturalistic representationalist hypothesis of this kind.

As to the second, “selection space parasitic” form of representational system (i.e., (2b)), Sellars’s proposal was that our ostensible perception that there is a red cube on the table will also normally incorporate – given the way that we are in part built *by nature* as well as by culture – a nonconceptual *sensory* representation or “sense impression” of a red cube. The theorist models the content of this “of-a-red-cube” manner of sensing on aspects of its typical outer physical cause (again, with the usual theoretical burdens that spelling out such a view entails).

This hypothesis is without detriment to the fact that we rational beings must also *conceptually* represent the presence of the red cube, as a functionally (normative-inferentially) constrained element within the very same cognitive response, *if* we are to perceptually recognize the cube in the way that situates us *as knowers* within a logical space of reason-giving. This of course points the way toward resolving the supposed inconsistency between Sellars’s views in EPM and in *Mental Events* mentioned earlier: namely, concerning our awareness of “sorts” as somehow both necessarily linguistic and yet also, in some cases and in some respects, innately biological and non-linguistic. Note also that on this view, as Sellars continually stressed throughout his works, the “of” of nonconceptual sensory representation is not the same as the “of” of conceptually contentful intentionality. Nonetheless the former non-conceptual representations, too, are genuinely *representational contents* that can correctly or incorrectly represent the presence of their corresponding objects on any given occasion. The point is that nonconceptual sensory representations, like all representations for Sellars, are “of” their corresponding objects – whether veridically or non-veridically – only in virtue of their place within a wider pattern-governed system or “ambience of norms”. In this case the relevant norms derive from a selection space of nature rather than a logical space of reasons – although in human cognition the one has become systematically integrated with the other.

But perhaps this happy stereoscopic picture puts things too neatly. Bob Brandom's recent Sellars MA Seminar in 2009 at Pittsburgh¹ rightly raises the question of whether Sellars in *Mental Events* really did intend to offer at all the sort of *ur*-Millikan account of biological "selection space normativity" for which I have suggested he was explicitly making coherent philosophical room (at the appropriate level). It is true that "below the line" of conceptual representation proper (to borrow a useful metaphor from McDowell, 2009, chapters 1–3) Sellars discusses animal representational systems primarily in terms of Humean-style associative uniformities and propensities. Brandom in his seminar suggests that by itself this would support only a non-normative *causal functionalist* account of animal cognition, not the more promising account of biological *proper functioning* that Millikan went on to develop in order to fill the sort of crack she finds in Sellars's account. I think it is at least clear, however, that Sellars in the article argued for a distinction between *correct* and *incorrect* "below the line" animal representations that is possible only as a result of such behavior or events being embedded within wider systematic patterns of behavior and cognition, some of which are innate due to a history of natural selection and some of which are learned through associative mechanisms. Sellars was proposing that the systematicity that is required in order to generate the constitutive normative correctness involved in any representational state in general, is at this most basic biological level to be explained by patterns of functioning that have resulted from a history of natural selection. Millikan then takes that vague idea and provides a more comprehensive theory, one that also seeks to explain how various learned associative patterns in animal cognition should be viewed as *derived proper biological functions* based ultimately on naturally selected mechanisms. This is stereoscopic progress in line with the account of Sellars's philosophy that I have been giving, both in general meta-philosophical respects (as a mother of naturalistic hypotheses as well as a chaste muse of conceptual clarity), and in relation to the specific issue of the normativity of non-conceptual representations *qua representations*. Millikan then attempts to push this model all the way up "above the line" into the space of reasons, and here matters do threaten to jar with both Sellars's and Brandom's accounts, as I shall briefly discuss below in relation to the second stereoscopic task.

¹ Kindly made publicly available on his website: <http://www.pitt.edu/~brandom/phil-2245/>.

But Brandom in the seminar also raises a further potential problem with the picture Sellars presents in *Mental Events*. Brandom suggests that something like his own normative pragmatic account of *assertion* (cf. Brandom, 1994, ch. 3) is needed in order to fill a serious gap that he finds in Sellars's remarks on how his complex *nominalist theory of predication* is supposed to support his naturalistic theory of basic representational relations (cf. (2a) above). But in this case I suspect that a specific misconstrual or running together of distinct levels and aims might be involved in this objection, and that no worry of *this* kind need chasten the attempt to stereoscopically fuse Brandom's own more developed pragmatics of assertion with Sellars's "below the line" account of representational relations to the world.

Brandom in this regard focuses on the following remark of Sellars's during his brief discussion of his nominalist theory of predication in *Mental Events*:

[I]t is a truism that the concatenation of 'red' with 'a' tells us that *a* is red.
(Sellars, 1981, §43)

Now, Brandom correctly points out that this remark is not only *nota* truism – it is not even true. For in order for such a concatenation to achieve the status of being a predicative *telling* that *a* is red – as opposed to such tokenings occurring in various contexts in which they accomplish no such thing – we need a more basic account of what it takes for such a concatenation to constitute an *assertion*, in those cases where it does have that force. And this, he suggests, Sellars nowhere provides – or at least certainly not when Sellars goes on from this remark to give his nominalist account of the role of predicates as in principle *dispensable* devices for getting names to have a certain "counterpart" extensional character (such as being-joined-to-the-right-with-an-"is red") that is supposed to be uniformly correlated with specific complexes and sequences of corresponding objects in the world (in this case, with red objects).

I think Brandom's account in *Making It Explicit* of the social pragmatics of assertion – including also the various "above the line" sub-sentential roles of subjects and predicates and other functional elements in assertions, and of the *de re* idioms that serve to underwrite a conception of representational objectivity – constitutes a welcome and substantial development of Sellars's basic account of assertion as, roughly speaking, the norm-governed taking of a position within a wider language "entry/inference/exit" game of giving and asking for reasons. It is a detailed explanation of the force and content of such

Sellarsian assertional standings that Brandom's *Making It Explicit* has attempted to provide, displaying Sellars's philosophy in this respect, too, as a fruitful mother of hypotheses.

Despite the insufficient remark from Sellars quoted above, however, Sellars would surely agree that concatenating is not sufficient for saying or asserting. The relevant "above the line" normative semantic rule, in this case, is roughly that it ought-to-be the case (*ceteris paribus*) that speakers respond to red *a*'s in relevant circumstances by uttering "*a* is red". If such an utterance-uniformity is the result of the right sort of wider space of reasons, which involves the satisfaction of various other important pragmatic presuppositions (many of which only Brandom actually analyses and explains), such concatenations or utterings can then constitute *tellings to* or *assertings that*. As Sellars explains in the article, in one primary sense *what the job is* of any given predicate is to play a certain normatively functionally classifiable (i.e., •dot-quotable•) conceptual role of this kind within a language game or space of reasons. But in another related sense, Sellars hypothesizes, what *that* predicational job succeeds in generating at the most basic empirical level (rather than in the domains of mathematics or morality, for instance, on Sellars's view) is to give a name a certain empirical or "natural-linguistic" character that *thereby*— thanks to the very behavioral uniformities that have been generated by the higher level rule-governed language game — has in fact become systematically causally correlated with a different but corresponding character in the object that is *thereby* represented. It is perhaps Brandom's objection to this account in Sellars, rather than the latter itself, that has run together the two different levels that both he and Sellars in general correctly recognize need to be carefully distinguished from one another.

What Sellars's sloppy remark above is designed to remind us of, then, is that *assuming* that such an above-the-line inferential practice is in place and having real effects, what will be systematically generated below-the-line at the most basic level, concerning those very same rule-governed linguistic tokenings, is a highly complex set of language-world mapping and tracking relations. The sloppiness reflects Sellars's ambitious but inevitably awkward attempt to envisage the interplay of those two levels simultaneously (the difficult Virgin Mary task, as it were, of being chaste muse and mother of hypotheses in one expository go), in what I nonetheless continue to think remains a philosophically coherent stereoscopic approach to the natural-

representational upshots of certain regions of our normative-conceptual achievements as thinkers.

A similar kind of response, I think, should be made to a third general worry raised in Brandom's seminar in relation to Sellars's naturalistic account of representation in *Mental Events*. This objection concerns Sellars's "extended" application in *Mental Events* of the dot-quoting device, which is appropriate to linguistic role-players within a logical space of reasons, to non-language using animal representational systems (cf. Sellars, 1981, §§76–77). The worry is that this extension confuses between normative space of reasons dependent *meaning* or "signification" on the hand (the conceptual "of" of intentionality), and the supposed underlying naturalistic dimension of picturing-representation on the other (the nonconceptual "of" of sensory content, causal covariance, and tracking-isomorphism), in just the way that Sellars had accused the Thomistic philosophers of doing in his article, *Being and Being Known* (Sellars, 1963 ch. 2). But again rather than confusion what we have in Sellars's account above is a distinction between two kinds of normative space and two correspondingly different kinds of representation: namely, (2a) logical "space of reasons parasitic" *linguistic* representations or "pictures" (*qua* "natural linguistic objects", as Sellars calls them), and (2b) biological "selection space parasitic" nonconceptual representations or "pictures" (exploited, for example, by an animal's "mapping and tracking" sensory cognition and pursuit of its prey). The dot-quoting device would seem to be intelligibly and fruitfully extended to refer to the proper biological functions that constitute the sort of selection space within which nonconceptual animal representations, on this view, are possible.

The deservedly influential appropriations of Sellars's views on the myth of the given and the logical space of reasons by Rorty, Brandom, and McDowell have unfortunately carried the suggestion that the underlying naturalistic dimension of *representation* discussed in this section must be discarded if we are to be able coherently to preserve the irreducibly normative dimensions of Sellars's conception of our empirically contentful thought and perceptual knowledge. I have argued above that this is not true. At any rate, without a firm grip on Sellars's simultaneously naturalistic and normative conception of representation one cannot understand what he meant when he wrote in *Philosophy and the Scientific Image of Man* that

[W]hatever else conceptual thinking makes possible – and without it there is nothing characteristically human – it does so by virtue of containing a way of representing the world. (PSIM, p. 17)

II

Which brings us finally to a second main stereoscopic task explored by Sellars, one that is no less controversial and, I think, conceptually cloudier than the first. While the first task attempted to integrate a causal-naturalistic conception of representation *within*, and partly parasitic upon, a conception of rule-governed conceptual representation, the second stereoscopic task involves taking a naturalistic explanatory stance *on* our higher-level rule-governed conceptual activities themselves. This second task breaks into two: one is to clarify the *irreducibility* of the relevant normative-pragmatic phenomena within the space of reasons; the other is to consider whether, and if so in what sense, it might make coherent sense to aspire to a fully adequate *naturalistic explanation* of those same irreducibly normative phenomena.

On the irreducibility question Millikan in this article correctly gestures toward at least two senses in which, for Sellars, the normative conceptual domain remains *conceptually and pragmatically irreducible*:

Thus, normative rules, for Sellars, are not translatable into nonnormative terms. Accepting a normative rule is not believing a fact but tending to be motivated in a certain way. (Millikan, 2005, p. 80)

For present purposes I am not concerned with whether or not Millikan has this just right, but rather want simply to endorse the idea that, in some sense, normative ought-to-be rules have an irreducible pragmatic status and functional role within Sellars's overall view. I have elsewhere emphasized the importance and pervasiveness of this dimension of normative-pragmatic irreducibility throughout my interpretation of what I call Sellars's *naturalism with a normative turn* (O'Shea, 2007, 2009). This comes out most clearly in Sellars's conception of what, in *Philosophy and the Scientific Image of Man*, he calls "the radical difference in level between man and his precursors", where this difference is conceived (following Kant and the later Wittgenstein) in terms of the irreducible normativity of human conceptual thinking and intentional action in general: «To be able to think is to be able to measure one's thoughts by standards of correctness, of relevance, of evidence» (PSIM, p. 6). On Sellars's stereoscopic view, however, this irreducibly and

constitutively normative dimension of human conceptual thinking and rationality *itself* calls for a careful scientific naturalist explanation:

There is a profound truth in this conception of a radical difference in level between man and his precursors. The attempt to understand this difference turns out to be part and parcel of the attempt to encompass in one view the two images of man-in-in-world which I have set out to describe. For, as we shall see, this difference in level appears as an irreducible discontinuity in the *manifest* image, but as, in a sense requiring careful analysis, a reducible difference in the *scientific* image. (PSIM, p. 6)

What is difficult, however, is to see exactly what Sellars takes such a naturalistic explanation of the normative dimension of human conceptual thinking to be, somehow explaining or analyzing “what appears as an irreducible discontinuity in the *manifest* image” to be in fact “a reducible difference in the *scientific* image”. In the works cited above I have examined Sellars’s further articulations of this global stereoscopic task in terms what he characterized as the *conceptual* (normative pragmatic) *irreducibility* yet *causal* (scientific naturalist) *reducibility* of various aspects of the manifest image. But these conceptions only take Sellars so far in his attempt to articulate what the problem is and what general form any solution to it must take. Here I want to close with some thoughts on how this second stereoscopic task arises in the works of Millikan and Brandom discussed in Section I.

Millikan remarks as follows on the *naturalistic* side of Sellars’s explanatory approach to the irreducibly normative dimensions of human cognition and conduct:

It is one thing to *use* semantic language, for example, to say and mean or to understand “‘rot’ means red”. But you can also *describe* the use of semantic language without using it. You can describe what patterns of response in a language community, along with the origins of these responses in a history of language training, and training of the language trainers, and so forth, constitutes that ‘rot’ means red in that community. You can understand what the ‘means’ rubric does without indulging in it. You can understand specific forms of semantic assessment without participating in the particular practices being examined. There are *truth-conditions* for “‘rot’ means red” of a perfectly ordinary, if very complicated sort. It’s just that it’s not the job of the sentence “‘rot’ means red” to impart the information that these truth-conditions hold. Rather, its job is to get one to use ‘rot’ as one already knows to use ‘red’. (Millikan, 2005, pp. 80–81)

The idea is that in this way one can in principle naturalistically *explain* not only what gives rise to but also, in some sense – this is the more controversial and difficult claim – what *constitutes* meaning-classificatory statements as such, without *using* “means” statements in any non-eliminable way in the explanans. This is not a view from nowhere, of course, but from within an explanatory conceptual framework with its own space of reasonings, which a normative “means” vocabulary can in turn make explicit. As I conceive it – granting that some of Millikan’s own remarks in the passage might unfortunately suggest otherwise – this is also not a “sideways on” view in the sense that McDowell criticizes in *Mind and World* (1994, e.g., pp. 34–36). A sideways on view, as McDowell there explains it, mistakenly presupposes that both a targeted system of concepts and its relationship to the world can be understood separately from and independently of an “internal” and engaged knowledge of the normative functioning of the relevant system of concepts. Whereas I think Sellars would agree with the idea that, for example, the functioning of normative-classificatory semantic vocabulary *will not even be a target that is in explanatory view* unless the explainer also understands its specific normative-pragmatic functioning (either from the inside or by efforts of interpretation from a relevantly similar normative-pragmatic space of reasons). The task is precisely in this way conceived as *stereoscopic* rather than “sideways on”.

One way to put this second stereoscopic ambition, I think, is provided by Brandom himself in his Locke lectures, *Between Saying and Doing* (2008). Brandom’s key methodological innovation in the Locke lectures concerns what he calls *pragmatically mediated* semantic relations between vocabularies, enabling him to present his “analytic pragmatist” as a successor to the classical logicist, empiricist, and naturalist “core programs” of twentieth century analytic philosophy. An example Brandom gives of what he calls (strict) “pragmatic expressive bootstrapping” within this account is the case of providing «an extensional metalanguage for intensional languages, as in the case of possible worlds semantics for modality» (Brandom, 2008, p. 11). (It is “bootstrapping”, for example, in the sense that the metalanguage is expressively weaker than the target language it explicates.) And then Brandom adds the following interesting example concerning Huw Price’s naturalism (2011):

One example of a claim of this shape in the case of pragmatically mediated semantic relations [...] is Huw Price’s pragmatic normative naturalism. He argues, in effect, that although normative vocabulary is not *reducible to*

naturalistic vocabulary, it might still be possible to *say* in wholly naturalistic vocabulary what one must *do* in order to be *using* normative vocabulary. If such a claim about the existence of an expressively bootstrapping naturalistic pragmatic metavocabulary for normative vocabulary could be made out, it would evidently be an important chapter in the development of the naturalist core program of the classical project of philosophical analysis. It would be a paradigm of the sort of payoff we could expect from extending that analytic project by including pragmatically mediated semantic relations. (Brandom, 2008, pp. 10–11)

In his important collection of essays, *Naturalism Without Mirrors* (2011, p. 29), Price in this spirit comments on the ironic flavor that this stereoscopic explanatory project takes on when one applies it to one's own normative practices: as a social scientist does, for example, when she temporarily views herself during such explanatory bouts as an example of her own general object of inquiry. There would seem to be nothing *incoherent* in this sort of non-reductive, naturalistic-explanatory aspiration, as long as the vocabularies of the explainings and of their targeted practices are both kept clearly in view. (Incidentally, Price himself, like Sellars, and in *this* respect unlike Brandom and McDowell, appears to be willing in principle to incorporate into his global pragmatic *anti-representationalism* an underlying dimension of naturalistic “mapping and tracking” representations at least in certain domains, in roughly the ways I have sketched in relation to the first stereoscopic task in section I.)

But what exactly is Brandom's attitude toward what he here praises as Price's “would be” naturalistic analysis of the use of normative vocabulary (i.e., according to which, as quoted above, it would be «possible to *say* in wholly naturalistic vocabulary what one must *do* in order to be *using* normative vocabulary»), as part of what generally features in Brandom's book as the superseded “core program” of classical analytic naturalism? Not only here but also in various places in commenting on Ruth Millikan's very different biological naturalism, Brandom offers praise and does not outright reject but certainly does not endorse the proposed naturalistic explanations of our normative-linguistic behavior.

Brandom's earlier *Making It Explicit* had presented norms as at once irreducible to the causal order and yet also as non-mysterious from a naturalistic point of view: on the one hand it is «norms all the way down» (1994, p. 44), and Brandom asserts that «Norms [...] are not objects in the causal order» (1994, p. 626); but on the other hand «Normative statuses are

domesticated by being understood in terms of normative attitudes, which *are* in the causal order» (1994, p. 626). In various places in *Making It Explicit* Brandom puts forward considerations that would seem to be entertaining something *close* to our second stereoscopic aim, doing so primarily in order to gesture toward an in principle available *de-mystification* of our normative practices from a naturalistic explanatory perspective, rather than as part of his own self-described task of explaining «what it is to grasp a propositional content» *per se*:

Thus no attempt will be made to show how the linguistic enterprise might have gotten off the ground in the first place. But it should be clear at each stage in the account that the abilities attributed to linguistic practitioners are not magical, mysterious, or extraordinary. They are compounded out of reliable dispositions to respond differentially to linguistic and nonlinguistic stimuli. Nothing more is required to get in to the game of giving and asking for reasons – though to say this is not to say that an interpretation of a community as engaged in such practices can be paraphrased in a vocabulary that is limited to descriptions of such dispositions. Norms are not just regularities, though to be properly understood as subject to them, and even as instituting them by one’s conduct (along with that of one’s fellows), no more need be required than a capacity to conform to regularities. (Brandom, 1994, pp. 155–156)

This passage brings out the key difficulty that would seem to be involved in the second stereoscopic task, a task that I characterized above in terms of the idea of taking a naturalistic explanatory stance *on* our higher-level rule-governed conceptual activities themselves. In the passage Brandom suggests that the abilities required in order both *to institute norms* and *to be subject to norms* – «to get into the game of giving and asking for reasons» – requires «no more [...] than a capacity to conform to regularities» of certain kinds. But at the same time, he makes clear, to *interpret* a community as engaged in a normative practice of giving and asking for reasons of this kind, and hence to be able to understand the conceptually contentful activities as such that result from and are constituted by such norms, is already to be engaged at the level of attributing and evaluating such normative statuses (it is to be engaged in “deontic scorekeeping”, to use the terms of Brandom’s model), rather than to be at the level of attempting to explain what is required to generate such normative statuses in naturalistic terms.

I think that there is much that is fundamentally correct – correct both in itself and as an interpretation of Sellars’s position – in the view defended in

different ways by both Brandom and McDowell that, to put it metaphorically, the shape of the normative dimensions that are constitutive of human rationality are, in an important sense, discernible only from *within* the perspective of the rule-governed activities that constitute such a “space of reasons” as such a space. Nevertheless, the sympathetic dissatisfaction with that outlook that is frequently expressed by naturalistic philosophers such as Millikan (and also Dennett 2010) is a reflection of a legitimate explanatory aspiration that was also shared by Sellars. It seems to me that there are coherent grounds for hope for further progress on this last naturalistic front while simultaneously stereoscopically retaining a conception such as Brandom’s of the normative-pragmatic irreducibility of human sapience within the “logical space of reasons”. I will close with a few final remarks on this thought.

Millikan argues that both Sellars and Brandom fall short in relation to this second stereoscopic task, suggesting (as noted earlier) that where «there may be a crack [...] is in Sellars’s treatment of the nature of linguistic rules and the relation of these to conceptual roles and thus to intentionality»:

Putting things bluntly, it seems that Sellars understands accepting semantic norms as merely displaying certain dispositions, dispositions to make certain moves in language and thought and dispositions to sanction these moves in others. Brandom claims that this sort of analysis will not do.

[...] Now, I agree with Brandom that conceptual norms must be disposition-transcendent, hence with his rejection of Sellars’s view of norms as derived from meta-dispositions to sanction. (Millikan, 2005, p. 81)

However, the grounds stated here for Millikan’s resistance to what she takes to be Sellars’s own allegedly *merely dispositional* solution to the second stereoscopic task is arguably based on a confusion of the aims of the two interrelated levels in Sellars’s account (the normative and the natural) that is similar to what we found in relation to Brandom’s resistance to the first stereoscopic task in section I. Brandom’s worry about the first stereoscopic task, we may recall, was that Sellars’s naturalistic, nominalist theory of predication-as-picturing allegedly fails to account for the normative force of assertional “telling” – while in fact, as we saw, this is not the job of those underlying “language game parasitic” representations, on Sellars’s view. In relation to the second stereoscopic task, just before her negative verdict expressed in the passage above, Millikan had carefully distinguished (as we saw in the passage quoted at the outset of this section) between, on the one hand,

the normative pragmatic level of rule-governed practices *internal* to which we say that things mean so-and-so, are true, and so on; and on the other hand, the naturalistic attempt to explain those rule-governed practices in terms of a more parsimonious scientific theoretical vocabulary. But when she then accuses Sellars of failing to account for the “disposition transcendence” of conceptual norms, she chooses to focus on the latter naturalistic explanatory perspective, rather than on the internally engaged, normative-pragmatic perspective. Surely Sellars, however, in this respect like Brandom, would and should address the disposition-transcending rule-following difficulties from *within* the fray of critically reflective perspectives in terms of which we engage each other and attempt to fix belief about an objective world, rather than in purely dispositionalist terms. The rule-following issues are admittedly difficult ones, but I would think that it is a mistake to hold, with Millikan above, that Sellars should be interpreted as impaling himself on the “dispositionalist” horn of the classic rule-following dilemma.

In relation to Brandom’s own “scorekeeping in a game” model, for its part, Millikan comments as follows: «There must be a deep divide between language and ordinary games that we should try not to obscure with a metaphor but instead to keep in full view» (Millikan, 2005, p. 81). She goes on to suggest that it was Sellars’s competing conception of *natural selection based* patterns of cognition and learning, as discussed in section I above, that can provide the required disposition-transcendent source of normativity at the properly conceptual level, too. As is well known, Millikan’s own detailed and sophisticated teleosemantic theory thus attempts to push the natural biological “selection space” model *all the way up* to account for the disposition-transcendent normativity of distinctively human conceptual cognition within a logical space of reasons. Here is just a snapshot of Millikan’s overall outlook on the biological nature of the constitutively normative dimensions of human thought and action (which she here happens to state summarily in terms of the familiar computer metaphor):

The human mainframe takes, roughly, stimulations of the afferent nerves as input both to program and to run it. It responds, in part, by developing concepts, by acquiring beliefs and desires in accordance with these concepts, by engaging in practical inference leading ultimately to action. Each of these activities may, of course, involve circumscribed sorts of trial and error learning. When conditions are optimal, all this aids survival and proliferation in

accordance with a historically normal explanation, one of high generality, of course. (Millikan, 1993, p. 95)

Unlike the *biological* theory of the normativity of disposition transcendent concepts and practical inference that Millikan develops in impressive detail, however, Sellars's focus in relation to our second stereoscopic task was primarily on the possibility of a non-reductive yet fully naturalistic (and broadly *expressivist*) account of the *intentions*, both community and individual, in virtue of which norms are instituted and rule-following behavior takes place (cf. O'Shea 2007, ch. 7, and 2009 for more on this). The challenge in this admittedly murky domain is for the account of norm-instituting intentions to succeed in being both *substantively naturalistic*, going beyond a mere "token-token physicalism" to explain causally the genesis and maintenance of norm-governed patterns of thought and behavior *as such*, while also successfully accounting for the dimension of normative-pragmatic irreducibility such that the relevant patterns can, in another sense, only be discerned by those engaging in them. Although this second stereoscopic task remains unfinished business, this is the place to which Sellars's own explanatory ambitions took him in his attempt to imagine how the gap might be bridged between the sorts of views later defended by Millikan and by Brandom.

We have seen that there are crucial theoretical domains in which "the son" and "the daughter" have each, from different directions, substantially improved upon Sellars's underdeveloped sketches in those regions, those fruitful sketches having served as the philosophical mother of their more detailed explanatory hypotheses. But we also saw that the son and the daughter each attempts – as one does in philosophy – either to have the social normativity "all the way down", in the case of the son, or the biological normativity "all the way up", in the case of the daughter. We found, further, that the grounds for the resistance of the son to Sellars's naturalistic hypothesis of both "logical space parasitic" and "selection space parasitic" *representational systems* – a hypothesis broadly embraced by the daughter – were arguably based on insufficient exploration of the shape of Sellars's stereoscopic proposal at this level. And likewise, the resistance of the daughter to Sellars's account of disposition-transcendent conceptual norms in terms of pragmatically irreducible yet naturalistically non-mysterious rule-following behavior – a hypothesis broadly embraced by the son – is also arguably based on an insufficient exploration of how a more indirect, demystifying naturalism about our institution of and conformity to social norms might be defensible.

Much more work needs to be done on each of the two stereoscopic tasks only briefly adumbrated here. But what has implicitly emerged by implication is that perhaps the hypothesis of a stereoscopic reconciliation of at least the hearts if not the full ambitions of these two sibling Sellarsian perspectives is worth pursuing further.

REFERENCES

- Brandom, R. (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA and London: Harvard University Press.
- Brandom, R. (2008). *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.
- Dennett, D. (2010). The Evolution of “Why?” In B. Weiss & J. Wanderer (Eds.), *Reading Brandom: On “Making It Explicit”*. London: Routledge, 48–62.
- McDowell, J. (1994). *Mind and World* (2nd ed. with new introduction, 1996). Cambridge, MA and London: Harvard University Press.
- McDowell, J. (2009). *Having the World in View: Essays on Kant, Hegel, and Sellars*. Cambridge, MA and London: Harvard University Press.
- Millikan, R. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Millikan, R. (2005). The Son and the Daughter: On Sellars, Brandom, and Millikan. In R. Millikan, *Language: A Biological Model*. Oxford: Clarendon Press, 77–91.
- O’Shea, J. (2007). *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge: Polity Press.
- O’Shea, J. (2009). On the Structure of Sellars’s Naturalism with a Normative Turn. In W. de Vries (Ed.), *Empiricism, Perceptual Knowledge, Normativity, and Realism: Essays on Wilfrid Sellars*. Oxford: Oxford University Press, 187–210.

- O'Shea, J. (2011). How to be a Kantian *and* a Naturalist about Human Knowledge: Sellars's Middle Way. *Journal of Philosophical Research*, 36, 327–359.
- Price, H. (2011). *Naturalism Without Mirrors*. Oxford: Oxford University Press.
- Sellars, W. (1968). *Science and Metaphysics: Variations on Kantian Themes*. London: Routledge & Kegan Paul.
- Sellars, W. (1981). Mental Events. *Philosophical Studies*, 39, 325–345.
- (PSIM) Sellars, W. (1963). Philosophy and the Scientific Image of Man. In W. Sellars, *Science, Perception and Reality*. New York: Humanities Press, 1–40. [First ed. by University of Pittsburgh Press, Pittsburgh, 1962]
- (EPM) Sellars, W. (1963). Empiricism and the Philosophy of Mind. In W. Sellars, *Science, Perception and Reality*. New York: Humanities Press, 127–196. [First ed. by University of Minnesota Press, Minneapolis, 1956]
- Sellars, W. (1963). *Science, Perception and Reality*. New York: Humanities Press.

Quine and Wittgenstein on the Science/Philosophy Divide *

Diego Marconi[†]
diego.marconi@unito.it

ABSTRACT

In this article I first sketch what I take to be two Quinean arguments for the continuity of philosophy with science. After examining Wittgenstein's reasons for not accepting the arguments, I conclude that they are ineffective on Wittgenstein's assumptions. Next, I ask three related questions: (a) Where do Quine's and Wittgenstein's philosophical views essentially diverge? (b) Did Wittgenstein have an argument *against* the continuity of science with philosophy? (c) Did Wittgenstein believe until the end of his philosophical career that scientific results are philosophically irrelevant? It will be seen that all three questions are related with Wittgenstein's distinction between conceptual and factual issues. I conclude that the opposition between Quinean philosophy and Wittgensteinian philosophy is genuine.

1.

In his book *Wittgenstein's Place in Twentieth Century Analytic Philosophy* (Hacker, 1996), P.M.S. Hacker set up a very sharp opposition between Wittgenstein and analytic philosophy, on the one side, and Anglo-American philosophy drawing inspiration from Quine on the other. As a way of identifying analytic philosophy, the opposition is unconvincing. Hacker rightly insists on the diversity of the analytic tradition, pointing out that different notions of philosophy's role and even different notions of analysis prevailed

* This paper was presented as the Presidential Address to the Third Conference of the Italian Society for Analytic Philosophy (SIFA), Bologna, September 24-26, 1998. Though long available from the author's web page, it was never published in a journal.

[†] University of Torino, Italy.

with different philosophers at different moments. But then, he wants to exclude Quine and other philosophers he regards as Quinean from the analytic tradition, without it being quite clear why the cleavage between Quine and the later Wittgenstein, or between Quine and Austin, should be so much wider or more crucial than the difference between, say, Austin and Russell (who are both included in the tradition).

Anyway, in drawing the opposition Hacker focusses on one aspect that I would also like to concentrate upon. According to him, post-Quinean philosophy appears to be dominated by «modes of thought that emulate the forms of scientific theories, the jargon and formalization of respectable science, without the constraints of systematic data collecting, quantitative methods and experimental testing» (Hacker, 1996, p. 266); whereas analytic philosophy properly so called always conceived of itself as being other than science¹, and the later Wittgenstein insisted that the attempt to emulate or ape natural science typically produces bad philosophy. In Hacker's own words,

A fundamental tenet of analytic philosophy, from its post-*Tractatus* phase onwards, was that there is a sharp distinction between philosophy and science. Philosophy [...] whether or not it is conceived to be a cognitive discipline, is conceived to be a priori and hence discontinuous with, and methodologically distinct from, science. Similarly, analytic philosophy in general held that questions of meaning antecede questions of truth, and are separable from empirical questions of fact. If Quine is right, then analytic philosophy was fundamentally mistaken. (1996, p. 195)

Now, indicting post-Quinean philosophy for rejecting any sharp demarcation with respect to science (i.e., for seeing itself as continuous with science) is not the same as indicting it for being pseudo-science, or, to borrow Putnam's word², parascience. Hacker does not seem to distinguish clearly between the two charges. That one doesn't recognize a sharp divide between philosophy and science doesn't seem to entail the assertion that philosophy *just is* science (that baldness is vague does not entail that everybody is bald); even less does it oblige one to practice philosophy as quasi-science, i.e., as something that imitates certain superficial features of genuine science, though it is not really science. These are three different things: the continuity of philosophy with

¹ It should be noted that this does not apply to Russell, for one.

² Putnam, 1992, p. 141.

science, the identity of philosophy with science (or, perhaps more plausibly, the inclusion of philosophy within science), and the faking of science by philosophy. What I particularly want to underscore is that continuity – the lack of a sharp demarcation – entails neither identity nor partial identity, i.e., inclusion. Analogously, political theory is not the same as politics, yet a particular speech or a particular essay by a political leader can be both a piece of political theorizing and a political act – an act of practical politics. Or again, theoretical physics is not the same as mathematics; however, there are contributions that one wouldn't know where to range, whether in physics or in mathematics. Moreover, if we take *relevance* as a criterion of continuity, so that discipline A is continuous with B if there are results of B that can be appealed to in order to establish theses belonging to A – perhaps not a bad way of identifying continuity – then many sciences turn out to be continuous with many others: biology with chemistry, sociology with psychology, perhaps every natural science with physics. Thus continuity does not seem to imply identity or inclusion.

2.

It scarcely needs arguing that Wittgenstein, early and late³, strongly opposed both the idea that philosophy is part of science, or one of the sciences, and the idea that philosophy is in any way analogous to science. Perhaps Wittgenstein would have agreed with Putnam in addressing the charge of “parascience” against several of today’s philosophers: for example, against philosophers working in the neighbourhood of cognitive science such as Jerry Fodor, Ruth Millikan, or Daniel Dennett.⁴ On the other hand, his opposition to the idea of continuity is not equally clear, not immediately at any rate. Ultimately, I will claim that Wittgenstein did indeed reject both the idea of continuity and the arguments supporting it. It is, however, interesting to examine such arguments together with Wittgenstein’s reasons for rejecting them, for it allows us to highlight some crucial points that are relevant to the structure of today’s

³ See T, 4.111; PI, §109.

⁴ Putnam’s own criticism appears to be aimed at both Fodor and Millikan and at “analytic metaphysics” as practised by D.Lewis and B.Williams: «Most constructions in analytic metaphysics do not extend the range of scientific knowledge, not even speculatively. They merely attempt to rationalize the ways we think and talk in the light of a scientific ideology» (1992, p. 141).

philosophical discussion. In this article, I will first sketch what I take to be two Quinean arguments for the continuity of philosophy with science. In each case, I will present Wittgenstein's reasons for not accepting the argument. I will conclude that the Quinean arguments are ineffective from Wittgenstein's point of view. Next, I will ask three separate though related questions: (a) where do Quine's and Wittgenstein's philosophical views essentially diverge, (b) did Wittgenstein have an argument *against* the continuity of science with philosophy, (c) did Wittgenstein really believe, to the very end of his philosophical career, that scientific results are philosophically irrelevant. It will be seen that all three questions are related with Wittgenstein's distinction between conceptual and factual issues. I will then conclude that the opposition between Quinean philosophy and Wittgensteinian philosophy is quite real. I do not intend to provide a solution to the problem (or family of problems) on which they took opposite sides; I only want to bring out the structure of their disagreement.

So, let us first look at Quine's arguments for the continuity of philosophy with science. In *Natural Kinds*, while discussing the legitimacy of appealing to empirical generalizations or to scientific theories such as Darwin's in order to justify a philosophical principle, Quine says the following:

I see philosophy not as an *a priori* propaedeutic or groundwork for science, but as continuous with science. I see philosophy and science as in the same boat – a boat which, to revert to Neurath's figure as I so often do, we can rebuild only at sea while staying afloat in it. There is no external vantage point, no first philosophy. All scientific findings, all scientific conjectures that are at present plausible, are therefore in my view as welcome for use in philosophy as elsewhere (Quine, 1969b, pp. 126–127).

As Barry Stroud pointed out (1995, p. 38), Quine never made clear which conception of philosophy he thought he was attacking here or in other similar texts, i.e., what we should understand by an “*a priori* propaedeutic” or by “first philosophy”. Perhaps, as Stroud suggests and Hacker would gladly go along with, he had in mind

something that philosophers for many years certainly said they were doing, or said they ought to be doing: “analyzing” the concepts and principles of science or of everyday life [...] an *a priori* unpacking of the empty form or structure of our thought, or the discovery of the formal principles which any respectable inquiry must follow, quite independently of whatever “content” might come to

fill that form. (Stroud, 1995, p. 38)

Maybe Quine was thinking of some such Kantian enterprise. On the other hand, sticking to the letter of what he says and keeping in mind the positivistic legacy that lies at the root of his reflection, it is more plausible to suppose that he is distancing himself from a *puristic* conception of philosophy: i.e., a conception on which no empirical or factual assumption has a legitimate place in a philosophical argument. The crucial expression in the text I quoted is “external vantage point”: a puristic conception of philosophy appears to presuppose an external vantage point, what he elsewhere called “cosmic exile” (Quine, 1960, p. 275). Positivism is relevant here, for Quine’s view can be seen as a radicalization of Carnap’s thesis of the meaninglessness, or “lack of cognitive content”, of so-called external questions. It is well known that Carnap regarded questions that are not raised within some language – questions, i.e., that do not presuppose the rules of some language or other – as more or less disguised questions concerning the aptness of adopting one language rather than another (a practical, not a theoretical issue according to Carnap). Such questions, e.g., “Are there numbers?”, are not amenable to a formulation «in terms of the common scientific language» (Carnap, 1950, p. 209) Quine’s view can be seen as Carnap’s view *minus* the conventionalist framework that Carnap was taking for granted. Like Carnap, Quine believes that all meaningful questions presuppose the rules of some language or other (there is no external standpoint, no cosmic exile); unlike Carnap, however, Quine regards it as mistaken even to imagine oneself in a position of uncertainty, or indeterminacy, or freedom of choice among different languages (where, as Carnap says, no meaningful questions could be asked). For we are all the time speaking within a language, our common language, which is the background of all scientific theories. To be sure, Quine is not talking in terms of the *rules* of a language – he is not saying that any meaningful question presupposes the rules of some language – for he regards the distinction between rules and statements or propositions as dubious, and that since the mid-Thirties.⁵ Thus taking a language for granted, or speaking from within a

⁵ I.e., since *Truth by Convention* (Quine, 1936). In that article, Quine examined the suggestion that logical principles such as «(II) *Let any expression be true which yields a truth when put for ‘q’ in the result of putting a truth for ‘p’ in “If p then q”*» might be conventions that are «adopted through behavior, without first announcing them in words». If we accepted such a suggestion, «the conventions [would] no longer involve us in vicious regress»: i.e., it would no longer be true that we

language, does not amount to presupposing the rules of that language *as opposed to* presupposing the truth of certain statements couched in that language.

Thus it appears that, for Quine, rejecting the idea of a first philosophy does not so much amount to rejecting transcendentalism (the “a priori unpacking of the empty form or structure of our thought”, in Stroud’s words), nor does it exactly coincide with rejecting epistemological foundationalism (“the discovery of the formal principles which any respectable inquiry must follow”). What Quine is rejecting is, more generally, the idea that one could do philosophy without assuming whatever presuppositions are implicit in the adoption of a language; or perhaps we should say, in order to avoid all conventionalistic overtones, that they are implicit in the very fact of having, and using, a language. Occasionally, Quine referred to such presuppositions by the phrase “conceptual scheme” (Davidson’s Third Dogma). We inevitably speak and argue from within a conceptual scheme. Consequently, Quine concludes, we might as well go all the way:

No inquiry being possible without some conceptual scheme, we may as well retain and use the best one we know – right down to the latest detail of quantum mechanics, if we know it and it matters (1960, p. 4).

This is, then, Quine’s essential motivation for the continuity of philosophy with science: as we are anyway speaking and arguing from within some conceptual scheme – *our* conceptual scheme – we might as well exploit the whole of science, “right down to the latest detail of quantum mechanics”.

Notice that Quine is not here saying that, speaking as we are from within our conceptual scheme, we are *as a matter of fact* assuming the whole of science and we simply ought to acknowledge the fact. That would be an obvious *non sequitur*: it is surely not immediately clear that the adoption of any conceptual scheme whatever involves the adoption of science, indeed, of the

need logic to infer logical truths from conventions such as (II), as Quine shows we do by a Lewis-Carroll-like argument (1936, pp. 96–97). However, Quine is suspicious of the idea of a convention that is adopted before it is formulated: «When a convention is incapable of being communicated until after its adoption, its role is not so clear» (1936, p. 99). For Quine, only behavior that is explicitly based on an explicitly formulated rule can be described as “rule-following”; behavior allegedly based on unformulated conventions «is difficult to distinguish from that in which conventions are disregarded» (1936, p. 99). But if rules coincide with their formulations, the very distinction between rules and (other kinds of) propositions or statements is at risk.

whole of science. And it would be odd for Quine to claim that *our* conceptual scheme involves the whole of science: scientific knowledge, with or without quantum mechanics, is neither so widespread nor so effectively influential to be plausibly regarded as part of our conceptual scheme. In this respect, the literary tradition of the West (within which science only plays a minor role) would be a more plausible candidate. Anyway, Quine is not claiming that science is our conceptual scheme, or part of it; Quine is saying, rather, that *we would do well* to adopt science as our conceptual scheme, for, as conceptual schemes go, it is the best available.

One could object to Quine that the grounds he gives for the adoption of science as a conceptual scheme do not really justify such a commitment. Granted, we are anyway speaking and arguing from within a conceptual scheme – *our* conceptual scheme. But why should we saddle ourselves with the whole of science, down to the latest details of quantum mechanics, rather than keeping our conceptual-schematic commitments to a minimum? Can't we rest content with adhering to the grammar and semantics of our mother tongue (which does not appear to involve explicit or tacit knowledge of quantum mechanics)? The common ground of philosophical discourse – it could be argued – is, and ought to be simply our semantic competence: there is no reason to load philosophical discourse with all sorts of obscure, poorly understood, and often controversial presuppositions.

However, such a *prima facie* reasonable objection clashes with a now long tradition of philosophical arguments challenging the distinction between semantic competence and the acceptance of theories. An early and crucial episode in that tradition was Quine's own article *Two Dogmas of Empiricism*, with the criticism of the analytic/synthetic distinction and the connected claim that there cannot be any principled reason to exclude any statement from counting for or against the truth of any other statement ("confirmation holism"). If one goes along with Quine's rejection of the analytic/synthetic distinction, then semantic competence cannot be identified with knowledge of certain truths *as opposed to* full-fledged scientific knowledge. Notice, however, that such is the case only if semantic competence is identified with some kind of propositional knowledge to begin with. Wittgenstein, for one, did not see the matter along such lines at all: for him, semantic competence was rather to be equated with a practical ability, the command of certain rules and techniques. Carnap, on the other hand, had interpreted semantic competence

in terms of knowledge of meaning postulates (plus logic): this is the conception of semantic competence that Quine is challenging by his criticism of the analytic/synthetic distinction. I.e., Quine shows – if he is right in his criticism – that semantic competence *à la* Carnap cannot be demarcated from general knowledge. His criticism is not immediately effective against a different conception of semantic competence, such as Wittgenstein's. But on the other hand, we saw that Quine himself doubted that the command of rules could plausibly be *contrasted with* the acceptance of certain propositions as true⁶: for him, adhering to certain rules must consist, ultimately, in taking certain propositions to be true. So Wittgenstein would have seen no reason to regard Quine's criticism of the analytic/synthetic distinction as a challenge to the opposition of semantic competence and factual knowledge (including, of course, scientific knowledge), while Quine, in turn, would not regard Wittgenstein's notion of competence as safe from his criticism.

Thus, according to Quine, we cannot easily identify the shared ground of philosophical argument with common semantic competence *as opposed to* more or less controversial scientific theories. It then becomes more plausible to hold that, as we are bound to be involved with all sorts of factual assumptions anyway, we might as well buy the whole lot, i.e., science to the latest detail of quantum mechanics. Such a conclusion is reinforced by the second main point of *Two Dogmas*, i.e., confirmation holism. If any statement can be relevant to the confirmation or disconfirmation of any other (at least in principle), it follows that scientific statements can be relevant to philosophical arguments. It is, of course, assumed that *there are* philosophical arguments; more precisely, it is presupposed that philosophical research aims at establishing theses. If there are philosophical theses that are up for confirmation or disconfirmation, then confirmation holism instructs us not to rule out any statement – not even quantum-mechanical statements – as possibly relevant.

However, as is well known, this is not how Wittgenstein saw the matter.

In philosophy we do not draw conclusions – he wrote in the *Investigations* – “But it must be like this!” is not a philosophical proposition. Philosophy only states what everyone admits. (PI, §599)

It is not so much that there are no philosophical theses; it's rather that there are no *controversial* philosophical theses, theses that one could think of giving

⁶ See fn.7 above.

grounds for by appealing to this or that fact, perhaps to this or that scientific result. Whenever something controversial is stated in philosophy, that is a sure sign that something went amiss in the philosopher's work. Now, there is the temptation not to take such pronouncements of Wittgenstein's seriously; one is tempted to say that such a contention cannot be upheld, and that Wittgenstein's own philosophical work does not bear it out.⁷ I believe the temptation should be resisted, whether or not we eventually agree with Wittgenstein on this, and whether or not Wittgenstein himself actually stood by his tenet. Wittgenstein's controversial thesis that there are no controversial theses in philosophy is of a piece with much else in his philosophy, for example, with his adoption of the "morphological method".⁸ And if we take him seriously on this, then Quine's continuity argument based on confirmation holism is devoid of any efficacy, from Wittgenstein's standpoint.

3.

We found in Quine two arguments for the continuity of philosophy with science. The first is based on the impossibility of cosmic exile and the rejection of the analytic/synthetic distinction: it is only plausible, at least *prima facie*, if semantic competence is a form of propositional knowledge. The second derives from confirmation holism, and it requires that philosophy be conceived as an argumentative discipline, in which theses are put forth and accepted or rejected depending on the evidence. Neither presupposition was acceptable to Wittgenstein, so this is, in a sense, the end of the story. However, there are three more points I would like to raise.

⁷ See Glock, 1996, p. 294: «This picture seems to impoverish philosophy, and is generally considered to be the weakest part of Wittgenstein's later work – slogans unsupported by argument and belied by his own "theory construction"». Glock goes on to argue that such methodological views are, however, «inextricably interwoven with the other parts of his work» (1996, pp. 294–295), a point with which I fully agree.

⁸ Wittgenstein never gave up the *Tractatus* insight that philosophy is an activity, not a doctrine (4.112), though he regarded the book as partly unfaithful to it. Even in later years, he wanted «to replace wild conjectures and explanations by quiet weighing of linguistic facts» (Z, §447; see BT, p. 432). The "morphological method" was his way of generating clarification without undertaking theoretical commitments: «I lay down the games as such, and let them spread their clarifying effect upon the several problems» (BT, p. 202). On these issues see Andronico, 1998, ch.2, and Marconi, 1997, pp. 89–95.

First point. It is quite clear that Wittgenstein shared some of the premises of Quine's first argument. There is a sense of "first philosophy" in which Wittgenstein, like Quine, does not believe in first philosophy: he does not believe in what he calls "metaphilosophy", or philosophy before philosophy.⁹ Like Quine, Wittgenstein does not believe that philosophy could start by shaping its tools – concepts such as "rule", "proposition", or "language" – in some pre-theoretical or meta-theoretical space ("cosmic exile", in Quine's terminology). On the contrary, philosophy essentially takes the ordinary use of such concepts for granted: that ordinary usage has no precise boundaries does not make such concepts less viable for philosophy than they are for everyday life (PG, §73). Like Quine, Wittgenstein regards ordinary language as the background of philosophical discourse: the concepts that philosophy employs are ordinary words in their ordinary use: «When I talk about language (words, sentences, etc.) I must speak the language of every day», he says (PI, §120). Where, then, do they part company exactly? Or, rather, why does Quine want to include science – indeed, the whole of science – in what he calls "our conceptual scheme", whereas Wittgenstein will have none of that? Why are the results of science, i.e., scientific propositions, of no special interest for philosophy according to Wittgenstein, although he admits that ours is «a community which is bound together by science and education» (OC, §298)? We already saw Quine's reasons to some extent. Wittgenstein's reasons are to be found in his definition of philosophy as a grammatical enterprise. Most of the time, science is for Wittgenstein just a collection of factual hypotheses that have no grammatical import, and *therefore* are of no interest for philosophy. More precisely, their grammatical import is independent of their truth or falsity: whatever grammatical import a scientific statement may possess is shared by its negation. This is one consequence of philosophy's «transition from the question of truth to the question of meaning».¹⁰ Obviously, this presupposes exactly the sort of distinction between the conceptual and the factual that Quine denies.

Thus, even Quine's and Wittgenstein's agreement on ordinary language as the background of philosophy is deceptive to some extent. For Quine, acquiescing in ordinary language¹¹ does not involve ordinary *concepts* more

⁹ BT, p. 67, PG, §72d; cf. PI, §121.

¹⁰ MS 106 46, quoted in Glock, 1996, p. 294.

¹¹ «Acquiescing in our mother tongue»(Quine, 1969a, p. 49).

than, or as opposed to, ordinary *truths* (moreover, there is no sharp line to be drawn between such ordinary truths and less ordinary truths, such as scientific truths). For Wittgenstein, to start with everyday language is to start with our customary use of ordinary words in everyday life (indeed, as part of everyday life): it is not to start with some body of common sense knowledge – such as could be represented by G.E. Moore’s truisms – for which the question could arise of its continuity, or discontinuity, with scientific knowledge.

Aside from not accepting Quine’s argument for the continuity of philosophy with science, does Wittgenstein have an argument against it? This is the second point I would like to raise. The answer is that he does have such an argument. It is based on the principle that “Nothing is hidden”¹²: the “data” that philosophy needs are all under our eyes. In philosophy we never need to *wait* until certain facts are established; there are neither discoveries nor surprises in philosophy. Wittgenstein says,

What is hidden is of no interest to us. One might give the name “philosophy” to what is possible *before* all new discoveries and inventions. (PI, §126)

If scientific results were relevant to philosophy, then there could be discoveries in philosophy, or something in philosophy could depend on a discovery: something in philosophy could be one way or the other depending on whether science has established, or discovered, that things are thus and so rather than otherwise. However, it can never be crucial for philosophy that *facts* are one way rather than the other, for, as we already saw, philosophy deals with possibilities, not with facts; its investigations are grammatical, not factual. As he went back to the *Tractatus* in the early Thirties, Wittgenstein denounced the “dogmatism” of his former theory of elementary propositions and logical analysis precisely because it made logic dependent on the discovery of certain facts; in that case, facts concerning the form of elementary propositions.¹³ The logical notions of analysis and elementary proposition had to wait for their full determination until “further research” had determined what the bottom level of reality was, and, consequently, what an elementary proposition looked like. According to the later Wittgenstein, this stemmed from a total misunderstanding of the nature of philosophical research:

¹² WWK, p. 183; BT, §89; PI, §126.

¹³ See Marconi (1995).

The truth of the matter is that we have already got everything, and we have got it actually *present*; we need not wait for anything. We make our moves in the realm of the grammar of our ordinary language, and this grammar is already there. Thus we have already got everything and need not wait for the future (WWK, p. 183)

The conceptual domain – the “realm of grammar” – is not something that science could make discoveries about, for two reasons: first of all, because it is entirely open to view, so that it is not something that one could think of making discoveries about (“Nothing is hidden”); secondly, because science only discovers *facts*, and facts – their being one way or the other – are grammatically indifferent. The only philosophical use of scientific discoveries is to make the philosopher better aware of possibilities:

Is scientific progress useful for philosophy? Certainly. The realities that are discovered lighten the philosopher’s task, imagining possibilities (LWPP I, §807)[Variant: Realities are so many possibilities for the philosopher].

Thus, even Wittgenstein’s argument against the continuity of philosophy with science ultimately depends on the dichotomy between the conceptual and the factual: it’s because philosophy is confined to the conceptual that, as far as philosophy is concerned, “nothing is hidden”.

4.

But then, are facts – their being one way or the other – really indifferent for grammar, hence for philosophy? This is the third and last point I would like to raise. The very late Wittgenstein – the author of *On Certainty* – appears to have had occasional doubts about the philosophical irrelevance of facts. It is sometimes pointed out that, in the notes *On Certainty*, certain facts acquire grammatical import, at least in the sense that they are assumed or presupposed by a language game, so that e.g., calling them in question is not really compatible with playing that particular game. It is perhaps not entirely clear whether the facts themselves are regarded as preconditions of the language game (OC, §618), or our certainty that such facts hold (OC, §§446, 519, 579); but anyway, Wittgenstein appears to be saying that certain facts, as laid out e.g., by physics (OC, §600) or anatomy (OC, §666), play a special role – a role that is close to that of a rule. Propositions expressing such facts – we are

tempted to call them “basic” propositions – are like rules in that they are as well-founded as any grounds one could give for them (OC, §111), and also in that they cannot be given up «without giving up all judgment» (or so one would be inclined to say)(OC, §494). Perhaps, Wittgenstein says, there is no sharp boundary between propositions of logic and empirical propositions (OC, §319); perhaps «the same proposition may get treated at one time as something to test by experience, at another as a rule of testing» (OC, §98).

All this is quite well known. And the conclusion is easily drawn that the very late Wittgenstein was indeed relaxing the distinction between the conceptual and the factual, between rules and propositions, or between grammar and experience, thereby coming closer to views such as Quine’s (or Davidson’s). Given enough time, he would have been brought to regard much of natural science as having grammatical import, hence to the continuity of philosophy with science.

As a speculation concerning the possible evolution of Wittgenstein’s thought, this is of course untestable. However, as an interpretation of what Wittgenstein, even very late, did actually say it is, I believe, one-sided and misses at least one important point. Wittgenstein is not saying that the facts of nature – “facts of (our) natural history”, as he calls them – are constitutive of concepts; he is saying that they motivate *our* particular use of certain concepts within particular language games. Let us read once more a very famous text in the *Investigations*, Part II:

I am not saying: if such-and-such facts of nature were different people would have different concepts (in the sense of a hypothesis). But: if anyone believes that certain concepts are absolutely the correct ones, and that having different ones would mean not realizing something that we realize – then let him imagine certain very general facts of nature to be different from what we are used to, and the formation of concepts different from the usual ones will become intelligible to him. (PI, II, xii).

That footrules are rigid, for example, does not *determine* our concept of measurement, but if they were not rigid a different notion of measurement would probably prevail. Here, the important point is that it would still be a notion of *measurement*, though different from ours. When Wittgenstein says that «we can also invent fictitious natural history for our purposes» (PI, II, p. xii), he appears to be suggesting that what is important for philosophy are the several possibilities of employment of certain words and the circumstances in

which such employments could turn out to be motivated and “natural”; *not* which of such circumstances do hold, which natural history is the true one. What contributes to the clarification of a concept such as *pain*, for example, is an examination of the different uses the word ‘pain’ can be put to for different purposes or in widely different circumstances, both natural and social; not, in and of itself, the association of our use of ‘pain’ with our physical and psychical constitution. Once again, philosophy is concerned with possibilities, not with actualities: the actual facts of nature – even “very general” facts – are not in themselves philosophically crucial. Reference to facts of nature does not settle philosophical disputes: it is a heuristic device whose purpose and effect is to make us realize the *contingency* of even the deepest features of our use of language. This makes science *useful for* philosophy – as the quotation about scientific discoveries clearly shows – but not *continuous with* philosophy, in the sense that scientific results could be premises to philosophical conclusions. Science stimulates philosophical fantasy, it does not establish, or help establish, philosophical conclusions (there are no such things, anyway).

Here, an objection could be raised against Wittgenstein. If philosophy is essentially interested in our own use of language (for, after all, that is where the philosophical malady is generated) then it would seem to be philosophically crucial that *one particular* natural history is true, rather than another – for example, one of the imaginary histories that Wittgenstein is fond of telling. For, when all comparisons and contrasts have been set up and duly experienced in imagination, it is after all in the light of the facts of *our* natural history that we make sense of our use of language. Suppose we were utterly ignorant of such facts: suppose we didn’t know whether footrules are rigid or not; or whether people usually remember their names (or only occasionally, or never); or whether physical bodies keep disappearing and reappearing rather than just being there most of the time. There may be something – perhaps a lot – that we could say about language under such a veil of ignorance, but we could hardly make sense of *our* use of language. Not knowing whether footrules are rigid or not, for example, we would entirely miss the *point* of our use of concepts of measurement. Counting the way we count in a world of stable objects is one thing, counting in the same way in a world of vanishing objects is a different thing. And so forth. So it seems that the facts being one way rather than another does make a difference, if philosophy is intended to make sense of our use of language.

Wittgenstein might have conceded this point; however, he would have argued that the rigidity of rulers, or the relative permanence of everyday objects, can hardly be seen as facts that science establishes; rather, science itself *presupposes* such “facts”. Therefore, their putative philosophical relevance does not involve the philosophical relevance of *science* – of scientific propositions, or of the facts such propositions are meant to establish. Quine, in turn, would point out that what we have here is just one more difference of degree: it is not easy to separate the facts that science (as a whole) presupposes from the facts that science establishes. Wittgenstein, on the other hand¹⁴, saw the difference between the bed of the river and the water flowing in it as one of kind, not of degree (OC, §§97, 99). So, once more, what is in question is the distinction between two kinds of propositions, whatever the two kinds are called.

REFERENCES

- Andronico, M. (1998). *Antropologia e Metodo Morfologico. Studio su Wittgenstein*. Napoli: La Città del Sole.
- Carnap, R. (1950). Empiricism, Semantics, and Ontology. *Revue internationale de philosophie*, 4, 20–40. Later in: R. Carnap (1956). *Meaning and Necessity: A study in semantics and modal logic*. Chicago and London: University of Chicago Press, 205–221.
- Gibson, R. (1996). Quine, Wittgenstein and Holism. In R.L. Arrington, & H.-J. Glock (Eds.), *Wittgenstein & Quine*. London and New York: Routledge, 80–96.
- Glock, H.-J. (1996). *A Wittgenstein Dictionary*. Oxford: Blackwell.
- Hacker, P.M.S. (1996). *Wittgenstein's Place in Twentieth Century Analytic Philosophy*. Oxford: Blackwell.
- Marconi, D. (1995). Predicate Logic in Wittgenstein's *Tractatus*. *Logique et Analyse*, 38(150), 179–190.

¹⁴ As Roger Gibson recently noted: Gibson, 1996, p. 93.

- Marconi, D. (1997). Transizione. In D. Marconi (Ed.), *Wittgenstein*. Roma-Bari: Laterza, 59–101.
- Putnam, H. (1992). *Renewing Philosophy*. Cambridge, MA: Harvard University Press.
- Quine, W.V.O. (1936). Truth by Convention. In O.H. Lee (Ed.), *Philosophical Essays for A.N.Whitehead*. New York and London: Longmans, 90–124. Later in: W.V.O. Quine (1966). *The Ways of Paradox, and Other Essays*. New York: Random House, 70–99]
- Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Quine, W.V.O. (1969a). *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Quine, W.V.O. (1969b). Natural Kinds. In Quine 1969a, 114–138.
- Stroud, B. (1995). Quine on Exile and Acquiescence. In P. Leonardi, & M. Santambrogio (Eds.), *On Quine*. Cambridge: Cambridge University Press, 37–52.
- (T) Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.
- (WWK) Wittgenstein, L. (1967a). *Wittgenstein und der Wiener Kreis*. Conversations recorded by F.Waismann; ed. by B.F.McGuinness, Oxford: Blackwell.
- (Z) Wittgenstein, L. (1967b). *Zettel*. Ed. by G.E.M.Anscombe & G.H. Von Wright. Oxford: Blackwell.
- (PG) Wittgenstein, L. (1969). *Philosophical Grammar*. Trans. by A. Kenny; ed. by R.Rhees. Oxford: Blackwell.
- (OC) Wittgenstein, L. (1974). *On Certainty*. Trans. by G.E.M. Anscombe and D. Paul; ed. by G.E.M.Anscombe and G.H.Von Wright. Oxford: Blackwell.
- (LWPP1) Wittgenstein, L. (1982). *Last Writings on the Philosophy of Psychology*, vol.1. Trans. by C.G. Luckhardt and M.A.E. Aue; ed. by G.H. von Wright and H. Nyman. Oxford: Blackwell.

- Wittgenstein, L. (1998). *Wittgenstein's Nachlass. The Bergen Electronic Edition*. Oxford and New York: Oxford University Press.
- (BT) Wittgenstein, L. (2005). *The Big Typescript* [=TS 213]. Oxford: Blackwell.
- (PI) Wittgenstein, L. (2009). *Philosophical Investigations*, 4th edition. Trans. and ed. by P.M.S. Hacker and J. Schulte. Oxford: Blackwell.

Commentary
The Bounds of Sense^{*}

Peter F. Strawson
Methuen, London, 1966

Massimo Marraffa[†]
marraffa@uniroma3.it

In the English-speaking philosophy of the second half of the 20th century, Peter F. Strawson's *Individuals* and *The Bounds of Sense* mark a return to metaphysics following the attacks on it by the logical positivists. In so doing, Strawson's two masterly books played a complex mediating role between analytical philosophy and Kantian criticism. In particular, the Strawsonian defences of the concepts of material body and person – often in association with the Wittgensteinian arguments against the possibility of a private language and for the necessary intersubjectivity of concept-application – were hailed by many analytical philosophers as new paradigms of a transcendental method of arguing against epistemological scepticism.

An insightful historical sketch of the reception of Strawson's neo-Kantianism in analytical philosophy is in Rorty (1971). In an earlier phase of analytical philosophy, a standard reply to the sceptic envisaging the possibility that material objects or other persons merely are the content of her own consciousness (her own representations) had been the phenomenalistic, "if-you-can't-beat-her-join-her" strategy consisting in saying that to have such-and-such representations just *was* to be seeing a material object or another person. But the meaning of that "was" turned out to be elusive. Furthermore, such an anti-sceptical strategy had an unpleasant air of idealism about it, though it was firmly claimed to have a "logical" as opposed to "metaphysical" nature. It's no surprise, then, that «when, on the heels of Austin's attack on "the ontology of the sensible manifold", Strawson revived the distinctively Kantian anti-idealist thesis that "inner experience requires outer experience", the shift in strategy was welcomed» (Rorty, 1971, p. 4). It seemed to be

^{*} I thank Carlo Gabbani, Matteo Morganti and Alfredo Paternoster for helpful comments on earlier drafts of this article.

[†] Università degli Studi Roma Tre, Rome, Italy.

possible then to combine Kant's Transcendental Deduction and Wittgenstein's private language arguments into one single argument against the legitimacy of the sceptical challenge of justifying our belief in the external world basing on the private data of individual consciousness. Some philosophers (e.g., Harrison, 1974; Wilkerson, 1976; Stevenson, 1982) began to hope that this "analytic Kantianism"¹ would be successful where Kant had failed, namely in truly transcending classic empiricism, with all its obsession for the private contents of experience, and thus orienting analytical philosophy towards a metaphysics of experience, provable *a priori* by *transcendental arguments*.

In *The Bounds of Sense* Strawson develops his distinctly analytic interpretation, defence, and elaboration of Kant's ideas (his "analytic Kantianism") by trying to disentangle a genuine metaphysics of experience from the allegedly incoherent framework of transcendental psychology (the theory of synthesis)². More precisely, Strawson develops "austere" versions of six Kantian theses about objectivity, space, the unity of space and time, substance and causation (1966, p. 24). These theses can be seen as the core of the Kantian "descriptive" metaphysics, as opposed to the "revisionary" metaphysics of transcendental idealism. Thus Kant's first Critique is reconstructed as an analysis of the concept of possible experience that hinges on the Transcendental Deduction (an analytical argument aimed to prove the "objectivity thesis"³). Such a reconstruction selectively includes themes from the Aesthetic, the Refutation of Idealism added in B, and the Analogies of experience.

In the course of years various objections have been raised against Strawson's reconstructive project in *The Bounds of Sense*. In what follows the focus will be only on two criticisms: that the various Kantian transcendental arguments reconstructed by Strawson could only ever establish that we must

¹ The term was coined by Glock (2003).

² As Strawson himself puts it, *The Bounds of Sense* was a «somewhat a historical attempt to recruit Kant to the ranks of the analytical metaphysicians, while discarding those metaphysical elements that refused any such absorption» (Strawson, 2003).

³ The objectivity thesis – according to which «experience must include awareness of objects which are distinguishable from experiences of them in the sense that judgements about these objects are judgements about what is the case irrespective of the actual occurrence of particular subjective experiences of them» (Strawson, 1966, p. 24) – is a reconstruction of Kant's assertion that the Transcendental Deduction aims to show how «*subjektive Bedingungen des Denkens sollten objektive Gültigkeit haben*» (A89/B122).

believe certain things to be the case, not that they are the case; and that Strawson wrongly dismisses transcendental idealism.

The definition of a transcendental argument (TA) is often drawn from a passage in *Individuals* (1959, p. 35) that can be paraphrased as follows: a TA is an attempt to show that the sceptic who doubts the existence of material objects and other minds is self-defeating since her doubts amount to the rejection of some of the necessary conditions of the conceptual framework within which alone such doubts make sense.

Confronting with arguments conforming to this format, some authors (Thomson, 1964; Stroud, 1968; Gram, 1974; Walker, 1978) made an objection that can be traced in Schulze's *Aenesidemus*. The objection is that a TA can at most establish statements on what we must think or believe there is in the world, not statements on what there is. Let us suppose that Strawson succeeded in developing a TA in favor of our conception of the world as a spatiotemporal system of objects conceived as distinct from the temporal series of our experiences of them. He would prove, then, that the sceptic, as everyone else, must remain within that conceptual framework, i.e., must apply the categorial concepts of the framework, and believe the propositions in which such concepts are used; or else she would be forced to stop thinking and speaking in an intelligible way. In so doing, however, Strawson would not prove that it is true – not even that it is possible to know that it is true – that there are objects that can exist independently from the framework and have properties conforming to its categorial features. This means that the sceptic can concede the *subjective necessity* of the framework while still insisting that its *objective validity* cannot be established. But *this* – the objectivity thesis – is what a TA is supposed to prove if it must be faithful to its anti-sceptical aim. Consequently, the critics argue, Strawson (and any “transcendentalist”) will be obliged to introduce in his TA a Verification Principle – a principle to the effect that one cannot apply a concept, or have a belief, without the availability of epistemic procedures that permit one to establish if there are cases instantiating the concept, or to ascertain the belief's truth-value; a principle that grounds the possibility of knowing that the reality (the *Ding an sich*) conforms to our categorial apparatus. But then, it could be that Oxford in the 1960s is not so far from «Vienna in the 1920s» (Stroud, 1968, p. 256).

The verificationist deadlock in which Strawson's analytic Kantianism seems to be trapped can raise the suspicion that he has got things the wrong way

round in putting all his faith in the use of TAs and totally rejecting transcendental idealism. Let us see why.

To refute the sceptic Strawson must prove the objectivity thesis, according to which there cannot be something that conforms to the subjective conditions of experience without being at least in part experience of objects that exist independently of the cognizing subject. But – here is the point – the sceptic doubts within a metaphysical paradigm that Kant called “transcendental realism”, where what is real is supposed to be defined quite independently from our discursive cognitive structures (the pure categories) and our forms of sensible intuition (space and time). In other terms, the transcendental realist uses non-epistemic notions of object and truth: the object subsists in itself, out and independently of the cognitive relationship; truth is correspondence to the thing-in-itself, and thus implies that a sentence could be true even if it was not possible, not even in principle, to come to know that it is true.⁴ Within the paradigm of transcendental realism, then, the sceptic (the “empirical idealist” in Kant’s idiom) cannot be refuted by either transcendentalism or verificationism.

Thus, a refutation of scepticism can pass through the delegitimization of the non-epistemic notion of objectivity on which it rests. This is just what Kant tried to do with the doctrine of transcendental idealism. The Kantian object is not the transcendental realist’s object. It is an epistemic product of the use of the categories. The categorial framework which the Metaphysical Deduction has tried to establish as a necessary presupposition of all possible interpretations of experience is not a necessary subjective apparatus which is then applied to a domain of non-epistemic transcendent objects by virtue of some kind of ontological guarantee or pre-established harmony (Genova, 1984, p. 493). Kant himself warns against such a misunderstanding of his thought by defining his transcendental idealism as «ein System der Epigenesis

⁴ In this formulation the reader will see the traces of how, in the early 1980s, Hilary Putnam revived the Kantian distinction between empirical realism and transcendental realism. In *Reason, Truth, and History* (1981) he endorses the former (= *internal* realism) and rejects the latter (= *metaphysical* realism). According to Putnam, Kant was the first philosopher to assume an *internalist* perspective, within which objectivity has an unbreakable bond with human cognitive activity. This perspective is not far from Michael Dummett’s antirealism. Whereas the metaphysical realist presupposes a God’s Eye point of view to survey the world as it really is, Kant (the transcendental idealist/empirical realist), Putnam (the internal realist), but also Dummett (the antirealist) hold that the only available point of view is that of a human being as a finite and natural entity. In this “secularization” of the philosopher’s point of view lies the gist of what Putnam calls “internalism”.

der reinen Vernunft» (B167). In this perspective, the object of knowledge is not something that is given irrespective of the human cognitive activity. What is given are only *Vorstellungen*, which must be submitted to the epistemological test as to whether or not they can be “referred to an object”, i.e., be connected with each other in certain ways, in accordance with the rules derived from the schematization of the categorial concepts of the framework. The object is here «an epigenetic product of the use of intelligence» (Genova, 1984, p. 493; see also Genova, 1974). And any other notion of object, including that of a *Ding an sich*, will be derived from the epistemic notion of object that is constituted by the categorial framework.

The moral is, then, that Strawson has actually got things the wrong way round in putting all his faith in the use of TAs and totally rejecting transcendental idealism. Kant’s TAs are underpinned by his transcendental idealism – otherwise they would fail. No analytic argument can defeat scepticism without first of all challenging its underlying non-epistemic notion of objectivity. One can definitely think that Kant’s transcendental idealism failed to offer a notion of objectivity that is able to supersede that of the sceptic/transcendental realist.⁵ But then, it will be necessary to devise a more plausible alternative to the non-epistemic objectivity, which Strawson does not do in *The Bounds of Sense*.

REFERENCES

- Genova, A.C. (1974). Kant’s Epigenesis of Pure Reason. *Kant-Studien*, 65(3), 259–273.
- Genova, A.C. (1984). Good Transcendental Arguments. *Kant-Studien*, 75(4), 469–495.
- Glock, H.-J. (2003). Strawson and Analytic Kantianism. In Id. (Ed.), *Strawson and Kant*. Oxford: Oxford University Press.
- Gram, M. S. (1974). Must Transcendental Arguments be Spurious? *Kant Studien*, 65, 304–317.

⁵ For example, Stroud (1984, ch. 4) argues that Kant puts forward only a “transcendental” version of scepticism. From an empirical point of view, the world of phenomena is real and empirical idealism is false; from a transcendental point of view, the world of appearances is ideal and transcendental idealism is true.

- Harrison, R. (1974). *On What There Must Be*. Oxford: Clarendon Press.
- Kant, I. (1781-1787/1956). *Kritik der reinen Vernunft*. Hamburg: Felix Meiner.
- Putnam, H. (1981). *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Rorty, R. (1971). Verificationism and Transcendental Arguments. *Noûs*, 5(1), 3–14.
- Stevenson, L. (1982). *The Metaphysics of Experience*. Oxford: Clarendon Press.
- Strawson, P.F. (1959). *Individuals*. London: Methuen.
- Strawson, P.F. (1966). *The Bounds of Sense*. London: Methuen.
- Strawson, P.F. (2003). *A Bit of Intellectual Autobiography*. In H.-J. Glock (Ed.), *Strawson and Kant*. Oxford: Oxford University Press, 7–14.
- Stroud, B. (1968). Transcendental Arguments. *The Journal of Philosophy*, 65, 241–256.
- Stroud, B. (1984). *The Significance of Philosophical Scepticism*. Oxford: Clarendon Press.
- Thomson, J. J. (1964). Private Languages. *American Philosophical Quarterly*, 1, 20–31.
- Walker, R.C.S. (1978). *Kant*. London: Routledge.
- Wilkerson, T.E. (1976). *Kant's Critique of Pure Reason*. Oxford: Oxford University Press.

Commentary

Scientific Explanation and the Causal Structure of the World

Wesley Salmon
Princeton University Press, Princeton, 1984

Raffaella Campaner *
raffaella.campaner@unibo.it

The time had come «to put the “cause” back into “because”» (Salmon, 1984, p. 96), and a theory of explanation had to be developed with an essentially *causal* character: it was together with his theory of explanation that Wesley Salmon developed a theory of probabilistic and mechanistic causation, outlined in detail in *Scientific Explanation and the Causal Structure of the World* (1984). His conception of causation aims to shed light on scientific explanation as a way to open the “black box of nature” and reveal its inner workings, and has ever since become a locus for anyone interested in dealing with causal explanation and putting forward a mechanistic notion of cause.

Salmon’s reflections stem from an interest in explanation and from radical criticism of and a number of counterexamples – well-known in the literature – to the “received view” of explanation. That is where the volume starts. Salmon’s Statistical-Relevance model (S-R), elaborated back in the early Seventies, is meant to overcome the limits of the Hempelian models, and, more specifically, the Inductive-Statistical model (I-S). The inferential form of explanation, the requirement of high inductive probability, the epistemic relativization of statistical explanation are strongly opposed, insofar as they are regarded as both inadequate to represent genuine explanations and implicitly committed to a deterministic view of the world. According to S-R, to explain an event is to identify all and only the factors that are statistically relevant to its occurrence, where a factor C is taken to be statistically relevant to the occurrence of an event B under circumstances A if and only if $P(B | A \cdot C) \neq P(B | A)$. The initial reference class is to be partitioned until a *homogeneous* reference class is obtained, i.e., a class that cannot be further partitioned by

* Department of Philosophy, University of Bologna, Italy

means of statistically relevant factors and does not include any irrelevant factor. The explanation is obtained by assigning the event to be explained to the appropriate final reference class.

While acknowledging epistemically and pragmatically homogeneous reference classes as well, Salmon's proposal is centered on the notion of *objectively* homogeneous classes, to warrant *genuine* statistical explanations, independent of the knowledge situation. The S-R model conveys explanatory information by providing relevant partitions and allowing prior and posterior probabilities to be compared. Statistical relevance, either positive or negative, is all that is needed in this respect, with no requirement whatsoever on the final probability value of the event to be explained. All events, highly probable as well as highly improbable ones, can be explained in the very same way, within a perspective that, without committing itself exclusively to either a deterministic or an indeterministic view, is compatible with both. The rebuttal of Hempel's position is accompanied by the proposal of an *ontic* conception of scientific explanation, capable of placing the events into networks of objective relations. Opponent approaches are described: the epistemic conception (subdivided into inferential, information-theoretic and erotetic) and the modal; the epistemic conception is held to be grounded on nomic expectability, the modal conception on nomological necessity. Whereas these views are judged inadequate to deal with an indeterministic perspective and unable to grasp the difference between explanation and description, the ontic conception is claimed to fulfill both tasks.

Soon aware that explanatory relevance cannot amount to just statistical relevance, but must be traced back to causal relevance, Salmon commits himself to the elaboration of an objective concept of causation on empirical grounds. The S-R model is by no means eliminated, but comes to play the role of the first level in a two-tiered model of explanation which has causation at its core. The notion of causation intends to implement the S-R basis and unravel the network of productive links underlying phenomena, often operating in stochastic ways. A first decisive step in this direction is the proposal of a "process ontology", substituting the "entity ontology" characterizing other views of causation, such as Reichenbach's (by which – on the other hand – Salmon is largely inspired). The causal relationships that scientific explanation must capture are clarified by the notions of "causal process", "causal production" and "causal propagation". Instead of starting off with a single definition of mechanism, like other subsequent mechanistic accounts, Salmon

builds up his theory piecemeal on these notions. Processes are defined as spatio-temporal continuous entities, such as waves and material objects persisting through time. Causal processes are distinguished from so-called pseudo-processes through the criterion of mark transmission: a causal process is a process which is able to transmit a mark, i.e., a modification of its structure, from the point in which it is imposed onwards without further interventions. Unlike pseudo-processes, causal processes can transmit information, energy, structure, and, in sum, causal influence. The causal network underpinning phenomena presents a conjunctive fork, where causal processes arise from a common cause and common background conditions, and interactive forks, where two causal processes directly intersect and produce mutual change. In the latter case, when two causal processes are both persistently modified in the interaction, causal production takes place, giving rise to statistical distributions of results. Causal influence is then propagated along causal processes, thanks to their spatiotemporal continuity. Propagation occurs according to the Russellian “*at-at*” theory of motion, according to which «to move from A to B is simply to occupy the intervening points at the intervening instants. It consists in being at particular points of space at corresponding moments» (Salmon, 1984, p. 153). Statistical relations provide hints of causal relations; causal processes, causal interactions, and the causal laws describing them provide the mechanisms by which the world works. The statistical-relevance and the causal-mechanical levels are regarded as equally crucial for the unravelling of the “causal structure of the world”, and hence for an adequate account of scientific explanation. To explain a phenomenon is to locate it at some point within the net of causal processes: «these processes are the physical mechanisms that are responsible – probabilistically – for the phenomena we are trying to explain» (p. 123).

After the crisis of the notion of cause due to advances in XX century physics, Salmon was one of the leading proponents of its revival. His theory constituted a breakthrough in the debate on scientific explanation, providing a fundamental contribution to reflections on causation and explanation and emphasizing the distinction between *statistical causation* (corresponding to type causation) and *aleatory causation* (corresponding to token causation), both necessary for an adequate account of causation but conceptually distinct. Making his point more strongly, he later came to state:

I believe that there is no such thing as probabilistic causality in the strict sense, because the probability relations require supplementation by such physical

entities as processes and interactions. Reichenbach evidently regarded probabilistic structures and physical structures as distinct ways of approaching probabilistic causality. I believe that they need to be combined to yield a satisfactory concept of causality. I'd call it "physical (indeterministic) causality" rather than "probabilistic causality". (Salmon, 2010¹, p. 11)

In Salmon's view, a «satisfactory concept of causality» is reached by appealing to theoretical notions in the framework of an empiricist and realist perspective. A far-reaching debate was thus engaged with other authors, and especially with Bas van Fraassen, whose constructive empiricism and pragmatics of explanation he largely contrasts: «scientific experience provides strong support – Salmon holds – for the appeal to unobservable common causes and causal processes when observable domains do not furnish the required causal connections» (Salmon, 1984, p. 228). Consistently with his ontic view, he argues for the reality of unobservables, defended by appealing to the common cause principle and consistent results of different experiments (as, e.g., in the determination of Avogadro number by Jean Perrin in the early 1910s), and to a combination of causal and analogical reasoning (see Salmon, 1984, ch. 8).

Salmon intends to capture causation as it manifests itself contingently in our world. The counterfactual formulations of the criteria of mark transmission and causal interaction could thus constitute a threat to the empiricist approach he wants to embrace. A causal process is such that, *had* a modification of its structure *been performed*, it *would have transmitted* it from that point onwards; a causal interaction is such that, *had* two causal processes *intersected*, both their structures *would have been modified* from that point onwards. Salmon (1984) appeals to counterfactuals «with great philosophical regret» (1998, p. 18), and was glad to abandon them in the Nineties for Phil Dowe's "conserved-quantity theory" (see Dowe, 2000), which gets rid of counterfactuals. However, Salmon's intuitions on how to deal with counterfactuals are worth recalling, and constitute a – largely neglected – antecedent of some very recent views. In the last few years, both actual and *hypothetical* interventions have been increasingly recognized as playing a crucial part in the identification of mechanisms' components and functioning (e.g., Woodward, 2003; Glennan, 2002; Craver, 2007). The role of "interventionist counterfactuals" has been emphasized. The usefulness of counterfactuals to tell genuine causal relations from non-causal ones and the

¹ Published posthumously.

possibility of interpreting them experimentally were already pointed out by Salmon. He believes that science has a direct way of dealing with the kinds of counterfactual assertions required for causal assessments: the ability to transmit a mark and the property of being a causal interaction are assessed by performing certain kinds of experiments (see Salmon, 1984, pp. 147–149). This interpretation is presented as fully objective. Even though we often play an active role, «human agency plays no essential part in the characterization of causal processes or causal interactions» (1984, p. 174), given that they would be such even if no human agent were to perform the experiments. Hints towards the way in which both the interventionist and the neo-mechanist perspectives are currently incorporating counterfactuals can thus be found in the very place where the “anti-counterfactualist tradition” (Woodward, 2004) arose, namely in Salmon’s probabilistic mechanicism and his conceptual apparatus as elaborated in the Eighties.

What about the applicability of Salmon’s theory? Both in its original form and – even more – in the “conserved quantity” version, it has been criticized for not being widely applicable, and possibly adequate only with respect to physical and chemical causation. It has been accused of imposing too strong requirements (e.g., homogeneity and spatiotemporal continuity of processes), of providing just some sort of abstract geometrical network of processes and interactions, adaptable only to idealized or very simple cases, and lacking of indications on how to identify the explanatorily relevant causal processes and interactions. As a matter of fact, the examples provided in (1984) cover an extremely broad set of phenomena, of both a commonsensical, everyday sort, and strictly scientific. They range from bacterial infections to food intoxication, from radioactive decay to delinquency acts, from the collision of billiard balls to the presence of a worked bone in an archaeological site. When actual science is referred to, however, physics is no doubt what Salmon has mainly in mind, and this also strongly affects his attitude towards the relationship between general and singular causation, which is seen as unproblematic. While no disciplinary restriction is drawn, quantum physics remains highly puzzling for Salmon’s view, which admittedly does not fit quantum phenomena (1984, pp. 247–259).

Whereas Salmon primarily questions *what* causal processes and interactions *are*, a major concern of neo-mechanists in the last decade has had to do with what mechanistic reasoning is *good for*, with a more in-depth focus on the disciplinary fields in which mechanist notions can be implemented, and

the purposes for which mechanistic models are employed. Aiming at capturing the actual use of causal notions, and especially interested in such fields as biology, medicine, cognitive science, economics, the recent mechanistic approach stresses the dynamic character of mechanisms, and their being complex, multilevel structures, whose overall behaviour strongly relies on the internal organization of component parts and that can be structurally and/or functionally decomposable. These features do not play a part in Salmon's view, and this can be undoubtedly regarded as one of its main limits. Nevertheless, Salmon's view has been recognized as capable of grasping causation in some of the fields which are the very target of neo-mechanist views, such as biology and medicine (Schaffner, 1993), epidemiology (Vineis, 2000), economics (Mäki, 1992). What seems to be really missing in Salmon's account is a two-level example of scientific explanation starting from the very question of explanation and the identification of the initial reference class, through the relevant partitions and the homogeneous reference class, up to the relevant net of processes and interactions underlying the phenomenon to be explained. If in principle Salmon insists on the complementarity of the statistical and the causal level, no instance of the complete construction of a two-tiered explanation is provided.

While definitely advancing an ontic, objectivist perspective on causal explanation, Salmon's last chapter also recalls contextual aspects, especially by referring to Peter Railton's position, thus anticipating some of the latest trends. Most recent mechanistic literature emphasizes the relationship between the level of graininess of a mechanistic description and the context in which it is drawn, recognizes the possibility of elaborating mechanisms' sketches or schemas, and acknowledges that causal accounts exhibit some perspectival aspect. After starting mildly admitting of some context-dependent aspects of explanation in Salmon, 1984², Salmon soon afterwards (1989) came to suggest that a "new consensus" *with regard to scientific explanation* could eventually be built, which might show how the causal-mechanical, unificationist and pragmatic accounts could be compatible with and possibly complement each other. No consensus on scientific explanation has been reached, but Salmon's work both fruitfully casts light on how causation and explanation are intertwined and already presents many interesting hints towards what are now regarded as some of the crucial steps forward in dealing

² Remarks on contextual aspects will get bolder in Salmon (2002) and (2010).

with causation. Far from converging into some form of “consensus” of causation, such elements as the appeal to counterfactuals to be interpreted experimentally and some recognition of the role of the context are increasingly setting the ground for the interaction between different (both mechanistic and non-mechanistic) theories of causation, a trend the last Salmon might have been sympathetic with.

REFERENCES

- Craver, C. (2007). *Explaining the Brain*. New York: Oxford University Press.
- Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science* (supplement), 69(3), S342–S353.
- Mäki, U. (1992). The Market as an Isolated Causal Process: A Metaphysical Ground for Realism. In B. Caldwell, & S. Bohem (Eds.), *Austrian Economics: Tensions and New Developments*. Dordrecht: Kluwer, 35–59.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. (1998). *Causality and Explanation*. New York: Oxford University Press.
- Salmon, W. (2002). A Realistic Account of Causation. In M. Marsonet (Ed.), *The Problem of Realism*. Aldershot, UK: Ashgate, 106–134.
- Salmon, W. (2010). The Causal Structure of the World. *Metatheoria*, 1, 1–13.
- Schaffner, K. (1993). *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.
- Vincis, P. (2000). Exposure, Mutation and the History of Causality. *Journal of Epidemiology and Community Health*, 54, 652–653.
- Woodward, J. (2003). *Making Things Happen*. New York: Oxford University Press.

Woodward, J. (2004). Counterfactuals and Causal Explanation, *International Studies in the Philosophy of Science*, 18(1), 41–72.

Commentary

Putnam on the Fact/Value Dichotomy and the scientific conception of the world

Gabriele De Anna *
gabriele.deanna@uniud.it

The philosophical journey of Hilary Putnam has been long and complex. Several entrenched threads have conditioned each other and the whole, but a preferential thread to consider in order to have a sense of the entire process is the relation between the scientific and the manifest conceptions of the world. From this point of view, his journey can be seen as going through (at least) three phases.

In the first phase, Putnam endorsed metaphysical realism and gave priority to the scientific conception of the world: science tells us what reality is ultimately made of, and all the non-scientific areas of our discourse have to be reduced to (or at least can be proven to be reducible to) scientific discourse (Putnam, 1975). Functionalism serves this purpose for one of the areas of our language which is traditionally particularly difficult to express in scientific terms: language about the mind (Putnam, 1960/1975). In the second phase, Putnam turned to the thesis that the scientific and the manifest conceptions are on a pair. The role of language is not to hook onto a mind-independent world of ready-made facts, i.e., scientific facts; indeed, although we can say that both a snooker ball and an electron exist, the electron is not just a very small ball: in these two cases, existence is a very different thing.¹ The point is that language serves the purpose of allowing us to interact with each other and with the world, as Wittgenstein pointed out, according to a normatively regulated form of life which we acquire from our community throughout the process of language learning. (In Putnam's reading, though, this thesis does not jeopardize the notion of truth).² Scientific language may well be part of our form of life, but it is not any more "referential" than other areas of language. Our language does not take us out of our minds into a world of real facts, but

* University of Udine, Italy

¹ See, for example, Putnam (1988).

² See, for example, Putnam (1994).

craves a cookie-dough-like world in pieces for us (Putnam, 1981). In the third and last phase, Putnam seems to give priority to the manifest conception. Now, he endorses a “second naivety” about reality, i.e., a self-conscious direct realism. Like Aristotle, Austin, the second Wittgenstein, and – to some extent – American pragmatists and John McDowell, he suggests that our common sense objects are what primarily exists. They do not exist as mind-independent scientific facts, neither are they artifacts made by our conceptual capacities. They simply exist and no further specification is needed. The manifest conception is prior, in this perspective, in the sense that scientific theories (by developing scientific instruments and conceptual tools) enlarge the cognitive capacities through which we think of common sense reality: the latter is thus a precondition of, and thus prior to, scientific reality (Putnam, 2001). (Naturally, here I am employing Sellar’s language in a sense that is far from his usages, in the hope to capture the distinctive character of Putnam’s position).

Putnam’s emphasis on the manifest conception of the world at his third phase opens, among others, the problem of explaining the exact relation between the scientific and the manifest conceptions: does saying that the manifest conception is prior imply that it is more inclusive, and thus the scientific conception leaves something out? And, in case, what is left out by science? A good point of view from which to address these questions is Putnam’s 2002 book *The collapse of the Fact/Value Dichotomy and other Essays*. Values are traditionally considered to be outside the scope of science, and, thereby, they are sharply separated from facts. Considering their role in the outlook of Putnam’s “second naivety” can be a good way to check what cognitive role science plays in that outlook.

The book is a collection of pieces which were written for various occasions: the first three chapters originated from the Rosenthal Foundations lectures given by Putnam at Northwestern University Law School in 2000; the rest are lectures and essays written for various occasions, which relate to the main topic of the Rosenthal lectures. For our concerns, three main traits of the book should be recalled: the criticism of the fact/value dichotomy, the criticism of value-free science, and the rejection of an ontology of values.

Criticism of the fact/value dichotomy

Putnam underlines that Hume’s claim according to which normative statements cannot be inferred from statements concerning matters of fact, has

led neo-empiricists and philosophers influenced by them to accept a sharp distinction between facts and values. According to this view, facts are dealt with by science and referred to by observational language (Carnap being one of the most influential figures in this respect), whereas values constitute an altogether separate realm, which is dealt with in different ways by different authors. Some deny that they exist and take evaluative language to be always erroneous or senseless, others take them to be subjective states or responses and take evaluative language to express or describe subjective emotions. According to Putnam, though, «the fact/value dichotomy is, at bottom, not a *distinction* but a *thesis*, namely the thesis that ethics is not about matters of fact» (p. 19); it originated from the persuasion that ethical language has neither an observational nor a theoretical purpose, and thus must be expelled «from the domain of knowledge» (p. 20). The word “cruel” has different senses, and in virtue of some of these it can be used to describe, in virtue of others it can be used to evaluate. The post-humean view is that the two senses must be kept separated on a cognitive level: by knowing facts we do not acquire grounds on how to evaluate those facts. (At the very best, we can acquire knowledge on how to evaluate other facts: e.g., by knowing facts about A’s psychology, i.e., that A dislikes that p , we can learn about the use of evaluative language by A, i.e., that he would justifiably and sincerely claim “it is bad that p ”). By contrast, Putnam contends that scientific practices, in particular choices among empirically equivalent theories, depend on judgments of coherence, simplicity, economy, etc., «yet, simplicity, coherence and the like are values» (p. 142), even if neo-empiricist philosophers and their followers (including Quine and Popper) «were determined to shut their eyes to [this] fact» (p. 31). *Pace* Rorty and other deflators of science and cognition, though, judgments of value are “judgments of reasonableness” and are “objective”: this requires «rethinking the whole dogma (the last dogma of empiricism?) that facts are objective and values are subjective» (p. 145). The upshot is that there are genuine “thick ethical concepts” that represent a deep “entanglement of fact and value”. E.g., the concept expressed by the word “cruel” «has normative, indeed, and ethical uses» (p. 39), and «simply ignores the supposed fact/value dichotomy and cheerfully allows itself to be used sometimes for a normative purpose and sometimes as a descriptive term» (p. 35).

The criticism of value-free science

One of the main purposes of the book is to offer a philosophical backup to the approach to economics supported by Vivian Walsh and Amartya Sen, both of whom were colleagues of Putnam and have been his long-term collaborators. The success of the neo-empiricist conception of knowledge, Putnam shows, had its influence also in economics: supporters of the neo-classical view thought that the description of the rational agent in economic theory had to be totally value-free. As Walsh has pointed out, though, this was not Adam Smith's original position. Furthermore, as Sen has argued, that approach completely misrepresents the actual thinking of agents: agents always make choices on the ground of values that they see in the foreseeable outcomes of possible actions open ahead of to them. A theory which does not take values into account cannot represent the actual process of choices of people. Hence, both Walsh and Sen have been working for the development of a new, value-sensitive economic theory, and, in this process, they have often discussed the faults of the received view about the fact/value dichotomy. The upshot about science does not concern only economics, but all practical sciences, including, for example, political science. In a discussion with Habermas, Putnam argues, for example, that «ethical values can be rationally discussed» (p. 133), in the public sphere.

The rejection of a metaphysics of values

In this book, Putnam stresses that his objection is to a *dichotomy* between facts and values, not to a parallel *distinction*, the difference being that a dichotomy is inflated with metaphysical contents, whereas a distinction is on a purely linguistic level: «nothing metaphysical follows from the existence of a fact/value distinction in [his] sense» (p. 19). Putnam's points are, thus, mostly linguistic. He claims that language about values is as legitimate as language about facts, and that the fact that there are these two areas of language does not imply that there are two sharply distinguished groups of statements, so that for any statement it must be the case that it fits either in one group or in the other. Rather, the borders between the two groups are vague, and some statements (e.g., thick ethical statements) may be in an intersection. The fact that we accept the legitimacy of value-language does not imply, though, that we need to admit the existence of referents of value-terms akin to referents of empirical or

factual terms. This is very much in the spirit of his return to pragmatism and to the second Wittgenstein: according to these approaches, linguistic meaning does not have an essence in virtue of which having a meaning is the same thing for all terms and statements. Rather, what having a meaning amounts to for a statement or a term varies from utterance to utterance and depends on the role it plays in each actual occasion when it is used. So having a referent is not identical to having a meaning, and thus we can grant that value-terms and statements have meanings even if we do not commit ourselves to an ontology of values. Indeed, another book which Putnam published around the time of *The collapse of the fact/value dichotomy* is entitled *Ethics without ontology*: here, Putnam offers an analysis of value-language which avoids ontological commitments.

In a very sympathetic review of these two books by Putnam, Sabina Lovibond remarked that Putnam's points do not imply that «we ought to avoid any ontological commitment» (Lovibond, 2006, p. 460). It seems to me that Lovibond's point is well put. The fact that not all language needs to be referential and that we do not need to introduce an ontology of values that mimics scientific ontology in order to secure a meaning to value language, does not imply that we cannot (or not even that we need not to) ask ourselves what are the experiential conditions for the assertibility of evaluative statements. By "experiential conditions" I don't mean just *empirical conditions*, i.e., the conditions which we can accept if we assume an empiricist conception of experience; I mean also conditions concerning our evaluative responses to facts that we describe through our non-purely factual language. It is true that Putnam rules out the possibility of an ontology of values, but he means that in a very strict sense of "ontology". In that context, by "ontology" he means a discipline which originated from neo-empiricists, and which aims at giving an observer-free description of what there is, i.e., – as Putnam often puts it – a description of the world from a No-Eye point of view (see Putnam, 2004, p. 51). We can grant him that there might not be an ontology of values in that sense, and still insist that it is a legitimate question to ask what features the world (or even "man-in-the-world") should have in order for evaluative language to have the meaning it has, and thus what are the assertibility conditions of evaluative statements and terms. In senses different from Putnam's usage, but probably more traditional, this can be called "metaphysics" or "ontology". Putnam himself seems to be after a "metaphysics" in this sense. Indeed, he is sympathetic with Quine's

pragmatist-sounding claim that “the fabric of sentences”, which is our language, is not made of black threads (factual statements) and white threads (conventional statements), but of gray threads. He also regrets that Harvard pragmatists did not succeed in convincing Quine to take a step further and claim that some statements are gray also in the sense that they combine factual and evaluative meanings (2002, p. 138). If (at least some) statements describe and evaluate facts at the same time, it must be legitimate to ask what features of the realities they describe elicit the evaluations they state, and why; but this is a metaphysical question, in a wide, traditional sense of the word “metaphysical”. There is, however, a more fundamental reason why someone sympathetic to Putnam’s approach should be willing to be involved in metaphysics. He recognizes that most of the contemporary defenders of the fact/value dichotomy do not accept the old empiricist arguments he mostly dwells with, but embrace that position for other reasons:

today [the dichotomy] is defended more and more on metaphysical grounds. At the same time, even the defenders of the dichotomy concede that the old arguments for the dichotomy were bad arguments. The most common metaphysical ground is simply *physicalism* (Putnam, 2002, p. 40).

His main example is Bernard Williams, the criticism of whom is left to Walsh’s words against the possibility of a reductive account of agency in economics (2002, p. 42). If one wants to reject the dichotomy, then, one needs to show that an account of human action requires a reference to features of the world, and of the position of man in it, that cannot be given in purely physical terms. Again, this is metaphysics, albeit in a traditional and wide sense.

Putnam’s rejection of the fact/value dichotomy can suggest some speculations about his view, in his third phase, of the relation between the scientific and the manifest conceptions of the world. If by “scientific conception of the world” we mean the description of the world which is offered by physical theory, Putnam seems to think that that conception is not exhaustive, since it leaves out the features of the world and the features of man which make the world a possible object of evaluation and man a subject of evaluation. Thus, saying that the manifest conception is prior to the scientific conception implies that it is more inclusive: to it, the world is made of facts which can be evaluated (just as it is made of facts which are colored, tasty, etc.), whereas the scientific conception has no room for evaluable properties (although scientific practice *presupposes* values). This is by no mean to say that according to Putnam the scientific

conception of the world *is wrong*: what is wrong is to take it *to be exhaustive*, and this is the mistake of physicalism. Saying that the scientific conception of the world is not exhaustive is just to say that physics considers some features of our common sense reality and leaves others out. Economics should thus be autonomous from physics, since it considers a different set of features of reality, including (some) values. Direct realism, or “second naivety”, gives priority to the manifest conception in the sense that it does not take the fact that the sciences overlook some kinds of properties as a reason to consider those kinds of properties inexistent or illusory.

REFERENCES

- Lovibond, S. (2006). Review: “The Collapse of the Fact/Value Dichotomy and Other Essays”. *Mind*, 115(458), 457–460.
- Putnam, H. (1960). *Mind and Machines*, now in H. Putnam (1975), *Mind, Language, and Reality*. Cambridge: Cambridge University Press, pp. 362–385.
- Putnam, H. (1975). *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.
- Putnam, H. (1981). *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Putnam, H. (1988). *Representation and Reality*. Cambridge (MA): MIT Press.
- Putnam, H. (1994). *Words and Life*. Cambridge (MA): Harvard University Press.
- Putnam, H. (2001). *The Threefold Cord: Mind, Body, and World*. New York: Columbia University Press.
- Putnam, H. (2002). *The collapse of the Fact/Value Dichotomy and other Essays*. Cambridge (MA): Harvard University Press.
- Putnam, H. (2004). *Ethics without Ontology*. Cambridge, MA: Harvard University Press.

Book Review

Empiricism, Perceptual Knowledge, Normativity, and Realism: Essays on Wilfrid Sellars

Willem A. deVries (Ed.)

Oxford University Press, Oxford, 2009

*Luca Corti**

luca.corti@hotmail.de

«What makes a philosophical system interesting? After all is said and done, the answer seems quite simple and perhaps even trivial: it survives.» (Pitt, 1978, p. 1). When Joseph Pitt set out this criterion when introducing a series of essays on Sellars, in 1976, Sellars was still living. The survival capacity of Sellars' ideas could hardly be imagined.

The philosophy of Wilfrid Sellars stands, in the twentieth century Anglophone philosophical landscape, as a large and systematic whole. Its importance for the contemporary philosophical debate is nowadays widely acknowledged. The list of philosophers who owe much to Sellars' views is long and well-known: Daniel Dennett, Ruth Millikan, Michael Williams, Robert Brandom, John McDowell, etc. We can well imagine the difficult task that future historians of philosophy (or historians of ideas) will have in tracing back and disentangling the many threads connecting Sellars to the contemporary discussion in a great number of fields, such as philosophy of language, philosophy of mind and philosophy of science. At any rate, time has to pass before this can be done in a conclusive way, for Sellarsian insights do not just "survive", they are still living and exercising their influence. As often emphasized by his interpreters, Sellars' systematic attitude alongside his "all-encompassing" account of philosophy¹, led him to elaborate a unitary system of deeply interrelated positions. Despite this, one of the distinctive and

¹ «The aim of philosophy [...] is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term» (*PSM*, p.1).

probably most fertile features of Sellars' scholarship is its variety. Above all in Anglophone philosophy, his ideas have originated a number of different approaches among the two generations of Sellars-inspired thinkers. Once a gateway into the system is undertaken, different pathways open up, spreading in a wide range of exegetical and theoretical directions. Like John's necktie of his famous example (*EPM*, § 14), Sellars' philosophy seems to change according to the light under which it is observed.

The volume edited by De Vries, who is already author of a comprehensive monograph on Sellars, gives us an overview of the situation. It encompasses a series of contributions aiming to explore such an architecture, most of them stemming from a conference held in London 50 years after Sellars gave his famous lectures there on the "Myth of the Given". In order to find a way around them, I have chosen to focus on a few topics.

(i) "Perceptual experience" and "empiricism" are the first topics to be analyzed. The two opening essays are respectively John McDowell's and Robert Brandom's. In them we witness another episode of a dispute that has taken place in Pittsburgh for many years between the two philosophers on Sellars' work.² The question at stake is roughly the following: was Sellars an empiricist? Robert Brandom's answer is that he practically was not. His portrait of Sellars (Brandom, 1997, 2002) depicts a position close to his own project, guided by the motto «experience is not one of my words» (Brandom, 2000, p. 205, n. 7).

Pursuing this line of interpretation, Brandom focuses his essay on Sellars' treatment of modal vocabulary. He regards Sellars as conferring to modals a specific "expressive role"; they bring to light the inferential network that holds the skeleton of our space of reasons. Thus, as in Brandom's own system, Sellars develops a "two ply account" of perceptual experience. Our descriptive vocabulary derives its semantic and epistemic significance only from being caught in a «space of implication» (p. 50). Modals express the functional role a concept plays in such an inferentially articulated normative structure (pp. 57–58).

² The dispute is both theoretical and exegetical, see Brandom (1994, 2002) and McDowell (1994, 2009); for an overview see Machbeth (2009) and Rorty (1998, pp. 122-152).

Disputing this line of interpretation, John McDowell argues that Sellars' purpose in EPM is not to dismiss empiricism at all. In order to prove it, McDowell engages in an exegetical reading of the EPM text, which leads him to refute Brandom's "two ply" image, condemning his «attempt to read the project into Sellars» (p. 14). According to McDowell, Sellars' aim in EPM (and more generally) is just to revise empiricism in its traditional form, not to abolish it *tout court*. De Vries and Coates, in their joint essay, side with McDowell's criticism of Brandom, and focus on Sellars' treatment of the "looks"-statements in order to unmask what they call "Brandom's two-ply error". A passage in §26b of EPM – where Sellars seems to identify in *experiences* the object of a report – would particularly controvert the anti-empiricist reading:

When I say "X looks green to me now" I am reporting the fact that my experience is, so to speak, intrinsically, as an experience, indistinguishable from a veridical one of seeing that x is green. (p. 141)

This paragraph is particularly controversial and is quoted several times in the volume (pp. 18, 51, 124).

But De Vries and Coates take a further stance on the issue, since they do not follow McDowell in defending the idea that experience is conceptual throughout. On the contrary they argue that «there is a further dimension to experience [...]. In addition to the propositional content of experience, there is a further sensory (or phenomenal) nonconceptual component» (p. 133). Thus, the debate between conceptualists and nonconceptualists on experience comes into focus, providing an outlook on a further division among Sellars' interpreters.

ii) As stated, Coates rejects McDowell's idea of experience as conceptual "all the way down" (p. 86). Moving from his interpretative point and following the same intention of squeezing theoretical juice out of Sellars' views about perception³, he exploits – and partly modifies – Sellars' conception of the faculty of imagination. He regards imagination as a crucial moment in a gradual process of conceptualization that constitutively articulates our perceptual experiences; imagination is to be conceived as a dispositional understanding,

³ I borrow this expression from Brandom, p. 59.

which allows us to «have a set of *implicit* expectations about how the nonconceptual component of experience might change» (p. 76).

Snowdon is less indulgent with Sellars on several points. Sellars' «views in EPM about perception» (p. 129) is one of them. On the one hand, in Snowdon's eyes Sellars' arguments do , not conclusively refute the orthodox sense-datum theory, which Sellars meets to some extent (p. 127). On the other hand, Sellars, in elaborating “The Myth of Jones” takes a controversial path by modeling inner episodes on linguistic ones. «Anyone with children realizes that their earliest model of the world [...] includes such manifestly noticeable inner happenings as pains and itches» (p. 128), thus considering the existence of pain as dependent on language acquisition «is surely an indication that something is seriously wrong in Sellars's model» (p. 128). The reader faces here a typical objection concerning «children, mutes and animals», which often recurs in debates on Sellars' account of mental episodes, at least since his correspondence with Chisolm⁴.

This is not the place to contemplate a possible reply. Nevertheless, we can agree that in order to gain full understanding of those issues and their Sellarsian treatment, a consideration of other topics – as well as a contextualization within a wider perspective of Sellars' systematic thought – seems necessary. Partly this is done in the essays that follow.

iii) Both authors of a comprehensive monograph on Sellars⁵, O'Shea and De Vries' respective essays represent, in my view, the core of the volume. The previous considerations on empiricism and perceptual experience, and the following discussions on the status of norms, picturing and the two-images of the world find their context here.

As displayed in a famous passage of his autobiographical reflections, one of the topics Sellars himself considered crucial for his system is normativity

⁴ The argument seems literally drawn from Chisolm: «Surely it would be unfounded psychological dogma to say that infants, mutes, and animals cannot have beliefs and desires until they are able to use language» (Chisolm, 1972, p. 222). See also Peacocke against McDowell: «Cats, dogs, and animals of many other species, as well as human infants, perceive the world, even though their conceptual repertoire is limited, and perhaps even nonexistent.» (Peacocke, 2001, p. 260). See also Rorty, 1998. In his answer to Chisolm, Sellars urges a more sophisticated understanding of his “thinking-out-loud” experiment.

⁵ De Vries 2005, O'Shea 2007.

(Sellars, 1975). Supported by Sellars' own words, James O'Shea regards Sellars' reflections on that issue as a privileged gateway to Sellars' system. He focuses on the relation between natural and normative, presenting one of the central insights of his recent book (p. 113). How can we reconcile a normative domain of concepts with a natural causally and nomologically structured world? It is a difficult task, for Sellars seems to hold two contradictory views: he accepts the preeminence of the causal explanation (summed up in his *scientia mensura dictum*), but he also defends the irreducible autonomy of the normative order. It is possible for him to fulfill this task, O'Shea argues, thanks to a double point of view from which one can observe the natural-normative dichotomy. The solution proposed by Sellars would consist in a bifocal consideration of linguistic-conceptual activity: *sub specie norma*, as a performance subject to normative standards and criteria, and at the same time, *sub specie causa*, as a natural item subject to causal explanation (in her essay Seibst will appropriately label this conception as the «double-life of linguistic items», p. 248). According to O'Shea, this view «can be shown to hold across the board for Sellars's views on the nature of meaning, intentionality, knowledge and truth» (p. 205).

By acknowledging that the thesis of «the autonomy of reason flirts with idealism» (p. 239; itself a protean term, probably worthy of a closer distinction), DeVries tends to conceive the “autonomy” of the normative in a manner which is close to O'Shea's views. He sees in Sellars «the idea of a self-sustaining, holistic system of rule-governed, contextually dependent, normative *types* embodied in natural *tokens*» (p. 235). He stresses that «Sellars's epistemology is realistic, not idealistic, from top to bottom», such that «Sellars naturalizes spirit rather than spiritualizing nature» (p. 230). How can we connect the two apparently irreducible domains? Given the idea of an autonomous normative reign of meaning, DeVries notices: «the notion of a direct semantic relation (whether meaning or reference) is as mythological as the notion of a given» (p. 239). Thus, a way to connect our linguistic descriptions to reality is called for. We must take into account another aspect of the complex Sellarsian system, De Vries concludes: the central notion of *picturing*.

iv) Correlation could eventually be seen as a common issue to the last two essays, devoted respectively to the topics of picturing and the two-images of the world.

Often neglected by “conceptualist” Sellarsian followers, the category of *picturing* is shown to play an essential role in mediating between causes and reasons (p. 250). Seibst looks closer into this relation, which treats linguistic items as purely natural objects among others – items *in rerum natura*. This radical “*flatus vocis* approach” allows Sellars to remain faithful to his dichotomy between norms and causes, while introducing at the same time a peculiar correspondentist element to his system. Jay Rosenberg, in his posthumous essay, addresses the issue moving from the correlation of the two-images of the world. He points out that «on Sellars’ view the very existence of any normative order at all can ultimately be explanatory accommodated only within a mature scientific image» (p. 293). In this respect, Rosenberg identifies an element of similarity between the two images, namely a common explanatory strategy based on the distinction between appearances and reality (p. 291). On this basis we could conceive the self-superseding of the manifest image and the shift to the scientific image, making sense of Sellars’ answer to the crucial question of «locating the normative conceptual order within the causal order of [...] nature» (p. 295).

Rosenberg’s contribution closes this survey of Sellars’ philosophical architecture. As outlined, the volume covers a wide range of important topics. The specialist reader will certainly appreciate the detailed arguments carried out in the scholars’ essays, whilst the non-specialist reader will taste a sample of the arguments that the authors have broadened in their own monographic works (partly devoted to fulfill specific Sellarsian insights, partly aiming at understanding his philosophy as a whole). In both cases, the reader will gain an overview of many of the interpretative and theoretical directions taken by scholars in Wilfrid Sellars’ name. This could be considered the main function of the book.

As expected, those who are seeking a univocal portrait of Sellars will be disappointed. The lack of unity might seem paradoxical for a thinker whose philosophy is often regarded as a whole where *tout se tient*. Nonetheless, this is a fruitful and distinctive feature well mirrored by the volume. Those who expect to find insights on the “historically-oriented” part of Sellars’ thinking will also be disappointed. Although reference to classical thinkers (Kant above all) had notably a great influence on Sellars’ thought, this feature is absent here (except for some incidental notes). This element could have helped draw a full-blown portrait of Sellars even from a theoretical point of view; a conference on the matter is yet to be held. Even so, the book is so rich in arguments that the

absence of this feature is not to be blamed. Mirroring perhaps its function in the contemporary Sellarsian revival, this volume shows us Sellars' philosophy more as a Jamesian "great *corridor-theory*" than as a systematic building: «a corridor in a hotel from which a hundred doors open into a hundred chambers» (James, 1906, p. 339). Thus, it raises a number of original insights on topics whose relevance for contemporary philosophical debate is undisputed.

REFERENCES

- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.
- Brandom, R. (1997). *Study Guide*. In W. Sellars (1997)
- Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge: Harvard University Press.
- Brandom, R. (2002). *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge: Harvard University Press.
- Chisholm, R. (1972). Intentionality and the Mental: Chisholm-Sellars Correspondence on Intentionality. In A. Marras (Ed.), *Intentionality, Mind and Language*. Chicago: University of Illinois Press, 214–248.
- De Vries, W.A. (2005). *Wilfrid Sellars*. Chesham: Acumen.
- Machbeth, D. (2009). Un'antinomia del giudizio empirico: Brandom e McDowell. In L. Ruggiu, & I. Testa (Eds.), *Lo spazio sociale della ragione, da Hegel in avanti*. Milano: Mimesis, 343–360.
- McDowell, J. (1996). *Mind and World*. Cambridge: Harvard University Press.
- McDowell, J. (2009). *Having the World in View. Essays on Kant, Hegel and Sellars*. Cambridge: Harvard University Press.
- O'Shea, J.R. (2007). *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge & Malden, MA: Polity Press.
- Peacocke, C. (2001). Does perception have a nonconceptual content?. *Journal of Philosophy*, 98(5), 239–264.

- Pitt, J. (1978). *The Philosophy of Wilfrid Sellars: Queries and Extensions*. Dordrecht: Reidel.
- Rorty, R. (1998). *Truth and Progress, Philosophical Papers*, vol. 3. Cambridge, NY: Cambridge University Press.
- Sellars, W. (1975). Autobiographical reflections. In H.-N. Castañeda (Ed.), *Action, Knowledge, and Reality: Critical Studies in Honor of Wilfrid Sellars*. Indianapolis: The Bobbs-Merrill Company, 277–293.
- Sellars, W. (1991). Philosophy and the Scientific Image of Man [*PSIM*]. In W. Sellars., *Science, Perception and Reality*. Atascadero, CA: Ridgeview Publishing Company, 1–40.
- Sellars, W. (1997). Empiricism and the Philosophy of Mind [*EPM*]. (edited by R.B. Brandom). Cambridge, MA: Harvard University Press.

HUMANA.MENTE - ISSUE 21, MAY 2012

Between two images

Edited by Carlo Gabbani

The Manifest and Scientific Conceptions of the
Human Being, 50 Years On

This Issue will appear 50 years after Sellars' very influential essay "Philosophy and The Scientific Image of Man", and also 100 years after his birth. Now more than ever, the issues raised by this essay are at the very core of many debates in epistemology, philosophy of mind, and philosophy of science.

For more information about the journal
visit our website at:
www.humanamente.eu

www.humanamente.eu

Humana.Mente – Journal of Philosophical Studies

was founded in Florence in 2007. It is a peer-reviewed international journal that publishes 4 issues a year. Each issue focuses on a specific theme, selected from among critical topics in the contemporary philosophical debate, and is edited by a specialist on the subject, usually an emerging researcher with both philosophical and scientific competence.

IN THIS ISSUE PAPERS BY

Willem de Vries, Kyle Stanford, David Hodgson,
Steven Savitt, Keith Lehrer, Giacomo Turbanti, Jay
Garfield, Owen Flanagan-Stephen Martin, James
O'Shea, Diego Marconi

HUMANA.MENTE
Journal of Philosophical Studies

ISSN: 1972 - 1293